

Symposium i anvendt statistik

2022-2

SYMPOSIUM
I
ANVENDT
STATISTIK

2022-2

Redigeret af Peter Linde
på vegne af organisationskomiteen for
Symposium i Anvendt Statistik

Støttet af SAS Institute Inc.

Forord

Det er symposiets formål at fremme information om såvel anvendt statistik som statistisk databehandling. Symposiet er tværfagligt med særlig vægt på metodik, formidling og fortolkning af statistiske analyser. I år er SEGES Innovation, vært for symposiet, hvilket vi gerne vil takke for. Symposiet arrangeres af Symposium i Anvendt Statistik og SEGES Innovation, Landbrug & Fødevarer. Symposiet i Anvendt Statistik er ansvarlig for det faglige program og økonomien.

Symposiet skulle oprindeligt være afholdt fysisk i januar 2022, men pga. Corona bliver det i stedet afholdt 29.-31. august 2022. Stadig i Axelborg, København. Symposiet har til formål at understøtte delingen af statistiske analyser, og alle forfattere har mulighed for at få publiceret både til januar og til august 2022. Dette er symposiebog 2022-2, der indeholder indlæggene fra august symposiet. Den mindre publikation 2022-1 indeholder indlæggene fra de forfattere, der har brugt muligheden for også at få publiceret i januar 2022. Dette års indlæg spænder over mange forskellige fagområder og lægger derudover vægt på metoder og analyser. Som det er normalt ved videnskabelige indlæg, er bidragsyderne ansvarlige for indholdet af indlæggene, og spørgsmål herom kan rettes direkte til forfatterne.

Med symposiet tilstræbes det at skabe et forum for tværfaglig inspiration og dialog for at udbygge kommunikationen mellem personer, der arbejder med beslægtede metoder inden for forskellige fagområder.

Peter Linde, Organisationskomiteen

ISBN 978-87-989370-2-9

Trykt hos PRinfoTrekroner i 110 eksemplarer

Organisationskomiteen for Symposium i Anvendt Statistik 2022

Lisbeth la Cour
Økonomisk Institut
Copenhagen Business School
Porcelænshaven 16A
2000 Frederiksberg
llc.eco@cbs.dk

Peter Linde
Statistisk konsulent
Granparken 187
2800 Lyngby
Peter@Brede.dk

Anders Milhøj
Økonomisk Institut
Københavns Universitet
Øster Farimagsgade 5 – B26
1353 København K
Anders.Milhoj@econ.ku.dk

Esben Høg
Matematiske Fag
Aalborg Universitet
Fredrik Bajers Vej 7
9220 Aalborg Ø
esben@math.aau.dk

Gorm Gabrielsen
Institut for Finansiering
Copenhagen Business School
Solbjerg Plads 3
2000 Frederiksberg
stgg@cbs.dk

Søren Möller
Faculty of Health Sciences
Syddansk Universitet
J. B. Winslows Vej 19
5000 Odense C
Moeller@health.sdu.dk

Helle M. Sommer
SEGES Innovation
Landbrug & Fødevarer
Axeltorv 3
1609 København V
hmso@seges.dk

Niels Kærgaard
Fødevare- og Ressourceøkonomi
Københavns Universitet
Rolighedsvej 25
1958 Frederiksberg
nik@life.ku.dk

Mogens Dilling-Hansen
Institut for Økonomi
Århus Universitet
8000 Århus C
dilling@econ.au.dk

Klaus Rostgaard
Kraeftens Bekampelse
Strandboulevarden 49
2100 København Ø
klar@cancer.dk

Jørgen Lauridsen
Økonomisk Institut
Syddansk Universitet
Campusvej 55
5230 Odense M
jtl@sdu.dk

Sara Armandi
SAS Institute
Købmagergade 7-9
1050 København K
Sara.Armandi@sas.com

Birthe Lykke Thomsen
Afdeling for Børn og Unge
Rigshospitalet
Blegdamsvej 9
2100 København Ø
btho0101@regionh.dk

Indholdsfortegnelse

Økonomi, præferencer og adfærd

The New Statutory Audit Framework in Europe: Consistency of Implementation Rationale and Au-dit Fee Dependence in Denmark	
<i>Claus Holm, Department of Economics and Business Economics, Aarhus University</i>	1
Cruising for parking in Copenhagen	
<i>Ismir Mulatic, CBS, and Jos van Ommeren, Department of Spatial Economics, VU Amsterdam</i>	11
Who is the Favorite Child? A Preliminary Descriptive Analysis Communicating Science	
<i>Sara Armandi, SAS Institute</i>	19
Høj begavelse i et økonomisk perspektiv	
<i>Freja Falk, Institut for økonomi, Københavns Universitet</i>	35

Statistisk metode

Classification using binary and continuous variables	
<i>Guillermina Eslava and Gonzalo Pérez, Faculty of Sciences, National Autonomous University of Mexico Universitet</i>	45
Inferens for den relative risiko, når estimatet er 0	
<i>Søren Møller, SDU og Odense Universitetshospital, og Linda Juel Ahrenfeldt, SDU</i>	54
Studying Preference of Pork Hedonic or Paired comparisons	
<i>Gorm Gabrielsen, CBS, in cooperation with M. D. Aaslyng, S. Støjer, E. V. Olsen, D. V. Byrne, and W. L. P. Bredie</i>	57

Data, estimation og Survey

National Overvågning af Arbejdsmiljøet blandt Lønmodtagere (NOA-L)	
<i>Hans Bay, Arbejdstilsynet</i>	58
Risikovillighed - En analyse af faktorer, der påvirker unges villighed til at påtage sig risiko	
<i>Mogens Dilling-Hansen, Department of Economics and Business Economics</i>	69
Imputering af gennemsnit med regressionsmodeller. Et mål for kvaliteten af imputeringsmatricer fra et datasæt til et andet.	
<i>Peter Linde, statistisk konsulent</i>	80

Statistiske analyser

Google searches linked to Apple stock volatility	
<i>Niels Buus Lassen, Centre for Business Data Analytics, CBS</i>	86
Vil Polen opleve et Exit (Polexit)?	
<i>Agnieszka Szmydyska, Økonomisk Institut, Københavns Universitet</i>	109
En statistisk tilgang til Spritbådene mellem Danmark og Tyskland i et historisk perspektiv	
<i>Lars Holmguard Bak, Rigsarkivet, Aabenraa, og Nils Karl Sørensen Økonomisk Institut, Syddansk Universitet</i>	125

Landbrug og analyse

Modeller for social genetiske effekter hos grise og betydningen for dyrevelfærd <i>Bjarne Nielsen, QGG, Aarhus University, Lizette V. Horndrup, Danish Agriculture & Food Council, and Birgitte Ask, SEGES Innovation, Breeding and Genetics.....</i>	125
Deck height during transport of weaners – piglet height and microclimatic conditions inside the vehicle <i>Leslie Foldager, Marianne Kaiser, Guoxing Chen, Line D. Jensen, Jens Kr. Kristensen, Cecilie Kobæk-Kjeldager, Karen Thodberg, Li Rong and Mette S. Herskin, Aarhus University</i>	129
Er det ok at stoppe et forsøg før tid? <i>Helle Mølgaard Sommer, SEGES Innovation.....</i>	143

Undervisning, metode og SAS

En statistisk sammenligning af danske skoleelevers læsefærdigheder før og efter Folkeskolereformen 2014 <i>Josefine Palmø og Sophie A. Steiness, Økonomisk Institut, Københavns Universitet.....</i>	149
The Eigenvector-Eigenvalue Identity Applied to Fast Calculation of polSAR Scattering Characteriza <i>Allan Aasbjerg Nielsen, Compute, Department of Applied Mathematics and Computer Science, DTU.....</i>	161
Nyheder i SAS <i>Anders Milhøj, Københavns Universitet</i>	166

Økonomi og samfund

Danske økonometriske makromodeller – i medgang og modgang <i>Jesper Jespersen, Roskilde Universitet</i>	174
MAKRO-modellen <i>Peter Stephensen, Martin Bonde, Joao Ejarque, Grane Høegh og Anders Kronborg, DREAM</i>	182
Data og statistisk metode ved økonomisk modelbygning: Udviklingen fra 1970'erne til i dag <i>Niels Kærgaard, Fødevare- og Ressourceøkonomi, Københavns Universitet</i>	196
Estimering af input-output koefficenter fra aggregeret data på virksomhedsniveau <i>Rasmus Seneberg Zitthen, Arne Henningsen, Simon Alexander Andreasen, Mads Frandsen og Mathias Struck Jürgensen, Institut for Fødevare- og Ressourceøkonomi, Københavns Universitet</i>	204

Sundhed

Is stress regionally persistent? <i>Jørgen T. Lauridsen, Department of Economics, University of Southern Denmark</i>	218
Blev vi danskere mere eller mindre syndige under coronaen <i>Anders Milhøj, Københavns Universitet</i>	226
PTSD in school-age children: A nationwide prospective birth cohort study <i>Mogens Nygaard Christoffersen, VIVE, and Anne Amalie Elgaard Thorup, Region Hovedstaden</i>	238
Extreme group analysis of patient cost of antibiotic prescribing among general practitioners <i>Troels Kristensen, Danish Centre for Health Economics (DaCHE), SDU, Charlotte Ejersø, Department of Endocrinology, Odense Universitets Hospital, and Jens Søndergaard, Research Unit of General Practice, SDU</i>	248

The New Statutory Audit Framework in Europe: Consistency of Implementation Rationale and Audit Fee Dependence in Denmark

Claus Holm, Department of Economics and Business Economics, Aarhus University, email hoc@econ.au.dk.

Abstract: Individual EU Member States had options on how they implement the new statutory audit framework in Europe. They could introduce stricter rules or apply certain exemptions where deemed appropriate. The aim of this study is to contrast the arguments from the implementation process with an examination of fee dependence related to the provision of non-audit services to PIEs. The audit reform introduced ‘cap’ and ‘blacklist’ measures on non-audit fees, which implied a regulator-determined condition of non-independence. In a sample with 19.021 observations, Denmark, Finland, Germany, Sweden and the UK are compared to determine the consistency of implementation rationale and audit fee dependence. The findings support the regulators’ concern that auditors of public interest entities with high levels of non-audit services are more likely to have an economic dependence problem or at least face a problem with independence in appearance. Findings also suggest, that inconsistent with the implementation rationale, Denmark stands out as a country with greater challenges related to the auditors’ provision of non-audit services.

Introduction

A major conflict is at the heart of the provision of non-audit services to public interest entities (PIEs). In addition to the statutory audit, audit firms have traditionally provided an array of services based on their experience and expertise within business-related matters. The value of knowledge spillovers amongst various services provided by audit firms has been stressed by both corporations and the accounting profession (Francis, 2006; Daugherty and Tervo, 2008). Audit market regulators have taken the opposite stance, advocating the better interest of the financial statement users. Regulators overseeing the major financial markets (including the US and European Union) have strenuously questioned the possibility of audit firms retaining independence from, especially, the large PIE clients when these clients pay for both the statutory audit and non-audit services (DeFond and Zhang, 2014; Quick, 2012). In Europe, the recent audit reform has amended the statutory audit Directive (2014/56/EU), and the associated EU Regulation 537/2014 prohibits a number of non-audit services and further cap the volume of allowed non-audit services to PIEs.

The audit reform introduced ‘cap’ and ‘blacklist’ measures on non-audit fees, which implied a regulator-determined condition of non-independence. While the efficacy of the

intended audit reform represented by Regulation 537/2014 can be contested, individual countries also faced options in the national implementation for introducing stricter rules or applying certain exemptions where deemed appropriate. The latter national legislation process is sometimes termed ‘law in action’ as opposed to the transnational ‘law on the books’ (Samsonova-Taddei and Humphrey, 2015, 59). The aim of this study is to contrast the arguments from the implementation process with an examination of fee dependence related to the provision of non-audit services to PIEs. The aim responds to a general concern for regulatory implications of ‘mindless argumentation’ across different settings (countries). Thus, in an international sample, five countries are compared to determine whether independence issues are similar in nature and/or more pronounced in particular Member States otherwise comparable on different dimensions.

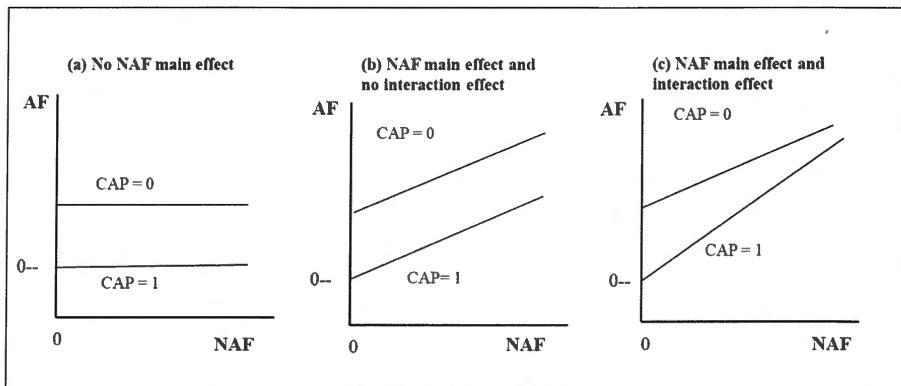
Empirical context

The countries considered in this study (Denmark, Finland, Germany, Sweden and UK) belongs to the group of 28 (now 27) Member States in which the audit reform was mandated by the European Commission to take effect no later than 16 June 2016. To contrast the extent of existing independence problems with the likely effect of the European audit reform, this study examines the relationship between the audit fee and non-audit fee starting before the audit reform. The expectation is that the introduced ‘blacklist’ will decrease non-audit fees, while increased fee competition leading to lower audit fees is one possible effect of new tendering rules. However, in terms of the effects of capping the level of fees for non-audit services, the wait for ex post regulation effects is longer because the auditor has to have been providing audit and non-audit services to the audited PIE for a period of three or more consecutive years before the cap applies. The established fee model methodology is chosen for this study. The regulators’ introduction of the ‘blacklist’ and ‘capping’ measures related to the auditors’ provision of the non-audit services has established a predefined threshold of ‘assumed non-independence’. Hence, the assumption of known threshold conditions for non-independence allows for a stronger controlled setting, which has not been available in previous studies of the relationship between audit fees and non-audit fees. In the years before the introduction, the threshold is counterfactual in terms of permissible fee dependencies.

The empirical tests in this study consider the possible fee dependency in the period 2010 to 2020. Specifically, an interaction effect approach is applied to identify whether and to what extent unwanted fee relationships existed. Figure 1 illustrates alternative projected relationships - *ceteris paribus* - between the audit fee (AF) and non-audit fee (NAF). Based on prior research, the absence of a direct relationship (main effect) between AF and NAF is considered indicative that the audits are without independence problems pertaining to

economic bonding. The loss-leader argument alone would imply that the absence of a negative association supports this argument while the possibility that the audit firm may retain cost savings from knowledge spillovers implies that a significant main effect in either direction potentially imposes such concern. To highlight the potential differences in the projected relationships, the intercepts for the two groups above and below the 70 percent cap in Figure 1 are drawn as being different. Because companies belonging to the group where the cap is exceeded ($CAP = 1$) by default acquire both audit services and non-audit services, the intercept is depicted as close to zero, whereas the other group ($CAP = 0$) will have some companies that do not acquire NAF and, thus, have a somewhat higher intercept with the AF axis.

Figure 1. Projected relationships between AF and NAF



In examining the importance of the cap measure, the existence of a NAF main effect is not the prime issue. The major interest is to see if there is a difference in slope between the two groups. The concern is whether the association between AF and NAF is stronger when a certain threshold is reached. The regulators' suggestion is that above the threshold the company–auditor relationship is tainted with an independence problem. The group of companies with a critical non-audit fee proportion ($CAP = 1$) will have a higher slope coefficient if the association between AF and NAF is stronger. The calculation is made both as a ratio in the individual year (current year AF) and as a ratio of NAF to 3 years' average AF (with the latter as the intended cap measure in the EU reform). A significant, stronger association (interaction effect) will suggest that the regulators' threshold is a good indicator of possible independence problems for this group. In order to examine whether the audit reform measures introduced changes in such fee dependencies a time indicator is also included in the testing. Due to the gradual introduction of the measures, such a POST

indicator is introduced and examined for the years available (i.e., post 2017, 2018 and 2019).

Methodology: Dataset, Model and Variables

The **dataset** contains company financial information and auditor-related data derived from Compustat Global, Audit Analytics and Orbis. For the years 2010–2020, the total sample population is 19.021 firm-year observations (after excluding financial institutions and observations due to missing variables). The sample observations are based on the 11 years and the five countries considered. The two largest subsamples are derived from the UK with a total of 8.753 observations and Germany with 4.301 observations. In comparison, the number of listed companies in each of the three Scandinavian countries is smaller. Sweden is the largest with 3.900 observations while the sample for Denmark holds 952 observations and Finland holds 1.115 observations. Fee dependence is examined using the **core audit fee determinants model** which has evolved from the seminal study by Simunic (1980). In the model audit fees (AF) are regressed on non-audit fees (NAF) as well as control variables, which in prior studies have been considered key determinants of audit fees.

$$AF = \beta_0 + \beta_1 NAF + \beta_2 CAP + \beta_3 POST + \sum \beta_i CONTROL\ VARIABLES_i + \sum \beta_j FIXED\ EFFECTS_j + \varepsilon \quad (1)$$

The **variables** identified in the fee models control for differences in client, auditor and engagement attributes. All fee measures are transformed using the natural logarithm to annual fees measured in thousand euros. The dependent variable in all the models is the fee for the statutory audit (AF) which in equation (1) is specified as a function of NAF (fee for non-statutory audit services); that is, the sum of all fees not related directly to the statutory audit (for example fee for other assurance services, fee for tax services and other consultancy services). In addition two indicator variables are included, that is CAP (dummy indicator for ratio of NAF to AF above 70 percent) and POST (dummy indicator for post observation year). The CONTROL VARIABLES can be subdivided into three types of attributes consistent with the typology used in the meta-analysis of prior fee studies by Hay (2013). First, the models include a number of variables that can be identified as *client attributes* related to size, complexity, performance and risk: the variable SIZE is measured as the natural logarithm of total assets in thousand euros; BUSSEG is the natural logarithm of the number of business segments of the company; GEOSEG is the natural logarithm of the number of geographical segments of the company; DEXDISC is an indicator variable

with the value of one if the company has extraordinary or discontinued items; INVREC is the ratio of the sum of inventory and receivables to total assets; ROA is the ratio of EBIT to total assets; LOSS is an indicator variable with a value of one if net income is negative; LEVERAGE is the ratio of debt to total assets; CAtoCL is the ratio of current assets to current liabilities. The models also include a set of industry indicators to control for industry-fixed effects. The industry indicators are based on the industry classification GICS (Global Industry Classification Standard) sectors. Second, the models include *auditor attributes* related to auditor quality: BIG4 is an indicator variable with the value one if a Big Four audit firm conduct the audit; SHORTTENURE is an indicator variable with the value of one if the auditor has conducted the statutory audit of the company for three years or less. Third, the models include *engagement attributes*: In addition to the control variables, it should be noted that the primary exploratory variable NAF can be classified as belonging to this category; BUZYSEASON is an indicator variable with the value of one if the company uses the calendar year as the financial year.

$$AF = \beta_0 + \beta_1 NAF + \beta_2 CAP + \beta_3 POST + \beta_4 CAP*NAF + \beta_5 POST*NAF + \beta_6 CAP*POST + \beta_7 CAP*POST*NAF + \sum \beta_i CONTROL\ VARIABLES_i + \sum \beta_j FIXED\ EFFECTS_j + \varepsilon \quad (2)$$

The fee dependency issue is tested using a three-way interaction model as shown in equation (2). To test whether the main effect association between AF and NAF is stronger when the regulatory threshold is reached, the NAF variable is interacted with the indicator variable CAP; to test whether the effect changes in the post period(s) the NAF variable is similarly interacted with the indicator variable POST; and finally the marginal effect is tested using the interaction with both CAP and POST.

Analyses

The descriptive statistics for the five countries and the pooled total sample are provided in Table 1. The top part of the table provides an overview of the distribution of the continuous variables (by mean and standard deviation). The following may be noted in terms of comparison of the individual countries. In terms of SIZE, the Swedish sample has the smallest companies on average, however probably due to more observations available than in Finland and Denmark. Leverage is, on average, about 47 percent, with UK slightly below average. The ROA is not consistent across the five countries. That is, the average ROA is -2.1 percent for the pooled sample but ranging from -5.3 percent in Sweden to 3.6 percent in Finland for the period 2010–2020. Complexity, as measured by LNBUSSEG and LNGEOSEG, seems to be the highest in Finland while the companies from Sweden and UK have low complexity with respect to these client attributes. In relation to the indicator

variables shown in the lower part of Table 1, some variation is discernible. The proportion of BIG4 audits is high (on average 66 percent). The German percentage is notably lower with 58 percent, whereas the Danish and Finnish companies in the sample are almost exclusively audited by the big four audit firms. In terms of LOSS, UK has the highest proportion (42 percent for the period) and Germany has the lowest (24 percent). Amongst the five countries, most audits are conducted with the calendar year as the financial year (BUZYSEASON). UK as the largest audit market in the sample has a considerably lower proportion of 51 percent calendar year audits.

Table 1. Descriptive statistics for 5 countries and total sample (2010-2020)

Variables	Denmark N=952		Finland N=1.115		Germany N=4,301		Sweden N=3,900		UK N=8,753		Total sample N=19,021	
Continuous	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
LNAF	5,310	1,537	5,407	1,412	5,456	1,528	4,581	1,587	5,055	1,508	5,082	1,555
LNNAF	4,808	1,846	4,285	2,010	3,479	2,497	3,216	2,142	3,503	2,286	3,550	2,302
SIZE	19,077	2,222	19,300	1,990	19,347	2,356	17,704	2,338	18,412	2,389	18,564	2,413
LEVERAGE	0,492	0,205	0,551	0,148	0,531	0,203	0,461	0,221	0,429	0,239	0,469	0,226
INVREC	0,284	0,174	0,331	0,180	0,312	0,180	0,298	0,210	0,237	0,193	0,274	0,196
ROA	0,026	0,236	0,036	0,185	0,033	0,153	-0,053	0,308	-0,044	0,397	-0,021	0,322
CAtoCL	2,271	3,902	1,646	1,236	3,117	23,403	5,268	81,500	3,673	14,772	3,685	39,845
LNBUSSEG	0,841	0,722	0,904	0,754	0,870	0,673	0,516	0,709	0,759	0,799	0,747	0,759
LNGEOSEG	0,872	0,817	1,176	0,760	0,894	0,760	0,805	0,868	0,773	0,792	0,835	0,806
Indicators	n	pct.	n	pct.	n	pct.	n	pct.	n	pct.	n	pct.
B4	903	95%	1,075	96%	2,485	58%	3,260	84%	4,762	54%	12,485	66%
SHORTTENURE	244	26%	243	22%	1,312	31%	941	24%	2,497	29%	5,237	28%
LOSS	271	28%	275	25%	1,033	24%	1,448	37%	3,638	42%	6,665	35%
BUZYSEASON	711	75%	1,080	97%	3,747	87%	3,586	92%	4,450	51%	13,574	71%
DEXTDIS	142	15%	124	11%	400	9%	204	5%	1,176	13%	2,046	11%

Table 2 provides further descriptive statistics related to fee dependency. The two variables considered are (1) NAF/AF measured as the average company ratio of NAF to 3 years average of AF, and (2) %>CAP measured as the proportion of companies with ratio of NAF to 3 years of AF above 70 percent. Over the sample period, the average for the total sample indicateds that NAF/AF has been below 70 percent and decreasing over til periode from approximate 60 percent to 28 percent. In view of the intention by regulators to diminish this ratio this seems to have worked. The proportion of companies above the cap level has also decreased – from about 26 percent to 9 percent in 2020. However, these average observations across the five countries are very different in the individual countries. Before the reform Denmark and Finland were at the high end in terms of NAF/AF ratio as well as the proportion of companies above the 70 percent benchmark. It is not until 2020 that Danish companies on average are below the intended cap level, while Finish companies on average accomplished this in 2015. In Germany, Sweden and the UK the average NAF/AF ratio has been considerably lower, even though a noticeable proportion of PIE audits in all countries has been over the cap level. As explained above, the blacklist measure which came into effect first had the intention to diminish the amount used on NAF; which the

development in the individual countries support. The combined regulatory effect of the blacklist and cap measures is strictly speaking only in place at the very end of the period. In terms of country differences it is noticeable that Denmark seem to be very accustomed to provide non-audit service by the statutory auditors, while the tradition in Germany seems to be quite opposite.

Table 2. Development in ratio NAF/AF and proportion of audits above 70 percent cap measure in 5 countries

Year	Denmark		Finland		Germany		Sweden		UK		Total sample	
	NAF/AF	% >CAP	NAF/AF	% >CAP	NAF/AF	% >CAP	NAF/AF	% >CAP	NAF/AF	% >CAP	NAF/AF	% >CAP
2012	0,83	40%	0,79	40%	0,42	19%	0,60	28%	0,63	26%	0,59	26%
2013	1,08	44%	0,67	35%	0,46	18%	0,49	22%	0,62	25%	0,59	25%
2014	0,95	44%	0,80	34%	0,47	22%	0,49	24%	0,65	26%	0,60	26%
2015	0,88	44%	0,66	32%	0,42	18%	0,50	22%	0,57	22%	0,54	23%
2016	1,04	56%	0,58	28%	0,40	16%	0,56	25%	0,44	16%	0,50	20%
2017	0,84	42%	0,59	25%	0,29	11%	0,43	18%	0,46	15%	0,44	17%
2018	0,74	44%	0,64	28%	0,31	13%	0,43	17%	0,36	12%	0,40	16%
2019	0,79	41%	0,51	15%	0,28	8%	0,43	14%	0,29	11%	0,36	13%
2020	0,61	27%	0,30	12%	0,20	5%	0,37	13%	0,23	6%	0,28	9%

Legend: NAF/AF measured as average company ratio of NAF to 3 years average of AF

%>CAP measured as proportion of companies with ratio of NAF to 3 years average of AF above 70%

Table 3 shows the multivariate audit fee models based on the total sample for the five countries. The OLS models are effective with an adjusted R-square of approximately 88 percent. The models are estimated based on robust standard errors; however, the adjusted R-squares are reported to compare explanatory power to previous research, i.e. based on OLS regression before the control for robust standard errors (Wooldridge, 2009). The model is not affected by serious multicollinearity amongst variables (max VIF of 3.24; not shown in table). Columns I and II use full annual observations presuming an indicator CAP measure calculated on an annual basis, while columns II-V are based on a reduced number of observations due to calculation of the CAP measure (indicator for ratio of NAF over average of 3 years AF above 70 percent). The main effects fee model shown in column I demonstrates that the fee model has an appropriate specification in terms of explanatory variables which all (except one) are highly significant and in the predicted directions. SHORTTENURE is the only exception without explanatory power in this sample. Columns II-V demonstrate the outcome of three-way interactions where the continuous NAF variable is interacted with the indicator CAP and POST variables. Ceteris paribus, the interaction models demonstrates that the fee dependence is best depicted as illustration (c) in Figure 1 when discarding the timing. The significant CAP-NAF interaction suggest, there is a stronger positive relationship between audit and non-audit fees for companies with a proportion of NAF to AF above the cap level than it is for companies below the cap level. The latter supports regulators' concern that companies with high levels of non-audit services may have an economic dependence problem or at least face a problem with independence in appearance. When timing is considered, the results seem to differ dependent on when post-regulation effects are isolated. Columns III-V use three different

POST measures 2017, 2018 and 2019 (e.g., POST 2017 is an indicator variable with the value of 0 for 2010- 2017 and 1 for 2018-2020). Only when the POST variable distinguish between years up till 2018 and the two remaining years, a negative interaction effect is identified. The negative signs in itself suggests the reduction of NAF fee dependencies as indicated by the ratios reported in table 2. However, for companies above the CAP level in the POST period the significant positive three-way interaction suggests marginally increased fee dependence as compared to companies below the CAP level.

Table 3. Main and interaction fee models (total sample)

	CURRENT CAP (NAF/AF>70)		AVERAGE CAP (NAF/3 YEARS AF >70)		
	PRE/POST 2017	PRE/POST 2017	PRE/POST 2017	PRE/POST 2018	PRE/POST 2019
			I	II	III
NAF	0.171***	0.169***	0.142***	0.142***	0.139***
CAP	-0.495***	-1.228***	-1.109***	-1.121***	-1.139***
POST	0.223***	0.283***	0.211***	0.230***	0.213***
CAP*NAF		0.136***	0.130***	0.132***	0.135***
POST*NAF		-0.0178***	-0.00926	-0.0143**	-0.00779
CAP*POST		-0.345***	-0.227**	-0.281***	-0.470***
CAP*POST*NAF		0.0685***	0.0369**	0.0404**	0.0677**
Constant	-4.623***	-4.391***	-4.810***	-4.809***	-4.806***
CONTROLS	Y	Y	Y	Y	Y
YEAR FE	Y	Y	Y	Y	Y
INDUSTRY FE	Y	Y	Y	Y	Y
COUNTRY FE	Y	Y	Y	Y	Y
F-Value	541.42***	551.22***	510.77***	516.14***	519.01***
Observations	19,021	19,021	14,485	14,485	14,485
R-squared	0.875	0.879	0.883	0.883	0.883

*** p<0.01, ** p<0.05, * p<0.1; Robust standard errors (cluster by company)

Table 4.A shows the multivariate audit fee models for each of the five countries and Table 4B summarizes the intercepts and NAF slopes for each combination of CAP and POST as demonstrated using the rearranged equation (3) model. Overall, the three-way interaction models demonstrate that across the countries the auditor fee dependency have similar patterns, with Denmark as a noticeable exception. First, a common observation is that the fee dependency is significantly higher above than below the 70 percent CAP level in all countries considered. For example for Germany the NAF slope coefficient is 0.290 for above CAP observations as compared to 0.129 below CAP for the pre-regulation period 2010-2017 and the slope coefficient is 0.339 for above CAP compared to 0.126 below CAP in the post regulation period 2018-2020 (using POST 2017 as indicator). Second, for Finland, Germany, Sweden and UK there is no POST*NAF interaction effects, while it is noticeable that Denmark stands out in terms of the effect shown for the POST period.

Specifically, the marginal effect shown by the three-way interaction suggests an increase in fee dependency in the POST period for those above the CAP compared to those below the CAP (positively significant effect at the 10 percent level for POST 2017 and at the 5 percent level for POST 2018). Third, Denmark stands out by having higher fee dependencies overall as expressed by significantly higher slope coefficients as compared to the remaining four countries across the four different settings, that is below CAP in the pre and POST period as well as above CAP in the pre and POST period.

Table 4.A Country comparison interaction models (PRE/POST 2017)

	DENMARK	FINLAND	GERMANY	SWEDEN	UK
NAF	0.346***	0.138***	0.129***	0.156***	0.129***
CAP	-1.085***	-1.038***	-1.308***	-0.998***	-1.040***
POST	0.476**	0.342*	0.134**	0.0439	0.318***
CAP*NAF	0.0859*	0.117***	0.161***	0.128***	0.116***
POST*NAF	-0.0839*	-0.0210	-0.00344	0.00201	-0.00832
CAP*POST	-0.589*	-0.268	-0.360	-0.149	0.221
CAP*POST*NAF	0.118*	0.0357	0.0526	0.0197	-0.0527
Constant	-3.553***	-6.785***	-4.527***	-5.260***	-4.163***
CONTROLS	Y	Y	Y	Y	Y
YEAR FE	Y	Y	Y	Y	Y
INDUSTRY FE	Y	Y	Y	Y	Y
F-value	94.17***	..	103.78***	359.67***	280.98***
Observations	748	860	3,285	2,982	6,610
R-squared	0.909	0.868	0.864	0.903	0.890

*** p<0.01, ** p<0.05, * p<0.1; Robust standard errors (cluster by company); AVERAGE CAP (NAF/3 YEARS AF >70)

Table 4.B Country comparison of intercept and slope coefficients (pre/post 2017)

CAP	POST	DENMARK		FINLAND		GERMANY		SWEDEN		UK	
		Intercept	Slope								
0	0	-3,553	0,346	-6,785	0,138	-4,527	0,129	-5,260	0,156	-4,163	0,129
0	1	-3,077	0,262	-6,443	0,117	-4,393	0,126	-5,216	0,158	-3,845	0,121
1	0	-4,638	0,432	-7,823	0,255	-5,835	0,290	-6,258	0,284	-5,203	0,245
1	1	-4,751	0,466	-7,749	0,270	-6,061	0,339	-6,363	0,306	-4,664	0,184

AF = intercept + slope * NAF + Controls + FE

AF = $(\beta_0 + \beta_2 \text{CAP} + \beta_3 \text{POST} + \beta_6 \text{CAP*POST}) + (\beta_1 + \beta_4 \text{CAP} + \beta_5 \text{POST} + \beta_7 \text{CAP*POST}) * \text{NAF} + \text{Controls} + \text{FE}$

Conclusion

In this study, the audit reforms' introduction of the 'blacklist' and 'capping' measures related to the auditors' provision of the non-audit services has established a predefined threshold of 'assumed non-independence'. In a sample with 19,021 observations, the five countries are compared to determine the consistency of implementation rationale and audit fee dependence. The saliency of the *minimum implementation rationale* in the EU audit reform implementation process is contested by examining the influence of the proposed

70% cap measure in five individual countries including Denmark, Finland, Germany, Sweden and the UK. In the public hearing process, it was argued that this rationale also was favoured by the countries and, thus, by regulators deemed appropriate due to similar circumstances for the transposition of the audit reform into national legislation. However, the preconditions for auditor independence problems seem to vary across these countries. Over the sample period 2010-2020, the average for the total sample indicateds that the ratio of NAF to three years average AF has been below 70 percent and decreasing over til period from approximate 60 percent to 28 percent. In view of the intention by regulators to diminish this ratio this seems to have worked. The proportion of companies above the cap level has also decreased – from about 26 percent to 9 percent in 2020.

Overall, the three-way interaction models demonstrate that across the countries the auditor fee dependency have similar patterns, with Denmark as a noticeable exception. The significant CAP-NAF interaction suggest, there is a stronger positive relationship between audit and non-audit fees for companies with a proportion of NAF to AF above the cap level than it is for companies below the cap level. The latter supports regulators' concern that companies with high levels of non-audit services may have an economic dependence problem or at least face a problem with independence in appearance. Findings also suggest that the implications of the harmonisation process has been and will continue to be different across countries. Inconsistent with the minimum implementation rationale, Denmark stands out as a country with greater challenges related to the auditors' provision of non-audit services.

References

- Daugherty, B. E. and Tervo, W. A. (2008) 'Auditor changes and audit satisfaction: Client perceptions in the Sarbanes-Oxley era of legislative restrictions and involuntary auditor change', *Critical Perspectives on Accounting*, 19, 931-951.
- DeFond, M. L. and Zhang, J. (2014) 'A review of archival auditing research', *Journal of Accounting & Economics*, 58, 275-326.
- Francis, J. R. (2006) 'Are Auditors Compromised by Nonaudit Services? Assessing the Evidence', *Contemporary Accounting Research*, 23, 747-760.
- Hay, D. (2013) 'Further Evidence from Meta-Analysis of Audit Fee Research', *International Journal of Auditing*, 17, 162-176.
- Quick, R. (2012) 'EC Green Paper Proposals and Audit Quality', *Accounting in Europe*, 9, 17-38.
- Samsonova-Taddei, A. and Humphrey, C. (2015) 'Risk and the construction of a European audit policy agenda: The case of auditor liability', *Accounting, Organizations & Society*, 41, 55-72.
- Simunic, D. A. (1980) 'The Pricing of Audit Services: Theory and Evidence', *Journal of Accounting Research*, 18, 161-190.
- Wooldridge, J. M. (2009) *Introductory Econometrics: A Modern Approach*, Ohio, USA, South-Western Cengage Learning.

Cruising for parking in Copenhagen

Ismir Mulalic*

Jos van Ommeren†

1 Introduction

Cities often use parking pricing to control the use of cars in urban areas (Marsden, 2006). Copenhagen is not an exception. Central Copenhagen is divided into tariff zones and parking fees vary across zones, time of day, and workdays vs. weekends (Gragera et al., 2021). The parking zones closest to city centre (the historic district) are the most expensive.

In order to successfully manage traffic congestion and to reduce air pollution, cities must limit excessive cruising for parking (Vickrey, 1954; Verhoef et al., 1995; Arnott et al., 2015; Zakharenko, 2016). This however requires knowing the occupancy rate and the level of cruising for parking at different times and locations in the city. Therefore some cities undertake manual or automatic counting of cars cruising for parking, or undertake surveys in which drivers are asked to report their cruising times (Shoup, 2006).¹ Estimating cruising by using administrative parking data is however a better alternative, as compared to more expensive and much more constrained methods based on surveys and experiments.

We extract two key variables (parking arrivals and vacancies) from administrative parking data for Copenhagen and estimate the external cost of marginally increasing parking duration (marginal external cost of (on-street) parking (MECP)) based on a new methodology introduced by van Ommeren et al. (2021).² This methodology has several advantages over other commonly applied methods. It is non-parametric and therefore the estimates do not rely on parametric assumptions. More importantly, this methodology provides estimates for the whole distribution of MECP over time and space and not only an average. Finally, it does not require causal inference. It is therefore easier to implement and requires fewer data.

We estimate the MECP for Copenhagen at the street level for half-hour intervals, and obtain time-varying and spatially detailed information. Our empirical findings suggest that at small reductions in parking fees might be welfare improving.

*Department of Economics, Copenhagen Business School.

†Department of Spatial Economics, VU Amsterdam.

¹Some cities undertake experiments. Drivers are asked to park and then cruising time is recorded.

²The MECP is an external cost because, while deciding their parking durations, drivers ignore cruising for parking by other drivers who aim to park at the same location (Inci et al., 2017).

2 Theoretical Framework

This section describes the theoretical framework introduced by van Ommeren et al. (2021) that guides our empirical methodology.³

In the model, time, denoted by t , is continuous and space is homogeneous. Drivers search in a street (or an area) with a given number of parking spots with a rate of $I(t)A(t)$. $A(t)$ is the *exogenous* rate of entry into the model at time t and $I(t)$ denotes the *endogenous* decision to search at time t . The total number of parking spots per street is N .

Drivers incur the search cost to park in this street. Parked cars leave at time t' after a (self-chosen) duration $\tau(t')$. Each driver has the same value of (search) time of c . A driver's walking time to the final destination is assumed to be proportional to in-vehicle search time with an exogenous walking-time multiplier of $\psi \geq 1$.⁴ A car that continues to park at time t during a certain day (from $t = 0$ to T) imposes a marginal search externality on arriving drivers (MECP).

The number of parked cars $n(t)$, is then

$$n(t) = \int_0^{\tau(t)} [I(t-\tau)A(t-\tau)] d\tau. \quad (1)$$

The vacancy rate is, consequently, $v(t) = 1 - n(t)/N$.

A driver randomly samples the parking spots at a rate of r (e.g., one parking spot per second) from a street with a vacancy rate of $v(t)$. The success of the parking search follows a Poisson distribution with a rate of $rv(t)$. The expected search time for an arriving driver at time t , $Z(t)$, is

$$Z(t) = \frac{\psi}{rv(t)}, \quad (2)$$

which includes walking time as well as in-vehicle search time.⁵ Notice here that the occupancy rate, $q(t) = n(t)/N$, and hence the vacancy rate, $v(t) = 1 - q(t)$, remains constant throughout the search process. Moreover, when drivers arrive at a street and decide to search, they also park in that street. Therefore the number of drivers parking in a given street at time t is equal to the number of drivers initiating search for parking in that street at time t . We use equation (2) to estimate the whole distribution of cruising time for each time and street.

van Ommeren et al. (2021) shows that the expected total search cost of all drivers who arrive at time t , $C(t)$, is

$$C(t) = \frac{c\psi}{r} \frac{I(t)A(t)}{v(t)}. \quad (3)$$

The MECP imposed by one additionally parked car during time t on searching drivers is then

³To simplify notation, we present a version of the model with identical agents. All the presented and discussed results also hold in the model with heterogeneous drivers, see van Ommeren et al. (2021).

⁴Notice that in the absence of walking, $\psi = 1$.

⁵The assumption of a Poisson process means that the parking search can be described independently of time and space (Arnett and Williams, 2017).

equal to

$$MECP = \frac{\partial C(t)}{\partial \tau^i(t)} = \frac{\partial C(t)}{\partial n(t)} = \frac{\partial C(t)}{\partial v(n(t))} \frac{\partial v(n(t))}{\partial n(t)} = \frac{c\psi}{r} \frac{I(t)A(t)}{Nv(t)^2}. \quad (4)$$

We use equation (4) to *non-parametrically* estimate the MECP for each time and street. Notice here that, if a parked driver decides to increase parking duration τ , the number of parked drivers increases marginally, which decreases the vacancy rate and marginally increases the search time of newly arriving drivers. The marginal effect of an increase in duration τ by all parked drivers at time t on $n(t)$ is equal to $[I(t - \tau)A(t - \tau)]$.

Administrative parking data typically includes information about the arrival rate of drivers per parking spot in the block, $I(t)A(t)/N$, and the vacancy rate, $v(t)$. Both vary across streets and over time. To estimate MECP we also need information on the sampling rate, r , the walking multiplier ψ , and the value-of-time c . We estimate a street-specific sampling rate using the spatial density of parking spots combined with assumptions on driving speed. The walking multiplier ψ depends on the ratio θ between the speed of driving (while searching) and the speed of walking, where $\theta \geq 1$. We follow Inci (2015) and assume that a driver can circle around the blocks to reduce walking time after reaching the destination. van Ommeren et al. (2021) shows that in this case⁶

$$\psi = (2\theta - 1) \ln \left(\frac{4\theta - 2\theta e^{-0.5vN}}{2\theta - 1} \right). \quad (5)$$

It is easy to see that ψ is increasing in vN . When the vacancy rate is high and the area is large ($vN \rightarrow \infty$), the driver searches on a straight line, i.e. $\psi = (2\theta - 1) \ln \left(\frac{4\theta}{2\theta - 1} \right)$. This is example of a linear search strategies, where drivers search along a street. However, when the vacancy rate or the search area is small ($vN \rightarrow 0$), drivers circle around to reduce walking time, and ψ approaches 1.⁷ The estimate of the value-of-time parameter c is available from literature. Given these information, we can estimate the MECP per street for any moment in time. van Ommeren et al. (2021) provides a detailed description of the calibration procedure of the model parameters.

The estimate of the MECP is important input for parking policy design and necessary for rigorous welfare analyses of parking policy reforms. The optimal parking fee per unit of time is effectively a Pigovian tax which internalizes the external search cost. van Ommeren et al. (2021) shows that the welfare-optimal parking fee is equal to the MECP,

$$p^*(t) = \frac{c\psi}{r} \frac{I(t)A(t)}{Nv(t)^2} = MECP. \quad (6)$$

⁶van Ommeren et al. (2021) assumes that drivers search only within a street and that the destination is at one end of that street. Therefore, a driver who obtains a random spot expects to walk half a street. If destinations are distributed uniformly along the street then the expected walking distance will be a quarter of the street.

⁷Assume e.g. that $\theta = 4$. For $\theta = 4$, ψ remains between 1.0 and 5.8. For example, when $\theta = 4$, street size is 20 and a vacancy rate of 0.1, $\psi = 4.4$.

Moreover, the optimal parking fee, $p^*(t)$, is inversely proportional to the square of the vacancy rate, $v(t)$. Therefore when vacancy is low, the marginal increase in parking search time caused by longer parking duration increases steeply. Consequently, parking fee should be higher when occupancy rate is high. It is also proportional to the number of searching drivers, $I(t)A(t)$. Therefore, when there is a low number of searching drivers (such as overnight or on Sunday), the parking fee should be lower.

Finally, it can be shown that, when the parking fee is optimal, the marginal benefit of a parking spot due to the reduction in the search cost must be equal to the marginal capital costs, see van Ommeren et al. (2021). This result can be used to evaluate parking supply. For example, if $p(t) > MECP(t)$, then the marginal benefit of a parking spot exceeds the marginal capital costs of on-street parking. This implies that the optimal parking supply is less than the observed parking supply. More importantly, using optimal time-varying parking fees, the total parking revenue is equal to the total cost of parking supply, including the cost of land.⁸

3 Data and Descriptive Statistics

We use data from the Municipality of Copenhagen to estimate the MECP. We have information for on-street parking on weekdays between 09:00 and 18:00 for the year 2010.⁹ We focus on 22 streets in the red parking zone closest to the historical city center that have no residential parking and at least 10 parking spots per street.¹⁰ There has been a zonal system, with three zones, since 2007 (Gragera et al., 2021).¹¹ The red zone is the city center with few residents and many shops, restaurants, and offices. On-street parking fees are close to the off-street parking fees in Copenhagen.¹² We also know the parking supply for each street. We observe the beginnings and ends of each individual *paid* parking duration (in total, we have 1,179,628 micro observations).

We have information about the occupancy rate and the number of arrivals for 1-minutes intervals. We aggregate the data and calculate the occupancy rate (the ratio of the number of arrivals to the number of parking spots) per interval of 30 minutes for each street. On average, the occupancy rate is 0.55 and the arrival rate in a street is 0.18 arrivals per parking spot per hour. Figure 1 provides information about the distribution of the occupancy rate. It shows that there is a lot of variation in occupancy levels. Occupancy is generally (for 83 percent of the observations) lower than 90 percent.

⁸This result holds only when assuming constant returns to scale in the supply of parking.

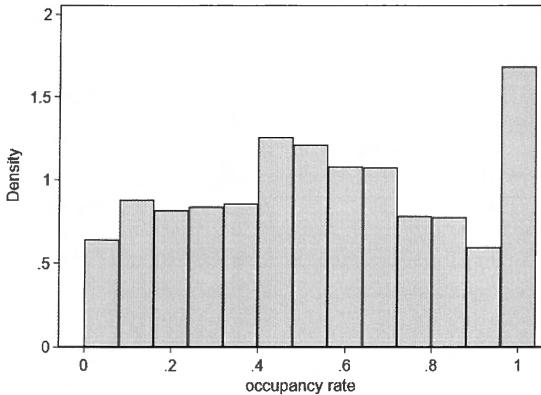
⁹About two-thirds of the parking spots in the Municipality of Copenhagen are on the street (Københavns Kommune, 2012).

¹⁰The considered streets are: Admiralgade, Christian IX's gade, Dyrkøb, Gammel mønt, Hovedvagtsgade, Kattesundet, Kgs. Nytorv, Laksegade, Landemærket, Lavandelstræde, Løngangstræde, Møntergade, Nikolajgade, Ny østergade, Nyborgade, Peder Hvitfeldts str., Rosenborggade, Skt. Peders stræde, Valkendorfsgade, Vestergade, Vognmagergade, and Åbenrå.

¹¹The zonal structure has been in use for years, with some adaptions, since its first introduction in the 1990s.

¹²The parking fee for the red zone is almost three times as high as for the blue zone (Gragera et al., 2021). Outside the three zones (the outer city) there are generally no fees for parking.

Figure 1: The distribution of the occupancy rate



The sampling rate, r , is derived from an assumed search driving speed of 20 km/h combined with information about the spatial density of parking spots within streets. The average sampling rate is about 0.75 parking spots per second. Finally, we allow for circular in-vehicle search and assume that the ratio of driving to walking speed is equal to 4. The walking time multiplier ψ has values between 1 and 5.8, with higher values when occupancy and search time are lower. For the value of cruising time, we use DKK 138.¹³

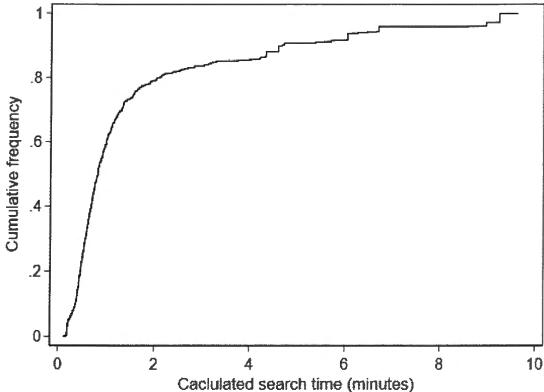
One difficulty here is that the paid parking duration typically exceeds the actual parking duration because on-street parking requires drivers to pay immediately after they park, and drivers tend to pay more than the actual duration to prevent fines. As a consequence, the occupancy rate that we calculate systematically *underestimates* the actual vacancy rate, and therefore we can only provide upper-bound estimates of the MECP. This turns out not to be problematic for the interpretation, because the calculated occupancy levels for the large majority of observations are low (the average occupancy rate is 0.55).

4 Main Results

We now discuss our main empirical results. Our upper bound estimate of the mean search time $Z(t)$, which is the sum of cruising and walking time, is 82 seconds. Figure 2 shows the cumulative distribution of the search time. It reflects that search time is short for a large majority of drivers: 60 percent of drivers search less than 1 minute. Very few drivers search for more than 5 minutes. This result is consistent with our descriptive finding that the occupancy rate is below 0.90 for most of the time.

¹³Transport economists typically find that the travellers value of time is about 50% of the gross wage rate (Small and Verhoef, 2007).

Figure 2: The cumulative distribution of search time



When we focus on occupancy rates between 0.90 and 0.95, the average search time is estimated to be about 3 minutes. Moreover, 25 percent of drivers search for more than 4 minutes, and only 1 percent more than 6 minutes.

Using these upper bounds, we find that the MECP is rather low for 16 out of 22 streets. The exceptions are Dyrkøb, Kattesundet, Lavendelstræde, Ny Østergade, Nybrogade, and Valkendorfsgade. We are particularly interested in whether the uninternalized MECP (i.e., the MECP minus the marginal parking fee) is positive or negative. This is basically a measure of to what extent the MECP has been internalized by the parking fee. In order to facilitate comparisons with hourly parking fees, we provide the estimates of the MECP for an hour of parking. In 82.2 percent of observations, the uninternalized MECP is less than DKK 29, so the parking fee is (approximately) equal to the MECP. In the remaining 17.8 percent of the sample, the uninternalized MECP exceeds DKK 29 such that the MECP exceeds the parking fee. Figure 3 shows the results as a function of the occupancy rate.¹⁴

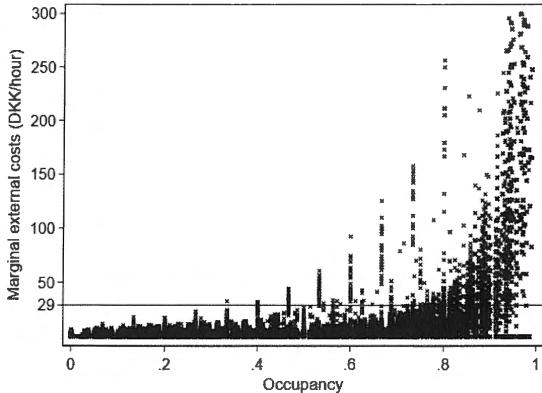
Summarizing, for the majority of streets in Copenhagen, cruising time is very limited. Our estimation results suggest that small reductions in parking fees might be welfare improving.

5 Conclusion

The optimal parking pricing policy is to set parking fees equal to the costs of cruising such that cruising and capacity underutilization are prevented. Practitioners therefore need to know the level of cruising for parking when designing parking policies. We apply the theoretical framework developed in van Ommeren et al. (2021) to Copenhagen and using administrative

¹⁴Using our data and the empirical strategy it is also possible to analyse the temporal-spatial distribution of the MECP, e.g. variation across hours of the day and days of the week.

Figure 3: The MECP as a function of occupancy rate



parking data estimate the whole distribution of the marginal external cruising time over time and space (streets). We find that the upper bound estimate of the MECP in Copenhagen is low for the majority of considered streets and often below the actual parking fee. This result is useful for the city parking authority to evaluate the optimality of the current parking fees and parking supply from a welfare perspective. Our empirical findings suggest that small reductions in parking fees might be welfare improving. More importantly, if the city authority introduces an optimal parking policy, then the cost of the parking supply is self-financing.

The ability to recover accurate real-time occupancy rates, e.g. from parking apps that allow drivers to pay via smartphones or from occupancy sensors, combined with the presented framework, can help parking authorities to implement optimal dynamic parking fees, where prices vary along different times of the day and across parking areas, and in this way to design optimal parking policy.

References

- Arnott, R., Inci, E., and Rowse, J. (2015). Downtown curbside parking capacity. *Journal of Urban Economics*, 86(C):83–97.
- Arnott, R. and Williams, P. (2017). Cruising for parking around a circle. *Transportation Research Part B*, 104:357–375.
- Gragera, A., Hybel, J., Madsen, E., and Mulalic, I. (2021). A model for estimation of the demand for on-street parking. *Economics of Transportation*, 28:100231.
- Inci, E. (2015). A review of the economics of parking. *Economics of Transportation*, 4:50–63.

- Inci, E., van Ommeren, J., and Kobus, M. (2017). The external cruising costs of parking. *Journal of Economic Geography*, 47:333–355.
- Københavns Kommune, . (2012). Københavns kommunens parkeringsinformation.
- Marsden, G. (2006). The evidence base for parking policies—a review. *Transport policy*, 13(6):447–457.
- Shoup, D. (2006). Cruising for parking. *Transport Policy*, 13:479–486.
- Small, K. A. and Verhoef, E. T. (2007). *The Economics of Urban Transportation*. Routledge, London.
- van Ommeren, J., McIvor, M., Mulalic, I., and Inci, E. (2021). A novel methodology to estimate cruising for parking and related external costs. *Transportation Research Part B: Methodological*, 145:247–269.
- Verhoef, E., Nijkamp, P., and Rietveld, P. (1995). The economics of regulatory parking policies: the (im)possibilities of parking policies in traffic regulation. *Transportation Research Part A*, 29:141–156.
- Vickrey, W. (1954). The economizing of curb parking space. *Traffic Engineering*, 25:62–67.
- Zakharenko, R. (2016). Time dimension of parking economics. *Transportation Research Part B: Methodological*, 91:211–228.

Who is the Favorite Child?

A Preliminary Descriptive Analysis

Sara Armandi

1 Introduction

“I couldn’t pick a favorite”. Most parents with more than one child would express this. Of course, at times, one kid is more lovable (or less annoying) than another, but luckily this is constantly changing. Additionally, children often have unique likable (and unlikable) features. Despite differences, it still might seem impossible to choose a favorite. However, in a world where data has the answer to almost anything, why should it not be possible to pick a favorite child, if not by heart, then by numbers and statistics? Hopefully, data will help answer the difficult question: “Who is the favorite child?”.

In this article, data from three different rounds are investigated, each round representing the behaviour of one child. In total, 5,125 records from three children are divided on three different behaviour types; feeding, sleeping and diapering. Using the concept of a parental utility function, descriptive statistics on the three children are compared, and ultimately, a favorite child is found.

2 Theory

It might not be obvious how one would choose one child over another. There are a lot of factors influencing this choice, some of them very straight forward, but others quite hard to measure. In this article, a child’s behaviour is expected to influence parent utility. The child who brings most utility is assumed to be the favorite.

2.1 Parental Utility from Child Behaviour

A child’s behavior affects parents well-being and inevitably, it affects to what extent parents like their child. Luckily, the dislike of a child only lasts in shorter periods of time and is often outweighed by the indescribable love parents have for their child. An attempt to quantify how much a parent likes a child, is carried out by creating a parental utility function as shown in Equation 1.

$$U = U(Behavior_{child}, Mood_{parent}) \quad (1)$$

The $Behavior_{child}$ parameter contains everything from the number of times the child pees or poos in its diaper, the child's sleeping pattern, the mood of the child including how much the child cries or smiles to its parents. Unfortunately, it's very difficult to quantify mood. It might be doable, by timing and noting each time a child cries or smiles, but these statistics are not available in the data material. Even more difficult to measure is the mood or general state of mind of the parent, $Mood_{parent}$. If a parent has sufficient mental resources, even a delinquent child might, contrary to all expectations, affect the parental utility positively. Of course, the opposite is the case when a parent has a low amount or no mental resources.

To simplify the utility function from Equation 1, the measures which are difficult to quantify are omitted, and the records which are available in data (feeding, sleeping, and diapering) are utilized. Hence, Equation 1 can be written as Equation 2.

$$\begin{aligned} \max_{F,S,D} U &= U(Feeding, Sleep, Diaper) \\ &= w_S \cdot Sleep - w_F \cdot (Feeding - T)^2 - w_D \cdot Diaper \end{aligned} \quad (2)$$

The parameter *Sleep* contains the child's sleeping patterns, including the total hours of sleep per day and the length of the naps. It is indeed important that a child sleeps a lot, but what might be even more important is the length of the naps. Especially at night these naps shouldn't be too short. The term is affecting the utility positively, as more sleep and longer naps for a child often entails the same for the parents, which is absolutely preferred.

In the second term of the utility function in Equation 2, *Feeding* represent the time between each feeding and T is a time constant which represent the ideal length between two feedings. When T is subtracted from the time between two feedings and both are squared and negative, the maximum utility from this term would be reached if the time between all feedings were equal to T . It is proposed, that an equal spacing between feedings is important, not only to be able to plan, but also because of the well-being of the mother's bosom. If the time between feedings is a lot longer than it usually is, it might feel like something is about to explode. On the contrary, it might feel like there is not enough food for the child if the time is too short compared to the normal [Armandi, 2022].

The parameter, *Diaper*, is the number of diapers changed each day with dirty diapers weighing a bit higher than wet diapers. It is included negatively in Equation 2 for obvious reasons. Finally, w_i are weights, as all terms in the function doesn't affect the utility equally. The size of the weights might differ depending on the mood of the parent, but the relation is very likely to be: $w_S > w_D \geq w_F$.

When maximizing Equation 2, the parent would have the perfect child which behave optimally. In this article, the focus is on comparing three children to find the favorite. A parental utility function for each child is investigated and the three are compared as shown in Equation 3.

$$U_{Child\ 1} \gtrless U_{Child\ 2} \gtrless U_{Child\ 3} \quad (3)$$

The child, generating the function with the highest parental utility, will be the favorite child.

3 Data

Data is collected in three rounds, each representing the behavior of a unique child. All three children included in data are carried and raised by the same mother (and the father has also taken care of the children). Additionally, all children are born as boys to exclude any possible gender effect. When data from all rounds are cleaned as well as restricted, they are combined to construct a final dataset, which is used to compare the behaviour of the three children.

3.1 Child 1 Data

The first data round was collected from August 2018 using a “Notes” app. To get more info on this data collection, refer to the article “Baby Patterns - A Mother’s Guide to Understanding Her Baby” from 2019. Only four months of daily records have so far been converted from unstructured “Notes” to structured data in a SAS dataset. More data is available and will be utilized in future articles [Armandi, 2019].

Table 1: Original Data by Child before Cleaning and Reduction

	Child 1	Child 2	Child 3
Total Number of Obs	2,798	7,881	3,497
First Record	30AUG2018:23:44	08JUN2020:23:10	25APR2022:02:40
Last Record	31DEC2018:22:00	11JUL2021:16:45	01AUG2022:06:21
Child Age Range (days)	22 - 145	12 - 410	1 - 99

Source: Calculations on original data.

Note: Each record is represented by one observation in the dataset.

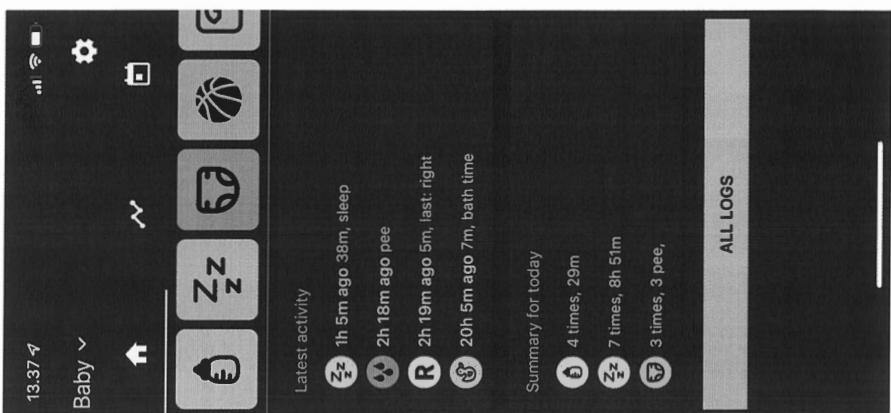
More records for Child 1 are available in archives. Data for Child 3 is still in collection process.

As shown in Table 1, the original data from child 1 contains 2,798 observations, which in this case correspond to a record noted in the “Notes” app. Records include feedings, baths, and diaper changes. Only from November 15th, 2018, have the child’s sleeping behavior been added to the data [Armandi, 2019]. This implies, that sleeping behaviour of the first child is not available in data used in this article. Finally, the age in days of the first record is 22 days, which is caused by the data collection process starting a lot later in the child’s life compared to the two subsequent rounds.

3.2 Child 2 & 3 Data

Data from the second and third child, was collected using the “Baby Tracker!” app. A screenshot of the app is displayed in Figure 1. In this app, it is possible to – as the name states – track a baby’s behavior. Besides tracking feedings (the main focus in round one), sleep and diapering, the app allows a wide range of other measures to be recorded, like tummy- and play time, weight, height, health and photos, just to mention some. However, feeding, sleep and diapering are the only records which have been collected consistently in round two and three, and thus, what are utilized in this article.

Figure 1: Data Collection App – “Baby Tracker!”



Source: Screenshot from “Baby Tracker!” app on mother’s mobile phone.

Data is easily extracted from the “Baby Tracker!” app using an export function, which allows the user to save the data in a spreadsheet. This spreadsheet is saved with the .abt file extension, an Annotated Braille Text File, created by Computer Application Specialties Company (CASC) Braille 2000, a braille text editor [FileInfo.com, 2014]. Knowing, that it

is possible to trick a computer, and that the .abt file is a kind of compressed file, the file association is changed to .zip, which makes it possible to unzip the file. The compressed file contains a .json file, which is easily imported to SAS®, using a JSON libname statement:

```
libname child JSON "...\\round.json";
```

Data from child 2 is currently the one with most records collected, as the time period extends the longest, from June 8th, 2020, to July 11th, 2021. As seen in Table 1 in Section 3.1, this corresponds to the child being 12 days to 410 days (approx. 1 year and 1,5 months) old and 7,881 unique records.

The third child is still quite young, and hence, the data collection is still in progress. The first record in data is from April 25th, 2022, when the child was only 1 day of age (Actually, the child was only 4 hours and 38 minutes old when the data collection started, as he was born 22:02 on April 24th). To write this article, data was extracted from the “Baby Tracker!” app on August 1st, 2022, which is also the latest record with the child’s age being 99 days.

3.3 Cleaning and Reduction

It is worth mentioning, that data might not be 100% reliable. Even though the data collector (the mother of the children) does everything in her power to take care of the children, it hasn’t been possible to guard them 24/7. Hence, some of the tracked points might not be precise and suffer from a mild degree of human error – especially when the father has been involved. Besides the “father effect” data might also be a bit more insecure when collected in the “late hours”. Even though the data collector might suffer from insomnia, mumnesia and other parental disorders, the accuracy of the data is quite precise, due to the persistence of the data collector.

When comparing data from the three children, data from the first child suffers from most human error, as the tracking system wasn’t optimal. This is handled by deleting some incomprehensible records, cf. Armandi, 2019. In the final data, only records that are of types feeding, sleeping, and diapering are included.

The time between two feedings and the time between two naps must be addressed as well. Some feedings of the child occur with only a few minutes apart and are therefore considered as a single feeding. This could be the case if the child feeds from both breasts, and there is a small pause in the middle when moving the child from one side to the other. In data, this would look like two different meals, which wouldn’t be right. To account for this, all feedings with less than 10 minutes to the subsequent, have been deleted. Note that this differs from the article from 2019, where the limit was 30 minutes. The reduction in time

is due to the age of the children in the data. As they are quite young, they often eat more frequent due to the size of their stomach. After deleting feedings, a new time between feeding records is calculated and utilized.

Likewise, it might be recorded in data, that the child is awake for only a couple of minutes, before continuing a nap. It is commonly known that babies have difficulties transitioning from one sleep cycle to the next. Often, they might need comfort before transitioning to the next cycle and continue their sleep. It might make sense to combine some of these naps, to account for these transition phases, where the child is only awake were shortly. In this data, however, if the naps are kept as individual records, as the duration of the naps are also of interest.

The final restriction that has been applied to data is the age of the children. To be able to make reasonable comparisons, the children should have the same age. In this case, 1 data from the first child sets the lower limit, as data collection wasn't systematical the first three weeks of the child's life. In the same way, data from the third child sets the upper limit, as this child is still quite young. The age in days for the children in this dataset are restricted to 24 to 98 days, both days included. This results in records for 75 days for each round.

In Table 2: , the frequency and percent of all records are displayed for each of the three children divided on the three different record types. In total, the final dataset contains 5,125 records.

Table 2: Frequency and Percent of Records by Child and Type in the Final Dataset

Type	Child 1		Child 2		Child 3	
	Frequency	Percent	Frequency	Percent	Frequency	Percent
DIAPERING	439	33.13	413	23.39	368	18.09
FEEDING	886	66.87	685	38.79	810	39.82
SLEEPING			668	37.83	856	42.08
TOTAL	1,325		1,766		2,034	

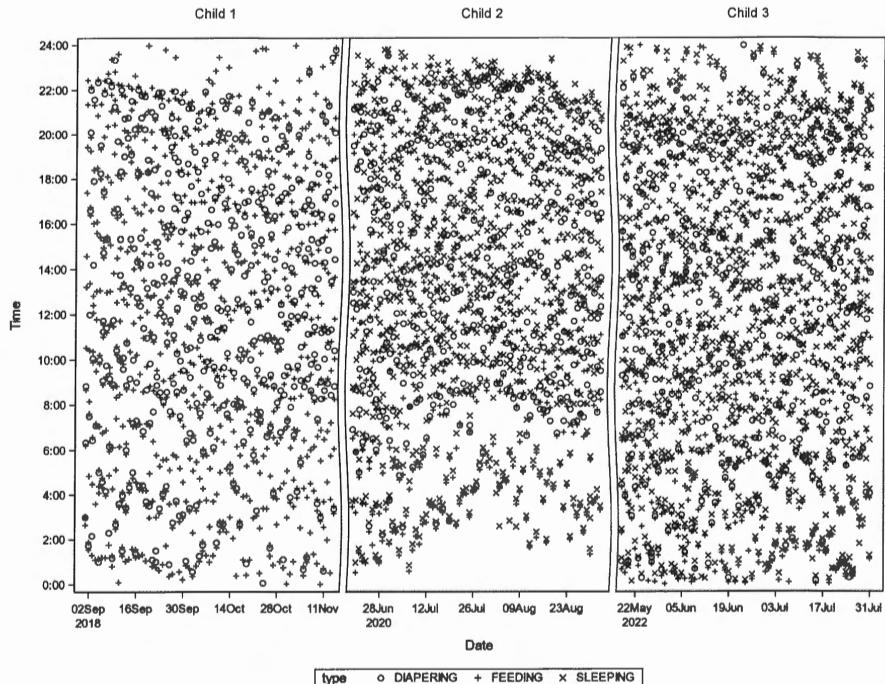
Source: Output from PROC FREQ, SAS® 9.4.

4 Results

An immediate visual analysis on data from the three is conducted and shown in Figure 2. The scatter plot displays records from each of the children at an age between 24 and 98 days. All 5,125 records are plotted with the date of the record on the x-axis and the time on

the y-axis. The records are divided into three different types, diapering \circ , feedings + and sleep \times . To elaborate, the circles are the starting point of a diaper change, plusses are when feedings are initiated, and crosses indicate when a child falls to sleep.

Figure 2: Date and Time for Records by Child



Source: Output from PROC SGPLOT, SAS® 9.4.

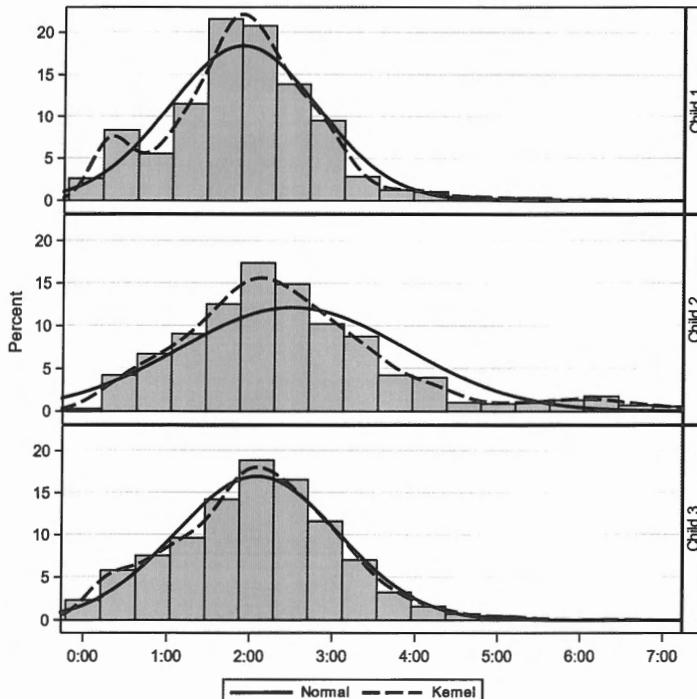
Note: The x-axis has been broken to fit records from all three children in one figure.

At a first glance, data from the second and third child seem denser than for the first. This is due to the lack of sleeping data for the first child. Further, there seems to be less activity in the nighttime hours for all children. This is especially evident from June 20th, 2020, to September 3rd, 2020, which is the time periods for data from the second child. Besides this, it's quite difficult to conclude anything else than, that there is a lot going on.

4.1 Feeding

To get insight into the eating behaviour of the children, a distribution of the time between feedings for data from all three children are shown in Figure 3.

Figure 3: Distribution of Time Between Feedings by Child



Source: Output from PROC SGANEL with HISTOGRAM statement, SAS® 9.4.

There are no immediate signs of huge differences between the three children. All are centered around a time spacing of two hours between feedings. However, it seems as if there are a few more instances for the first child, where the time between feeding is short. On the other hand, data from the second child is quite right skewed, with a lot of long hours between two feedings.

To get a better picture of what is going on in Figure 3, descriptive statistics are calculated and displayed in

Table 3. As seen, from the number of observations, child 2 eats less frequent than the two other's which is given by the number of observations in the data. The minimum lengths between feedings are all very close to 10 minutes in all three cases. This makes sense, as, described in Section 3.3., feedings with less than 10 minutes to the subsequent, have been deleted. When looking at the maximum time between feedings, once again, child 2 is the one of interest. This value is two hours longer than the maximum of both child 1 and 2. From this, and what is seen in Figure 3, it would be fair to conclude that child 2 is able to last longer without food than his siblings.

Table 3: Descriptive Statistics of Time Between Feedings (*H:MM:SS*)

Child	N Obs	Min	Max	Mean	Median	Std Dev	Range
1	886	0:10:00	6:11:00	1:53:09	1:53:00	0:54:03	6:01:00
2	685	0:11:00	8:22:28	2:30:21	2:16:15	1:22:06	8:11:28
3	810	0:10:01	6:24:54	2:05:20	2:06:01	0:58:49	6:14:53
All	2381	0:10:00	8:22:28	2:08:00	2:03:14	1:06:33	8:12:28

Source: Output from PROC MEANS, SAS® 9.4.

The mean and median in

Table 3 supports what we concluded from Figure 3, that data from child 2 is right skewed. This is due to the mean being higher than the median. Neither of the data from child 2 or three seems to be skewed and are both close to 2 hours, with child 1 feeding a bit more frequent than child 3.

What is easily seen from the numbers in

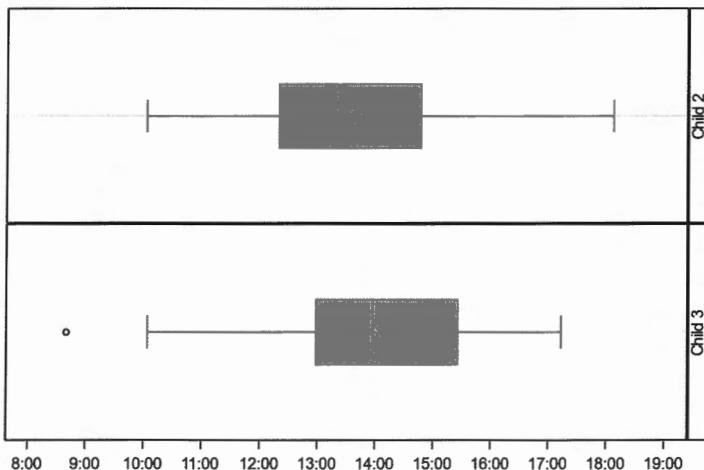
Table 3 as well as the histograms in Figure 3, the time between feedings is for none of children constant as wished for. To conclude something on whether the time between two feedings is “more constant” on child than another, the measures of dispersion are examined. Both the standard deviation and the range are significantly lower for child 1 and 3 than is the case for data from child 2. This suggests that child 2 is farthest from a constant time between feedings.

4.2 Sleeping

Sleeping patterns are only available for child 2 and 3 in this dataset. This could turn out to be a favor to child 1 if he wants to stand a chance to be the favorite child. When child 1 was a baby, he was often referred to as “Max30”, as he was notorious for his short naps, resulting in a mother which was not very well-functioning due to lack of sleep (Whether the mother's condition was due to extreme bad sleeping patterns from child 1, or a result of being a first-

time mother is unknown. This feeling of child 1 being terrible at sleeping might not have been the case, had he not been the first-born. With second- and third borns, mothers are used to not getting any sleep). The sleeping pattern of child 1 will be included in later work, as this is available in data from November 15th, 2018.

Figure 4: Box Plot with Daily Amount of Sleep by Child



Source: Output from PROC SGMPANEL with HBOX statement, SAS® 9.4.

The aggregated daily amount of sleep for child 2 and three is displayed in a box plot in Figure 4. Most days, child 2 sleeps between 12,5 and 15 hours, whereas child 3 sleeps between 13 and 15,5 hours a day. However, as seen in the figure and in Table 4, child 2 has the highest number of hours slept on one day, and child 3 is the owner of the record for the least hours of sleep on one day; 8 hours and 40 minutes, luckily, an outlier. Both the mean and the median are highest for child 3. Additionally, the distribution of child 2 is a bit right skewed, which is also clear from the distribution shown in the histogram in Figure A1 in Appendix.

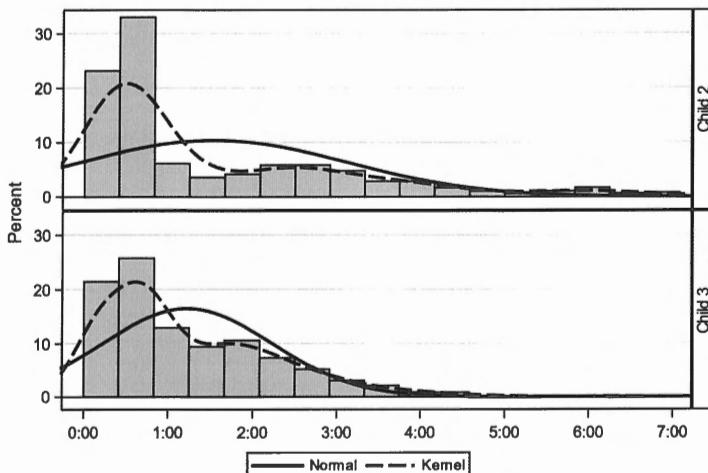
Table 4: Descriptive Statistics of Daily Amount of Sleep (*HH:MM*)

Child	N Obs	Min	Max	Mean	Median	Std Dev	Range
2	75	10:03	18:08	13:38	13:22	1:43	8:05
3	75	8:40	17:13	14:00	13:58	1:43	8:33
All	150	8:40	18:08	13:49	13:43	1:43	9:28

Source: Output from PROC MEANS, SAS® 9.4.

It is of course important for a child to sleep a lot, especially in the first months of life. What is also important, is to have long stretches of sleep. Figure 5 shows how the lengths of all naps are distributed. It is very clear, that most naps are below one hour for both children. Especially child 2 has a high percentage of short naps. At the same time, child 2 also has some of the longest naps. On the contrary, child 3 don't seem to have many naps longer than four hours but doesn't have as high percentage of short naps below one hour as child 2. According to Table A1 in Appendix, the average nap length is 1:32 for child 2 and 1:14 for child 3. Additionally, the difference between the longest lengths for the two children is more than two hours in favor of child 2.

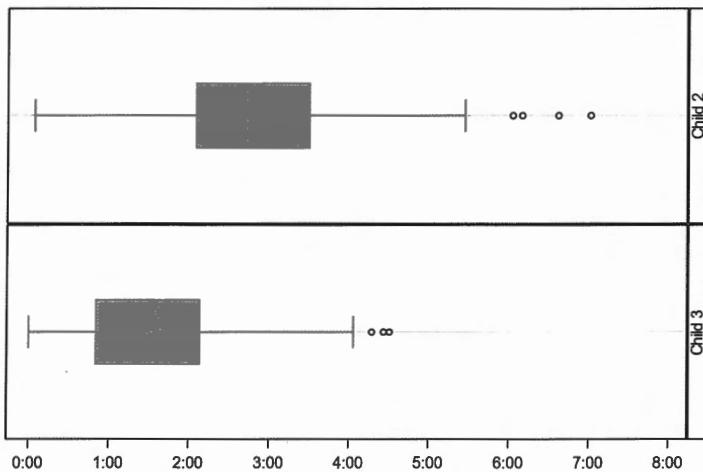
Figure 5: Distribution of Length of Naps



Source: Output from PROC SGMPANEL with HISTOGRAM statement, SAS® 9.4.

From a parental point of view, the ability to sleep long is especially important in the nighttime. In Figure 6, the distribution of the length of naps in the nighttime are displayed. It is easy to see, that child 2 takes longer naps during night than child 3. Close to 50 pct. of all naps are between two and three hours for child 3 with only three naps lasting longer than four hours. For child 2, the naps during nighttime lie between two and four hours with a few naps lasting longer than six hours. The distributions of the two children's nighttime nap length are shown as histograms in Figure A2 in Appendix with Table A2 showing the descriptive statistics. From these, it's easy to see, child 2 sleeps better at night than child 3, which is also clear from Figure 6.

Figure 6: Box Plot with Length of Naps in Nighttime



Source: Output from PROC SGANEL with HBOX statement, SAS® 9.4.

Notes: Nighttime is when a nap starts between 23:00 and 8:00.

4.3 Diapering

The final measure of interest for the parental utility is the number of diaper changes. The less diapers that need to be changed, the happier the parent. As seen from Table 5, the minimum and maximum are quite close. Note, that the minimum number of changed diapers would be higher than 2 and 3, if data did not suffer from the “father effect”. From the mean, it is seen that the average daily number of diaper changes decreases when moving from child 1 to 2 and again from child 2 to 3. This might indicate that children 2 and 3 does not pee and poo as much as child 1, or it could indicate that the parents get lazier as they get more children. The distribution of the diaper changes is shown in a histogram for each child in Figure A3 in Appendix.

Table 5: Descriptive Statistics of Daily Number of Diapers Changed

Child	N Obs	Min	Max	Mean	Median	Std Dev	Range
1	75	2	8	5.85	6	1.10	6
2	75	3	9	5.51	5	1.03	6
3	75	3	9	4.91	5	1.13	6
All	225	2	9	5.42	5	1.15	7

Source: Output from PROC MEANS, SAS® 9.4.

5 Conclusion

“Who is your favorite child?”. A lot of different factors seem to affect this question. Of course, the children’s behaviour is important, but also the mood of the parent who must answer this question, influences the choice. To make everything even more complicated, parental wisdom plays a role, as the knowledge a first-time parent has compared to a second- and third-time parent differs. This is of course related to the fact that a child’s behaviour is affected by the parental behaviour.

In this article, data from three different children were investigated to try to answer the question at state. For each child, the amount of sleep, number of feedings and diapers that need to be changed, are examined, as these measures are expected to affect how much utility, a parent would get from a child. Using the results from these basic descriptive statistics, and the formulation of the utility from Equation 2 in Section 2.1, a favorite child can be declared.

Initially, the results show that child 2 has the longest average time between feedings. But at the same time, child 2 is also farthest from a constant time between feedings. Of the three, child 1 might be slightly closer to a constant time between two feedings than the others. Looking at sleeping patterns, child 3 gets most sleep on average during the entire day, but when investigating the lengths of naps, child 2 outperform child 3 in nighttime. Finally, when considering results from diaper change, child 3 would affect the parental utility more positive than the others due to a lower average number.

From these results it’s difficult to choose a favorite, especially considering that the parents affect these measures a lot. As an example, if the parent does not feel like changing a diaper (s)he can just choose not to. Also, the results of the three different behaviour types affect each other. This is clear when seeing that the time between feedings is farthest from a constant level for child 2, which is mainly caused by child 2 taking longer naps.

What needs to be considered to make a final conclusion are the weights, w_i , from Equation 2. Asking any parent, you would always get the answer, that the baby’s sleeping pattern affects the parents most (implying that w_s is larger than the other weights). Because, as everyone knows, a good night’s sleep is the basis for everything else.

According to the results of this article, child 2 is the favorite child.

References

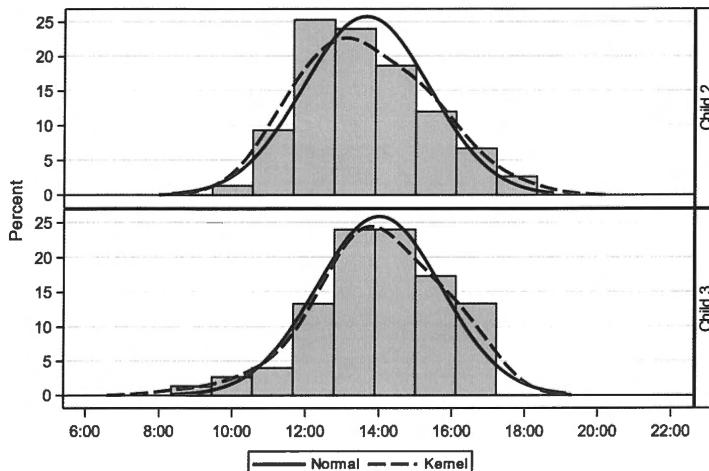
Daniel Armandi (2022). *Discussions with the Uncle of the Three Children*, August 2022

Sara Armandi (2019). *Baby Patterns – A Mother’s Guide to Understanding Her Baby*, Symposium i Anvendt Statistik 2019, red. Peter Linde, Det Nationale Forskningscenter for Arbejdsmiljø

FileInfo.com (2014). *ABT File Extension - What is an .abt file and how do I open it?*. Retrieved 7 August 2022, from <https://fileinfo.com/extension/abt>

Appendix

Figure A1: Distribution of Daily Amount of Sleep by Child



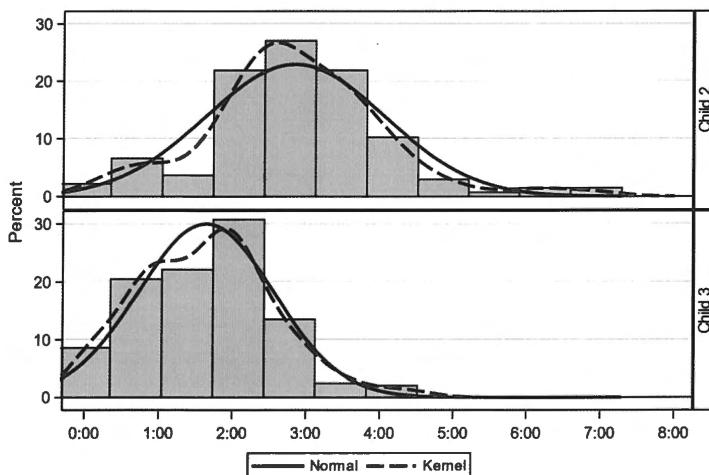
Source: Output from PROC SGMPANEL, SAS® 9.4.

Table A1: Descriptive Statistics of Length of Nap (*HH:MM*)

Child	N Obs	Min	Max	Mean	Std Dev
2	668	0:01	8:19	1:32	1:36
3	856	0:00	6:15	1:14	1:00
All	1524	0:00	8:19	1:22	1:18

Source: Output from PROC MEANS, SAS® 9.4.

Figure A2: Distribution of Length of Naps in the Nighttime



Source: Output from PROC UNIVARIATE, SAS® 9.4.

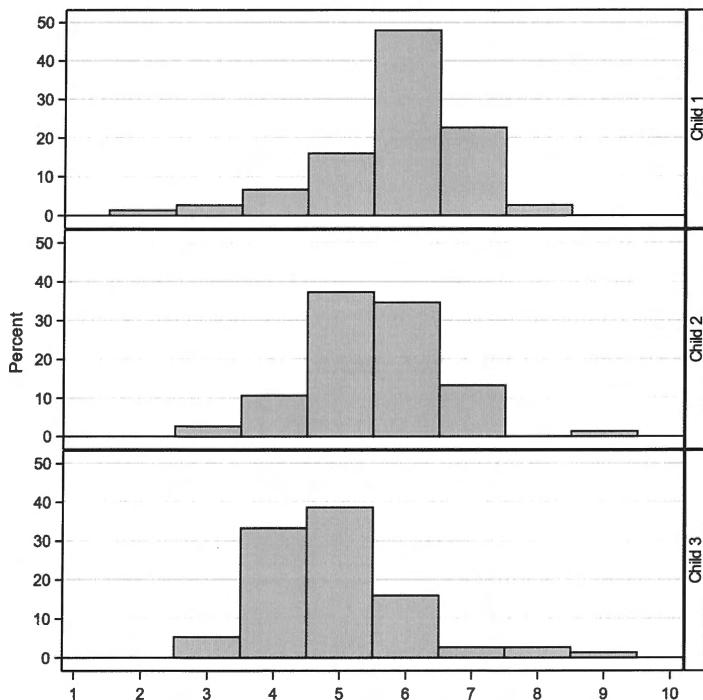
Table A2: Descriptive Statistics of Length of Nap at Nighttime (*HH:MM*)

Child	N Obs	Min	Max	Mean	Std Dev
2	145	0:04	7:01	2:47	1:14
3	274	0:00	4:31	1:34	0:56
All	419	0:00	7:01	1:59	1:12

Source: Output from PROC MEANS, SAS® 9.4.

Notes: Nighttime is when a nap starts between 23:00 and 8:00.

Figure A3: Distribution of Number of Daily Diaper Changes



Source: Output from PROC SGANEL with HISTOGRAM statement, SAS® 9.4.

Høj begavelse i et økonomisk perspektiv

En økonometrisk analyse af højt begavede respondenteres vurdering af deres trivsel i folkeskolen og korrelationen med senere uddannelsesniveau, social- og faglig trivsel samt tidsperspektivet i opfangelsen af den høje begavelse.

Freja Falk

Institut for økonomi, Københavns Universitet

Abstract

In the face of the competitive global economy, several sources have pointed out that Denmark's competitive advantage should be found in innovation. At the same time, organizational tasks have to a large extend changed to be more cognitive demanding. In this context, research has found that highly intelligent individuals in general are more innovative than their less intelligent peers. Therefore, it may be argued that it is crucial for politicians to ensure the wellbeing of this population. However, several studies have shown that highly intelligent individuals do not thrive in elementary school. This paper examines how several well-being variables in elementary school correlates with social- and professional well-being and later levels of education. Descriptive statistic shows a difference in the thrive between the highly intellectual individuals and the average student in elementary school. The multiple logistic regressions analysis of the highest intellectual class shows a significant negative correlation between boredom in teaching and later level of education. Moreover, the analysis shows a significant positive correlation between the evaluated ability to corroborate and the social well-being. Lastly it shows the importance from an economic perspective of discovering the high intellect early for the highly intelligent individuals to gain an increased possibility for completing a longer higher education or above. Many studies supports that a higher level of education generates a higher level of national income. These findings allow this report to suggest initiatives to support the highly intellectual children: 1) Early intelligence tests for children in elementary school, 2) Increased professional stimuli and 3) Relations formation with like-minded individuals. Funds for the support of the highly intellectual children is a political decision at last and this paper has proposed the potential economic gain from supporting this group but does not cover whether the funds (as it is by now) should be allocated different.

Kort økonomisk introduktion

Flere har påpeget, at Danmarks internationale konkurrenceevne i høj grad skal komme fra innovation, og de mest banebrydende innovationer er gennem tiderne blevet opfundet af individer med en intelligens væsentlig over gennemsnittet (Hart 2007) (Varemærkestyrelsen 2021). Teknologier som kunstig intelligens og robotteknologi har allerede påvirket den organisatoriske virkelighed, hvilket har skabt væsentlig forandring i arbejdsopgaverne (Erhvervsministeriet 2021). Flere nationale og internationale studier viser, at arbejdsmarkedet er i en bevægelse fra rutineprægede opgaver med et smalt kompetencefelt til mere komplekse og kognitive opgaver (Autor og Price 2013) (Jørgensen 2007). Denne udvikling medfører således en stigende efterspørgsel på arbejdskraft, der kan navigere i kognitiv kompleksitet, hvilket på mange måder er i lig med den flydende intelligens (IQ). Ydermere påpeger (Murtza, et al. 2020), at højt begavede individer alt andet lige er mere profitable for organisationer (end lavere intelligensgrupper) grundet deres øgede kognitive kompetencer.

Litteraturen peger således på, at højt begavede individer spiller (eller kan spille) end vigtig rolle for den danske konkurrenceevne på både det organisatoriske- og samfundsmæssige niveau. Desværre peger mange studier på, at denne gruppe individer allerede fra folkeskolens begyndelse tenderer til mistrivel, hvilket kan have stor negativ effekt på deres videre færdens i uddannelsessystemet samt deres faglige- og sociale trivsel (Mehlbye, Flarup og Iversen 2015) (Andreasen og Lausten 2019) (Niels Christian Vilman 2022)

Litteraturgennemgang

Dansk forskning af højt begavede individer er mangelfuld ifølge (Mehlbye, Flarup og Iversen 2015), hvorfor det er begrænset med kildemateriale, som behandler trivslen hos de højt begavede individer og deres videre færdens i uddannelsessystemet.

Statistisk set sidder der en højt begavet elev i hver anden skoleklasse med en IQ på 130 eller derover (Mehlbye, Flarup og Iversen 2015), hvilket svarer til to procent af alle folkeskoleelever. Disse elever besidder et højt intellektuelt potentiale, men kan være svære at opfange, da højt begavede børn har tendens til at underpræstere i undervisningen (Nissen, Kyed og Baltzer 2011). Forskning af (Rubenstein, et al. 2012) påpeger, at: "... *underpræstering især skyldes, at de højt begavede elever keder sig, fordi de ikke får tilstrækkelige faglige udfordringer*". Når de højt begavede børn ikke får nok faglige udfordringer, understimuleres deres hjerner (Schødt u.d.), hvorigennem de kan skabe dårlige læringsstrategier (Franck, Interview 2022). Som følge af disse

udfordringer vil der ifølge en nyere rapport fra KORA¹ være et stort mørketal af højt begavede elever, der ikke bliver opfanget i folkeskolen (Mehlbye, Flarup og Iversen 2015). De er ofte vant til at kende svaret i undervisningen og kan pludselig føle sig dumme, hvis de ikke kender svaret, eller skal anstrengte sig for at komme frem til svaret (Franck, Interview 2022). I det videre uddannelsesforløb kan denne tendens medføre, at de højt begavede individer forsøger sig med videregående uddannelser, men ender med at droppe ud, da udfordringerne er steget i sværhedsgrad, og de ikke har ”lært at lære” (Andersen 2020). Dea Franck uddyber: ”*Hvis vi ikke møder dem (red. de højt begavede børn), så lærer vi dem ikke at lære. Hvis de ikke kæmper i det de lærer, så behøves de ikke være bevidste om læringsstrategier, og så får de det svært, når de møder større udfordringer.*” (Franck, Interview 2022). Et amerikansk studie understøtter dette fænomen, hvor en undersøgelse fandt en overrepræsentation af højt begavede elever, der droppede ud af universitetet i forhold til størrelsen på gruppen af højt begavede elever (Mehlbye, Flarup og Iversen 2015).

For at komme understimulering i forkøbet bør de faglige udfordringer tilrettelægges således, at de svarer til den højt begavede elevs ”intellektuelle alder”, da de har tendens til asynkron udvikling (Borch u.d.). Den asynkrone udvikling skelner mellem intellektuel- og emotionel udvikling: ”*Det højt begavede barn løber stærkt intellektuelt, og der er tale om en decideret accelereret indlæring i forhold til gennemsnitlige jævnaldrene. På det emotionelle plan er der dog ikke tale om en accelereret udvikling. Det emotionelle udvikler sig i takt med alderen*” (Franck u.d.). Der sker altså en skævvriddning i forhold til den ”intellektuelle alder” og den reelle ”emotionelle” alder i folkeskolen, hvilket kan skabe social- og faglig mistrivsel for barnet. Den højt begavede elev kan fagligt tendere til at springe til resultatet uden at være i stand til at kunne redegøre for processen eller omvendt være så meget nede i detaljen, at de jævnaldrende ikke kan følge med – begge tendenser kan skabe problemer for samarbejdsevnen (Schødt 2016). Disse mønstre peger på, at den højt begavede gruppe muligvis oplever øgede trivselsudfordringer, hvilket denne rapport vil forsøge at afdække.

Variabler anvendt efter indsamling af data med stikprøveteori som metode

Data er indsamlet gennem et spørgeskema blandt individer, der er medlem af Mensa med en IQ liggende i top 2 procent af den danske befolkning. Undersøgelsen endte

¹ KORA er en forkortelse for: Det nationale institut for kommuners og regioners analyse og forskning.

med at tælle 185 respondenter. Variablerne er, hvis ikke andet angivet, dannet som en likertskala fra 1-5.

Variabler i datasæt til analyse		Type	Reference
Label	Variabel		
EHU	Hvad er din højeste gennemførte uddannelse	Binær	0=lav og 1=høj
Alder	Angiv din alder	Selvangivet	-
Gender	Angiv dit køn	Binær	0=mand og 1=kvinde
Tid	Hvornår fandt du ud af, at du var højt begavet?	Binær	0=før og 1=efter
Kedlig	"Undervisningen i folkeskolen var kedelig"	Likert-skala	1 (lav) til 5 (høj)
Faglig	"Jeg klarede mig fulgt godt i folkeskolen"	Likert-skala	1 (lav) til 5 (høj)
Lyst	"Undervisningen i folkeskolen gav mig lyst til at lære mere"	Likert-skala	1 (lav) til 5 (høj)
Sam	"Jeg var god til at samarbejde med andre i folkeskolen"	Likert-skala	1 (lav) til 5 (høj)
Solve	"Jeg kunne i folkeskolen finde en løsning på problemer så længe, at jeg prøvede hårdt nok"	Likert-skala	1 (lav) til 5 (høj)
STF	Hvordan vil du vurdere din sociale trivsel i folkeskolen?	Binær	0=lav og 1=høj
FTF	Hvordan vil du vurdere din faglige trivsel i folkeskolen?	Binær	0=lav og 1=høj
TTF	Hvordan vil du vurdere din samlede trivsel i folkeskolen?	Binær	0=lav og 1=høj
Geo	Hvor i landet er du vokset op?	Binær	0=forstad og 1=storbyp
MHU	Hvad er/er din mors højeste gennemførte uddannelse	Likert-skala	1 (lav) til 5 (høj)
FHU	Hvad er/er din fars højeste gennemførte uddannelse	Likert-skala	1 (lav) til 5 (høj)

Kilde: Datasæt fra opgavens spørgeskemaundersøgelse

Anm.: Variabler er farvekodet efter forklaringsformål: gul=personlige forhold, grøn=trivselsmåling, blå=trivselsoverblik og lilla=socioøkonomisk. Respondentens personlige IQ er fjernet fra datasættet (jf. afsnit: rensning af data).

Analyseværktøj

Variablerne *EHU*, *STF* og *FTF* blev omdannet fra Likert-skalaen til binære variable, der antager verdien høj (1) eller lav (0). Variablen *EHU* angiver ”egen højeste gennemførte uddannelse”, hvor scoren ”høj” er klassificeret efter de individer, der har angivet enten 4 eller 5 som svar, og de individer, der har responderet mellem 1 og 3 klassificeres som ”lav”. Dette skel betyder, at respondenterne i denne rapport skal have en længerevarende uddannelse eller derover for at ende i gruppen med ”høj” uddannelse, og er altså *ikke* et udtryk for en subjektiv holdning om, hvem der er højt eller lavt uddannet. Omvendt ønskes det at undersøge, hvorvidt højt begavede individer havde lav trivsel i folkeskolen, hvorfor denne skala blev justeret til at klassificere svarene mellem 1 og 2 som ”lav” og svarende mellem 3 og 5 som ”høj”. Det betyder, at de individer, der har angivet at deres trivsel var middel i denne undersøgelse, klassificeres som ”høj” trivsel. Ved at omdanne responsvariablerne til binære variable muliggøres den logistiske regression. Den generelle model anvendt til multipel logistisk regression (Forman 2012) er givet ved:

$$\log(Odds) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad , \text{hvor } odds = \frac{P}{1 - P}$$

Denne metode udleder blandt andet Maximum likelihood estimater, ROC-kurver og effektplots, der viser sammenhængen mellem variable, og er de værktøjer, der vil blive anvendt gennem denne rapport.

Deskriptiv statistik

For at blive klogere på trivslen hos den højt begavede, der kan opleve øgede trivselsudfordringer i folkeskolen, er der i Tabel 2 præsenteret en gennemsnitsbetragtning over udvalgte trivselsfaktorer. Gennemsnitsbetragtningen for den højt begavede gruppe er udledt ved et simpelt aritmetisk gennemsnit, hvor n angiver antal respondenter, k angiver den givne variabel og \bar{X}_k angiver gennemsnittet for k:

$$\bar{X}_k = \frac{1}{n} \sum_{i=1}^n k_i$$

Som sammenligningsgrundlag bruges data fra Undervisningsministeriets nationale trivselsundersøgelse (undervisningsministeriet 2018/2019). Data fra Undervisningsministeriet er trukket fra skoleåret 2018/2019 for at undgå mulige påvirkninger fra Coronaepidemien (Lund 2022).

Tabel 2

Gennemsnitsbetragning af fem trivselsfaktorer mellem den gennemsnitlige folkeskoleelever og den højt begavede gruppe			
Spørsmål fra spørskema	Højt begavede (H)	National (N)	Forskel=H-N
"Undervisningen i folkeskolen var kedelig"	3,91	3,29	0,62
"Jeg klarede mig fagligt godt i folkeskolen"	4,19	3,85	0,34
"Undervisningen i folkeskolen gav mig lyst til at lære mere"	2,70	3,21	-0,51
"Jeg var god til at samarbejde med andre i folkeskolen"	2,95	3,86	-0,91
"Jeg kunne i folkeskolen finde en løsning på problemer så længe, at jeg prøvede hårdt nok"	4,22	3,71	0,51

Kilde: Data hentet gennem undervisningsministeriet for de nationale tal og fra mit eget dataset fra spørskemaundersøgelsen for de højt begavede
Anm.: Der gøres opmærksom på, at spørsmålene fra undervisningsministeriet er formet i nutid, hvor mine respondenter har svaret i datid

Jf. Tabel 2 ses afvigelser omkring \pm et halvt- til et helt point (jf. Likert-skalaen) mellem den gennemsnitlige folkeskoleelever og den gennemsnitlige højt begavede respondent. Den højt begavede gruppe tilkendegav en øget kedsomhed i undervisningen (+0,62), en øget opfattelse af egne faglige evner (+0,34) og en øget problemløsningssans (+0,51) i forhold til den gennemsnitlige folkeskoleelever. Omvendt tilkendegav de en mindre lyst til videre læring på baggrund af folkeskoleundervisningen (-0,51) og en iøjnefaldende lavere evaluering af egen samarbejdsevne (-0,91). Resultater der peger i retning af, at der ligger nogle udfordringer hos den højt begavede gruppe, der adskiller sig fra den gennemsnitlige folkeskoleelever, som kan være relevante at sætte ind overfor.

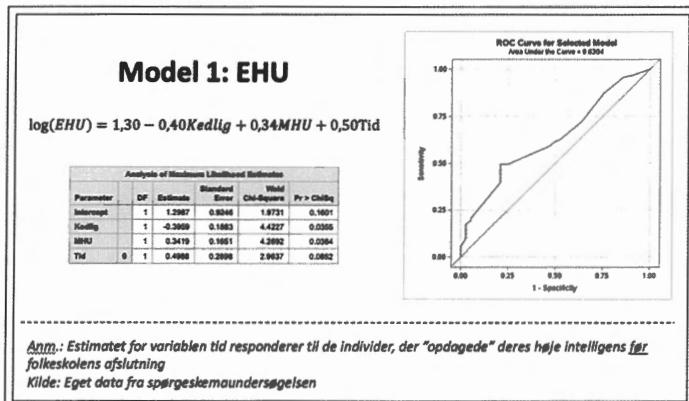
Analyse 1: Faktorers sammenhæng med respondenteres uddannelsesniveau

Den fulde model for faktorers sammenhæng med respondenternes uddannelsesniveau (FM-1) blev opstillet således:

$$\begin{aligned} \log(EHU) = & \alpha + \beta_1 Kedlig + \beta_2 MHU + \beta_3 FHU + \beta_4 Faglig + \beta_5 Lyst + \beta_6 Sam \\ & + \beta_7 Solve + \beta_8 STF + \beta_9 FTF + \beta_{10} TTF + \beta_{11} Gender + \beta_{12} Tid + \beta_{13} Geo \\ & + \beta_{14} Alder \end{aligned}$$

I denne delanalyse blev der korrigteret for alder for at sikre, at respondenten kunne have gennemført en længere videregående uddannelse, og dermed havde mulighed for

at ende i kategorien ”høj” uddannelse. Af den grund blev der indsats en skæringsgrænse på >25 år, hvilket muliggjorde, at alle respondenter burde have lige sandsynlighed for at ende i kategorien ”høj” og ”lav” uddannelse. Ved indførelse af skæringsgrænsen blev 13 respondenter ekskluderet, hvilket gav en populationsstørrelse på n=170. FM-1 blev estimeret i SAS, hvor variablerne *Tid*, *Kedlig* og *MHU* viste sig at være signifikante. Dermed blev EM-1 udledt med tilhørende output:



Jf. *Model 1* havde variablen *Kedlig* en signifikant negativ sammenhæng (-0,40) med respondenternes uddannelsesniveau. Analysen viste således, at desto mere respondenten fandt folkeskoleundervisningen kedelig desto lavere sandsynlighed ville respondenten have for at gennemføre en længere videregående uddannelse eller derover. Variablen *MHU* viste jf. *Model 1* en signifikant positiv sammenhæng (+0,34) mellem moderens højeste gennemførte uddannelse og respondenternes uddannelsesniveau. Resultatet indikerede således en betydning af socioøkonomiske forhold, såsom social arv og miljø. Selvom dette var interessant, vil denne rapport ikke behandle dette område yderligere grundet fokus på trivsel.

Den sidste variabel *Tid* var i første kørsel af FM-1 insignifikant ved $\alpha = 0.05$, og blev derfor udelukket ved *backward selection*-processen. Processen viste, at udelukkelsen skete ved $\alpha = 0.0852$, hvilket indikerede, at variablen muligvis var relevant at inddrage. Ved inddragelse af variablen *Tid* med $\alpha = 0.1$ steg forklaringsgraden af modellen gennem ROC-kurven, hvorfor inddragelsen af variablen til EM-1 blev accepteret. Estimatet for *Tid* viste jf. *Model 1* en positiv sammenhæng (+0.50) med respondenternes uddannelsesniveau ($\alpha = 0.1$). Variablen *Tid* var binær med reference til ”0”, som refererede til de respondenter, der opdagede deres høje begavelse inden folkeskolens afslutning. Jf. *Model 1* viste analyse 1 således, at der var en positiv

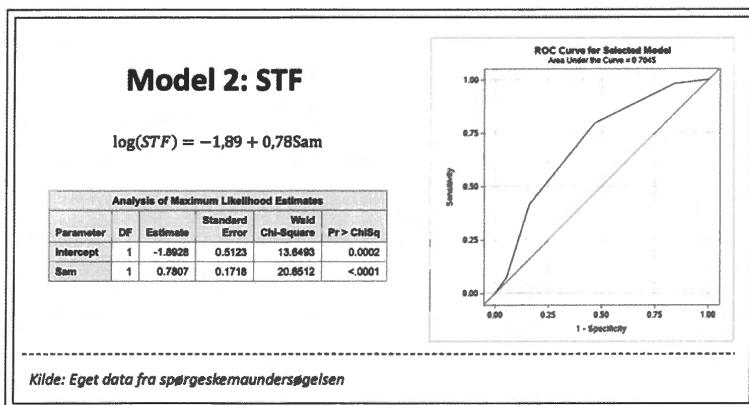
sammenhæng mellem opdagelsen af den høje begavelse inden folkeskolens afslutning og respondenternes uddannelsesniveau. Effektplottet for EM-1 viste, at sandsynligheden for at ende med en længerevarende uddannelse eller derover var for alle niveauer af respondentens evaluering af kedsomhed i undervisningen højere, hvis den høje begavelse blev opdaget inden folkeskolen afslutning. Der gøres opmærksom på, at *kun* 22 ud af 183 individer opdagede deres høje begavelse inden folkeskolens afslutning, hvorfor datagrundlaget for denne gruppe ikke var særlig stort. Men det peger på, at der kunne ligge en udfordring i at opfange de højt begavede børn i folkeskolen.

Analyse 2: Faktorers sammenhæng med respondenternes sociale trivsel

Den fulde model for faktorers sammenhæng med respondenternes sociale trivsel (FM-2) blev opstillet således:

$$\log(STF) = \alpha + \beta_1 Kedlig + \beta_2 MHU + \beta_3 FHU + \beta_4 Faglig + \beta_5 Lyst + \beta_6 Sam + \beta_7 Solve + \beta_8 FTF + \beta_9 TTF + \beta_{10} Gender + \beta_{11} Tid + \beta_{12} Geo + \beta_{13} Alder$$

Variablen *EHU* blev udeladt som forklarende variabel grundet omvendt kausalitet mellem social trivsel i folkeskolen og uddannelsesniveau. FM-2 udviste ”*quasi-complete separation*”, hvilket formentlig skyldtes, at den totale trivsel kunne forklare al variationen i den sociale trivsel. Variablen *TTF* blev derfor fjernet fra FM-2 og problemet løst. Størstedelen af variablerne var yderst insignifikante, hvorfor der ikke var belæg for at medtage andre variable end *Sam*, som var den eneste signifikante variabel i modellen.



Analysen viste en signifikant positiv sammenhæng (+0.78) mellem variablen *Sam* og respondenternes sociale trivsel. Analysen viste således, at desto bedre respondenten evaluerede deres samarbejdsevne i folkeskolen desto større sandsynlighed var der for,

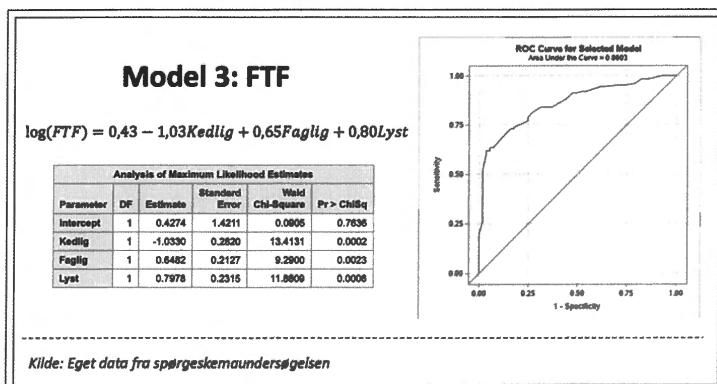
at respondenten ville opleve høj² social trivsel. Trods én enkelt signifikant forklarende variable var forklaringsgraden jf. *ROC-kurven* udemærket, hvilket peger på, at respondenternes evaluering af deres samarbejdsevne i folkeskolen danner en god indikation for deres sociale trivsel i folkeskolen. Effektplottet for EM-2 viste, at evnen til at samarbejde forudsiges at være essentiel for den sociale trivsel blandt den højt begavede gruppe. Hvis en respondent evaluerede egen samarbejdsevnen til laveste (1) niveau, blev der forudsagt en sandsynlighed på 25 procent jf. *Figur 2* for, at individet ville opleve høj social trivsel i folkeskolen. Omvendt er der forudsagt en sandsynlighed for høj social trivsel på omkring 90 procent, hvis respondenten evaluerede egen samarbejdsevne på højeste (5) niveau.

Analyse 3: Faktorers sammenhæng med respondenternes faglige trivsel

Den fulde model for faktorers sammenhæng med respondenternes faglige trivsel (FM-3) blev opstillet således:

$$\log(FTF) = \alpha + \beta_1 Kedlig + \beta_2 MHU + \beta_3 FHU + \beta_4 Faglig + \beta_5 Lyst + \beta_6 Sam + \beta_7 Solve + \beta_8 STF + \beta_{10} Gender + \beta_{11} Tid + \beta_{12} Geo + \beta_{13} Alder$$

Den totale trivsel kunne på samme måde som i FM-2 formodes at forklare al variationen i den faglige trivsel, hvorfor variablen *TTF* blev fjernet som forklarende variabel i FM-3 for at undgå *quasi-complete separation*. Med samme argument som i FM-2 blev variablen *EHU* også taget ud, for at undgå omvendt kausalitet. FM-3 blev estimeret i SAS, hvor variablerne *Tid*, *Faglig* og *Lyst* viste sig at være signifikante.



Arealet under ROC-kurven (AUC=0.86) jf. *Model 3*, antyder en høj forklaringsgrad af EM-3. Jf. *Model 3* viste analysen en signifikant negativ sammenhæng (-1.03) mellem variablen *Kedlig* og respondenternes faglige trivsel (*FTF*). Resultatet viste, at desto

² Høj: defineret som et svar mellem 3 og 5 på Likert-skalaen

højere evalueret kedsomhed en respondent havde i folkeskoleundervisningen desto større sandsynlighed var der for, at respondenten ville opleve lav³ faglig trivsel.

Ydermere viste analysen jf. *Model 3* positive signifikante sammenhænge mellem variablerne *Faglig* (+0.65), *Lyst* (+0.80) og respondenternes faglige trivsel (*FTF*). Resultater, der viste, at en øgning i responderne opfattelse af egne faglige evner eller øget lyst til læring på baggrund af folkeskoleundervisningen ville øge respondenternes sandsynlighed for at opnå høj faglig trivsel.

Konklusion

Denne rapport havde til formål at undersøge sammenhængen mellem den højt begavede gruppens evaluering af trivsel i folkeskolen og deres senere uddannelsesniveau samt tidsperspektivet i opsporingen af den høje begavelses. Data blev indsamlet ved en spørgeskemaundersøgelse blandt de to procent højst begavede individer i Danmark og analyseret ved brug af multipel logistisk regression. Den deskriptive statistik viste, at den højt begavede gruppe trivselsmæssigt adskilte sig fra den gennemsnitlige folkeskoleelever. Den videre analyse viste: 1) En signifikant negativ sammenhæng mellem en øget kedsomhed i folkeskoleundervisningen og uddannelsesniveau, 2) En signifikant positiv sammenhæng mellem tidlig opsporing af den høje begavelse og uddannelsesniveau og 3) En signifikant positiv sammenhæng mellem en øget samarbejdsevne og social trivsel i folkeskolen. Disse resultater peger på følgende initiativer, der med fordel kunne iværksættes for de højt begavede børn: 1) Tidlig intelligenstestning, 2) Øget faglige stimuli og 3) Relationsdannelse med ligesindede. Rapportens hensigt var at lægge et grundlag for videre forskning i understøttelsen af de højt begavede børn. Hvorvidt den almene folkeskole skal kunne understøtte de højt begavede børn, er i sidste ende en politisk beslutning om, hvordan de knappe ressourcer fordeles. Denne rapport kan ikke konkludere, hvorvidt midlerne fra politisk side skal allokeres anderledes (end de er nu), men har rettet fokus mod de udfordringer vi som samfund står over for i forhold til de højt begavede børn.

Afsluttende note: Dette er blot en sammenskrivning af en længere rapport, der ved interesse kan sendes ved kontakt til frejafalk@hotmail.com.

³ Lav: defineret som et svar på enten 1 eller 2 på Likert-skalaen

Bibliografi

- Andersen, Pia Lægaard. 2020. »LinkedIn.« *Hvis du er så klog..., hvorfor har du så ikke en uddannelse?* 20. marts. <https://www.linkedin.com/pulse/hvis-du-er-så-klog-hvorfor-har-ikke-en-uddannelse-andersen/?originalSubdomain=dk>.
- Andreasen, Asger Graa, og Mette Lausten. 2019. *Trivselssudfordringer ved skolestart.* Forskning og analyse, Vive.
- Autor, David H., og Brenden Price. 2013. *The Changing Task Composition of the US Labor Market.* The economist.
- Borch, Ingvild. u.d. *Autoriseret psykolog Ingvild Borch.* Senest hentet eller vist den 9. marts 2022. <https://psykologingvild.dk/hoejt-begavede-boern/>.
- Erhvervsministeriet. 2021. *Immatrielle aktiver.* Handlingsplan for et innovativt Danmark, Erhvervsministeriet .
- Forman, Julie. 2012. *Simpel og multipel logistisk reggression.* Forelæsningslides, Københavns Universitet, Biostatistisk Afdeling.
- Franck, Dea. u.d. *Følelser og asynkron udvikling.* Publikation , Dea Franck.
- Franck, Dea, interviewet af Freja Falk. 2022. *Interview* (14. februar).
- Hart, M. 2007. »Understanding human history .« *Washington Summit Publishers.*
- Jørgensen, Jan Reitz. 2007. *Danmarks strategi for livslang læring.* Redegørelse til EU-kommissionen, Undervisningsministeriet.
- Lund, Mikael. 2022. 8. marts. <https://vafo.dk/artikel/corona-har-påvirket-mange-unges-trivsel-negativt>.
- Mehlby, Jill, Lasse Hønge Flarup, og Katrine Iversen. 2015. *Indsatser målrettet højt begavede børns faglige udvikling og trivsel.* Det nationale institut for kommuners og regioners analyse og forskning, KORA.
- Murtza, Muhammad Hamid, Shahzad Ali Gill, Hassan Danial Aslam, og Amna Noor. 2020. »Intelligence quotient, job satisfaction, and job performance: The moderating role of personality type.« *Journal of Public Affairs* , august.
- Niels Christian Vilmann. 2022. »Ekspert kritiserer ministers lukkeplan for enhed til kluge børn: Vi risikerer, at de simpelthen giver op.« *Berlingske*.
- Nissen, Poul, Ole Kyed, og Kirsten Baltzer. 2011. *Identifikation af begavede og talentfulde elever - hvordan gør man?* Center for Grundskoleforskning.
- Rubenstein, Lisa Davia, Del Siegle, Sally M. Reis, D. Betsy McCoach, og Meredith Greene Burton. 2012. »A Complex quest: The development and research of underachievement interventions for gifted students.« *Psychology in the Schools*, 14. juni.
- Schødt, Ann C. u.d. *Belastende begavet.* Senest hentet eller vist den 31. marts 2022. <https://belastendebegavet.dk/anderledes-begavede-overset-understimuleret/>.
- . 2016. »Asynkronitet & intelligens .« *Potentiale fabrikken*, 19. oktober.
- undervisningsministeriet, Børne- og. 2018/2019. *Uddannelsesstatistik.* <https://uddannelsesstatistik.dk/Pages/Reports/1792.aspx>.
- Varemærkestyrelsen, Patent- og. 2021. »Ny global undersøgelse: Danmark har potentiale til at få mere værdi ud ad store investeringer i forskning og udvikling.« *Ritzau*, September.

Classification using binary and continuous variables

Guillermina Eslava and Gonzalo Pérez.

Faculty of Sciences, National Autonomous University of Mexico

Abstract

This work presents comparative results of some classification methods for the case of two populations and a set of binary and continuous variables. Methods used include linear and non-linear discriminant analysis, logistic discrimination, discriminant analysis based on conditional Gaussian distributions with a tree graph structure, and random forests. These methods are compared in terms of classification error rates on simulated data.

Key words. Conditional Gaussian distribution, Homogeneous mixed graphical models, Linear and non-linear discriminants, Logistic discrimination, Misclassification error, Mixed binary and continuous variables, Supervised classification, Random forests.

1 Introduction

Classification of labeled observations based on a set of measurements is a problem that has been developed within statistics in what is known as supervised classification. There are various methods that have been known for some time, like discriminant analysis in its different forms, parametric, non parametric, linear and non-linear, and logistic discrimination. More sophisticated techniques, such as probabilistic graphical models both with undirected or with directed graphs, corresponding to Markov and Bayesian networks, are parametric models which can also be used for classification. More recently, a surge of algorithmic methods has become popular for data-driven approach to classification, such as random forests, neural networks and more recently deep neural networks.

When dealing with numerical covariates, most classification methods apply. On the other hand, when all covariates are categorical, not all classification methods apply directly, as is the case of linear and quadratic discriminant analysis although they are often nonetheless successfully used. For a set of measurements which consist of both categorical and continuous variables, some methods become more difficult to apply, in particular those based on probabilistic graphical models. Although the theory for graphical models for mixed variables has been successfully developed, see e.g. Lauritzen (1996, Ch. 6), and some algorithms for model identification and model estimation exist, the use of these models in practical applications is still somewhat limited due to scarce availability of software. Note however, that some software has been available for some time, see e.g. Højsgaard et al. (2012, Ch. 5).

In this note we present a small simulation study where we consider classification between two groups on the basis of four binary and six continuous variables. We consider simulated datasets generated from a homogeneous conditional Gaussian distribution, with two independent paths as graph structure, one for the binary and one for the continuous variables. We apply three linear methods: linear discriminant analysis, naive discriminant analysis, and linear logistic regression. We also consider six non-linear classification methods: three modified discriminant functions, two logistic regressions with interaction terms, and random forests. We briefly present each method, and then provide the respective numerical results. This simulation study is part of ongoing research, where we explore other simulation settings and real data.

2 Classification methods

We consider the problem of classification between two well defined classes of individuals, Π_1 and Π_2 , on the basis of p variables measured on a sample of individuals from each class. Let $C \in \{1, 2\}$ be the class variable and $\mathbf{x} = (x_1, \dots, x_p)$ the random vector of p variables. Let $\pi_1 = P(C = 1)$ and $\pi_2 = P(C = 2)$ be the prior probabilities that an observed individual belongs to class Π_1 and Π_2 , and $P(C = 1|\mathbf{x})$ and $P(C = 2|\mathbf{x})$ be the posterior probabilities, respectively.

2.1 Bayes rule

The Bayes classification rule is to choose the class with the higher posterior probability. That is assign an observation to Π_1 if

$$P(C = 1|\mathbf{x}) > P(C = 2|\mathbf{x}). \quad (1)$$

If one assumes that \mathbf{x} has a density or probability function $f_c(\mathbf{x}|c) = f_c(\mathbf{x})$ in population c , $c = 1, 2$, the Bayes rule (1) is equivalent to assign an observation to Π_1 if

$$\log \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} - \log \frac{\pi_2}{\pi_1} > 0. \quad (2)$$

The rule (1) is optimal in the sense that minimizes the error rate or probability of misclassification $P(e)$ defined as

$$P(e) = \pi_1 P(2|1) + \pi_2 P(1|2), \quad (3)$$

where $P(i|j)$ denotes the probability of assigning an observation from population Π_j to Π_i .

The vector $\mathbf{x} = (x_1, \dots, x_p)$ may contain only continuous, only discrete, or a mixture of both types of variables.

2.2 Discrimination for normal populations

If one assumes that \mathbf{x} has a multivariate Gaussian density $f_c(\mathbf{x})$ with mean $\boldsymbol{\mu}_c$ and covariance $\boldsymbol{\Sigma}_c$ in population c , $c = 1, 2$, the left side of the equation (2) is the quadratic

discriminant function given by

$$\frac{1}{2} \log \frac{|\boldsymbol{\Sigma}_1^{-1}|}{|\boldsymbol{\Sigma}_2^{-1}|} + \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}_2^{-1} (\mathbf{x} - \boldsymbol{\mu}_2) - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) - \log \frac{\pi_2}{\pi_1}. \quad (4)$$

Assuming $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$, expresion (4) reduces to the following linear discriminant function:

$$\mathbf{x}^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \frac{1}{2} (\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) - \log \frac{\pi_2}{\pi_1}. \quad (5)$$

The naive classifier is obtained by assuming a Gaussian distribution on each population where $\boldsymbol{\Sigma}_c$ is a diagonal matrix with elements $\{\sigma_{1c}^2, \dots, \sigma_{pc}^2\}$. The derived function obtained from (4) becomes:

$$\sum_{i=1}^p \left[\frac{1}{2} \log \frac{\sigma_{i2}^2}{\sigma_{i1}^2} - \frac{1}{2} \frac{(x_i - \mu_{i1})^2}{\sigma_{i1}^2} + \frac{1}{2} \frac{(x_i - \mu_{i2})^2}{\sigma_{i2}^2} \right] - \log \frac{\pi_2}{\pi_1}. \quad (6)$$

2.3 Logistic discrimination

Logistic regression can be used for classification when variables are both binary and continuous. It provides the posterior probabilities which are needed to use the Bayes discriminant rule (1), namely by specifying

$$P(C = 1|\mathbf{x}) = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p) P(C = 2|\mathbf{x}), \quad (7)$$

$$P(C = 2|\mathbf{x}) = 1/[1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)]. \quad (8)$$

Logistic regression with interaction terms can similarly be used to compute the posterior probabilities.

2.4 Mixed graphical models

When both binary and continuous variables are present, we may consider specifying two conditionally independent graphical models, one for each of the two classes, resulting in a single classifier. Let $\Delta = \{i_1, \dots, i_q\}$ be the set of q binary variables and $\Gamma = \{y_1, \dots, y_r\}$ the set of r continuous variables. The set of $p = q + r$ variables $\mathbf{x} = (\mathbf{i}, \mathbf{y}) = (i_1, \dots, i_q, y_1, \dots, y_r)$ follows a conditional Gaussian density that satisfies the Markov properties with respect to an undirected marked graph $G_c = (V, E_c)$, with $V = \Delta \cup \Gamma$. The density function is

$$f_c(\mathbf{x}) = p_c(\mathbf{i}) f(\mathbf{y}|\mathbf{i}, c) \quad (9)$$

where

$$p_c(\mathbf{i}) = P(\mathbf{x}_\Delta = \mathbf{i} | C = c) > 0 \text{ and } f(\mathbf{y}|\mathbf{i}, c) = N_{|\Gamma|}(\boldsymbol{\mu}_c(\mathbf{i}), \boldsymbol{\Sigma}_c(\mathbf{i})). \quad (10)$$

The Markov properties with respect to $G_c = (V, E_c)$ impose some restrictions on $p_c(\mathbf{i})$, $\boldsymbol{\mu}_c(\mathbf{i})$ and $\boldsymbol{\Sigma}_c(\mathbf{i})$. The density functions in (9), for populations $c = 1$ and $c = 2$, can be used in the Bayes rule given in (2). In practice, when using a conditional Gaussian density, the graph structure $G_c = (V, E_c)$ should be identified or estimated, and the parameters

$p_c(\mathbf{i})$, $\mu_c(\mathbf{i})$ and $\Sigma_c(\mathbf{i})$ should be estimated.

Here, we restrict the identification of the graph structure $G_c = (V, E_c)$ to be a decomposable graph, specifically a tree structure. For this structure, the estimation of the density is simple, considering that $p_c(\mathbf{i})$, $\mu_c(\mathbf{i})$ and $\Sigma_c(\mathbf{i})$, $c : 1, 2$, have exact maximum likelihood expressions.

2.5 Naive Classifier

The naive classifier is obtained by assuming independence among the p variables x_1, \dots, x_p in each population. This classifier can be obtained considering a mixed graphical model in each population with the empty graph as the interaction graph, see for example Figure 1b), which corresponds to the empty graph for four binary and six continuous variables. Considering $\mathbf{x} = (\mathbf{i}, \mathbf{y}) = (i_1, \dots, i_q, y_1, \dots, y_r)$, the left side of the equation (2) is given by

$$\sum_{j=1}^q \left[\log \frac{p_{j1}^{i_j} (1 - p_{j1})^{1-i_j}}{p_{j2}^{i_j} (1 - p_{j2})^{1-i_j}} \right] + \sum_{j=1}^r \left[\frac{1}{2} \log \frac{\sigma_{j2}^2}{\sigma_{j1}^2} - \frac{1}{2} \frac{(y_j - \mu_{j1})^2}{\sigma_{j1}^2} + \frac{1}{2} \frac{(y_j - \mu_{j2})^2}{\sigma_{j2}^2} \right] - \log \frac{\pi_2}{\pi_1}, \quad (11)$$

where $p_{jc} = P(i_j = 1|c)$, $j = 1, \dots, q$, and $y_j|c \sim N(\mu_{jc}, \sigma_{jc}^2)$, $j = 1, \dots, r$, $c = 1, 2$.

2.6 Random forests

Decision trees together with the use of the bootstrap method are the main ingredients of random forests. This methodology is used as a classifier, often successfully and with computational efficiency. Such method can be applied to data with both binary and continuous variables, and can additionally provide a measure of the importance of each of the variables in terms of classification performance. Random forests are not invariant to transformations of the covariates, and some of their parameters require tuning, for instance: the number of fitted trees, the number of variables allowed to participate on each one, the maximum depth of each tree, and the minimum number of observations at each terminal node of the trees.

3 Simulation study

We simulate data from a conditional Gaussian distribution in each population. The interaction graph is the same for both populations and is composed of two unconnected paths, one for four binary variables and another for six continuous variables, as shown in Figure 1d). The density function of $\mathbf{x} = (\mathbf{i}, \mathbf{y}) = (i_1, \dots, i_4, y_1, \dots, y_6)$ is given by

$$\begin{aligned} f_c(\mathbf{i}, \mathbf{y}) &= p_c(\mathbf{i})f_c(\mathbf{y} | \mathbf{i}) \\ &= \frac{p_c(i_1, i_2)p_c(i_2, i_3)p_c(i_3, i_4)}{p_c(i_2)p_c(i_3)} f_c(y_1, \dots, y_6 | i_1, i_2, i_3, i_4), \end{aligned} \quad (12)$$

where $f_c(y_1, \dots, y_6 \mid i_1, i_2, i_3, i_4)$ is the density of a $N(\boldsymbol{\mu}_c(\boldsymbol{i}), \boldsymbol{\Sigma}_c(\boldsymbol{i})) = N(\mathbf{0}, \boldsymbol{\Sigma}_c)$, for $c = 1, 2$, and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}$ with $\rho = 0.3$ and $\boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$ with $\rho = -0.3$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & 0 & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & 0 & 0 \\ 0 & 0 & -\rho & 1 + \rho^2 & -\rho & 0 \\ 0 & 0 & 0 & -\rho & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & 0 & -\rho & 1 \end{pmatrix}. \quad (13)$$

The probability $p_c(\boldsymbol{i})$ is computed as in (12) considering that the bivariate probabilities associated with the connected binary variables in Figure 1d) are the same, that is, $p_c(i_1, i_2) = p_c(i_2, i_3) = p_c(i_3, i_4)$, where $p_1(0, 0) = p_1(1, 1) = 0.325$ and $p_1(0, 1) = p_1(1, 0) = 0.175$ for Π_1 , and $p_2(0, 0) = p_2(1, 1) = 0.175$ and $p_2(0, 1) = p_2(1, 0) = 0.325$ for Π_2 .

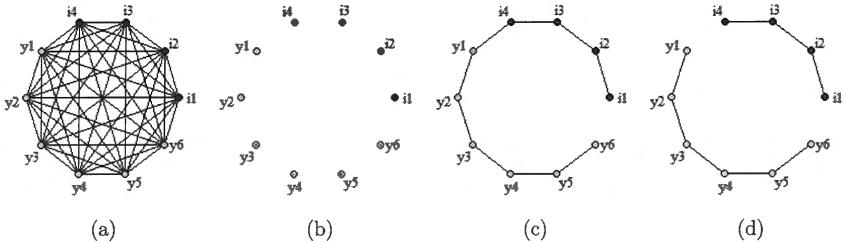


Figure 1: Interaction graph structures: a) complete: all interactions, b) empty: no interactions, c) path and d) two unconnected paths.

Table 1 and Figures 2 and 3 show the error rates obtained with 1000 training and test sets of size 50, 100 and 1000 in each population for nine methods, three linear and six non-linear. The computation of the errors was done with R (R Core Team, 2021), specifically with the following functions: *glm* to fit logistic regressions and, when selecting a model, *step* using BIC criterion; *naiveBayes* in *e1071* package (Meyer et al., 2020) for the naive classifier; *lda* and *qda* in *MASS* package (Venables and Ripley, 2002) for linear and quadratic discriminant analysis; *minForest* in *gRapHD* package (Abreu et al., 2010) to select a tree structure and various functions in *bnlearn* package (Scutari, 2017) to estimate the parameters and evaluate the discriminant function based on the CG distributions; and *randomForest* (Liaw and Wiener, 2002) to build a random forest using 1000 trees and tuning the *mtry* option.

Method	Error rates %					
	Training set			Test set		
Group sample size	50	100	1000	50	100	1000
Linear discriminant analysis LDA	37.1	41.0	47.2	50.2	49.8	50.0
Naive classifier	34.7	39.3	46.7	50.3	49.8	50.0
Logistic regression	37.1	41.0	47.2	50.2	49.8	50.0
CGD with tree structure	15.2	17.9	20.1	29.2	24.0	20.7
Quadratic discriminant analysis QDA	12.0	15.9	19.9	28.5	24.7	20.9
Best reduced Logistic with two-way interactions	0.0	15.2	20.3	35.8	26.7	20.9
Logistic with two-way interactions	0.1	9.6	19.6	38.4	29.4	21.2
LDA with two-way interactions	6.5	14.1	20.6	34.8	27.9	21.8
Random forests	0.0	0.2	0.0	37.3	33.0	23.3

Table 1: Training and test error rates for the simulated data based on 1000 training and test sets of size 50, 100 and 1000. The estimated Bayes error rate of 20.2% was computed using rule (1) with 20,000 observations from each population.

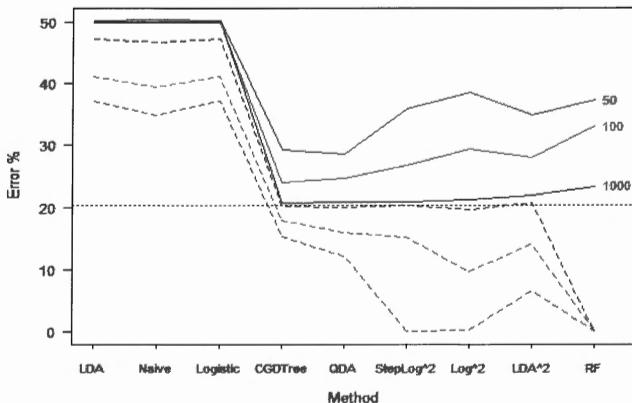


Figure 2: Estimated error rates for the simulated data based on 1000 training and test sets of size 50 (blue), 100 (red) and 1000 (green): test (solid lines) and training error (dashed lines). The dotted line shows the estimated Bayes error rate of 20.2%.

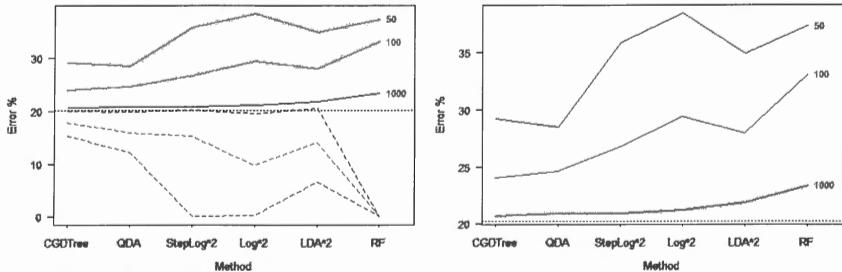


Figure 3: Estimated error rates for the simulated data based on 1000 training and test sets of size 50 (blue), 100 (red) and 1000 (green): test (solid lines) and training error (dashed lines). The dotted line shows the estimated Bayes error rate of 20.2%.

4 Discussion

The data set analysed here was generated from two homogeneous conditional Gaussian densities, one for each class. The simulation experiment was designed to show an instance where a linear discriminant function will not discriminate between the two populations. The main characteristics of the setting are the following.

- i) Two conditional Gaussian densities $f_1(\mathbf{i}, \mathbf{y})$ and $f_2(\mathbf{i}, \mathbf{y})$ were used.
- ii) There are $2^q = 2^4 = 16$ cells or different combinations of the values of the $q = 4$ binary variables in \mathbf{i} .
- iii) The conditional Gaussian distribution for each population is homogeneous, this implies that its mean vector and covariance matrix do not depend on the cell value or location, i.e. $\mu(\mathbf{i}) = \mu$ and $\Sigma(\mathbf{i}) = \Sigma$ for all cells.
- iv) $f_1(\mathbf{y} \mid \mathbf{i})$ and $f_2(\mathbf{y} \mid \mathbf{i})$ have mean vectors equal to zero and different covariance matrices, i.e. $\mu_1(\mathbf{i}) = \mu_2(\mathbf{i}) = \mathbf{0}$ and $\Sigma_1 \neq \Sigma_2$.
- v) Although $f_1(\mathbf{y} \mid \mathbf{i})$ and $f_2(\mathbf{y} \mid \mathbf{i})$ have different covariance matrices $\Sigma_1 \neq \Sigma_2$, the marginal density function of each variable is the same in the two populations, i.e. $y_j \mid \mathbf{i} \sim N(0, 1)$, $j = 1, \dots, 6$, for each cell in both populations.
- vi) The two marginal probability distributions of the four binary variables in (9) are different, i.e. $p_1(\mathbf{i}) \neq p_2(\mathbf{i})$.
- vii) The difference between $p_1(\mathbf{i})$ and $p_2(\mathbf{i})$ lies in the bivariate marginal distributions: $p_1(i_1, i_2) \neq p_2(i_1, i_2)$, $p_1(i_2, i_3) \neq p_2(i_2, i_3)$ and $p_1(i_3, i_4) \neq p_2(i_3, i_4)$. The univariate marginal probability functions are the same in both populations, i.e. $p_1(i_j) = p_2(i_j) = .5$ for $i_j \in \{i_1, i_2, i_3, i_4\}$.
- viii) The graph structure associated with $f_1(\mathbf{i}, \mathbf{y})$ and $f_2(\mathbf{i}, \mathbf{y})$ is composed of two unconnected paths as shown in Figure 1d).

The characteristics of the conditional Gaussian densities $f_1(\mathbf{i}, \mathbf{y})$ and $f_2(\mathbf{i}, \mathbf{y})$ for this setting were selected so that the difference between the two populations was due to interactions between pairs of variables. In this case a linear discriminant function will not discriminate between the two populations. Similar examples where linear discriminant

analysis does not work, have been shown for example by Krzanowski (1977). The fact that the two populations have the equal marginal means might not be interesting in most practical applications, though an interesting example is given in Bartlett and Please (1963) where they deal with the discrimination problem in the case of zero mean differences.

The numerical results in Table 1 for this particular example show that the three linear classifiers did not discriminate between the two populations, as expected by design. Test errors were around 50%.

Considering the non-linear classifiers we note the following.

- i) The five parametric classifiers for sample size 1000 on each group performed well. Their test error was at most one percentage point from the estimated Bayes error of 20.2%. The nonparametric classifier, Random forests, also performed well, its test error was three percentage points higher than the estimated Bayes error.
- ii) For sample size 100 on each group. The test error rate of each parametric classifier differs between three and nine percentage points from the Bayes error, and the test error for Random forests was about ten percentage points higher than the Bayes error.
- iii) For the small sample size of 50 observations from each population, the quadratic discriminant function and the discriminant function based on conditional Gaussian densities had the best performance with a difference of about nine percentage points from the Bayes error rate. All other parametric classifiers together with Random forests had a difference of between 14 and 18 percentage points from the Bayes error.
- iv) For the five parametric classifiers the difference between training and test errors for the sample size 1000 is less than two percentage points. However, for the sample size 50, the training error is very low, at most 15.2% and zero in some cases.
- v) Random forests, for this example, had zero average training error for the three sample sizes.

We conclude saying that the use of a classification method should be accompanied with training and test errors, resampling methods can be used for real datasets, and more than one classifier should be used to better assess the performance of a specific classifier. As Taylor mentions when commenting about the comparison of classifiers in the discussion of Ripley (1994, p. 441):

A comparison of methods ... is often difficult to interpret. Observed differences in goodness of result can arise from:

- (a) different suitabilities of the basic methods for given data sets,
- (b) different sophistications of default procedures for parameter settings,
- (c) different sophistication of the program user in selection of options and tuning of parameters and
- (d) the occurrence of effectiveness of processing of the data by the user.

Acknowledgements Guillermina Eslava gratefully acknowledges the hospitality of the Department of Applied Mathematics and Computer Science, Technical University of Denmark. This work was done while she was on Sabbatical leave from the Faculty of Sciences at the National Autonomous University of Mexico (UNAM), and gratefully acknowledges a six months grant from the program PASPA, DGAPA, UNAM.

References

- Abreu, G., Edwards, D., Labouriau, R. (2010) High-Dimensional Graphical Model Search with the gRapHD R Package. *Journal of Statistical Software* 37(1), 1–18
- Bartlett, M.S., Please, N.W. (1963) Discrimination in the case of zero mean differences. *Biometrika* 50, 17–21
- Højsgaard, S., Lauritzen, S.L., Edwards, D. (2012) *Graphical Models with R*. Springer, New York
- Krzanowski, W.J. (1977) The performance of Fisher's linear discriminant function under non-optimal conditions. *Technometrics*, 19, 191–200
- Lauritzen, S.L. (1996) *Graphical Models*. Clarendon Press, Oxford
- Liaw, A., Wiener, M. (2002) Classification and Regression by randomForest. *R News* 2(3), 18–22
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch F. (2020). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien
- R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ripley, B. D. (1994) Neural Networks and Related Methods for Classification. *J. R. Statist. Soc. B* 56, 409–456
- Scutari, M. (2017) Bayesian Network Constraint-Based Structure Learning Algorithms: Parallel and Optimized Implementations in the bnlearn R Package. *Journal of Statistical Software*, 77(2), 1–20
- Venables, W.N., Ripley, B.D. (2002) *Modern Applied Statistics with S*. Springer, New York

Inferens for den relative risiko, når estimatet er 0

Søren Möller^{1,2} og Linda Juel Ahrenfeldt³

¹ Klinisk Institut, Syddansk Universitet (moeller@health.sdu.dk)

² Open Patient data Explorative Network, Odense Universitetshospital

³ Epidemiologi, Biostatistik og Biodemografi, Institut for Sundhedstjenesteforskning,
Syddansk Universitet
(lahrenfeldt@health.sdu.dk)

Bidraget er baseret på artiklen [9] i IJERPH i 2021 af de samme forfattere.

Introduktion

Relativ risiko (RR) er det foretrukne associationsmål i mange sundhedsvidenskabelige studier, herunder kliniske interventionsforsøg og epidemiologiske cohortestudier. Inferens for RR, specielt i form af konfidensintervaller, kan bestemmes ved velkendte standardmetoder, hvilke der bl.a. undervises i på indledende biostatistikkurser. De fleste af disse metoder bygger dog på normalfordelingstilnærmelser eller andre asymptotiske forhold, og kræver derfor at både stikprøven og prævalensen af udfaldet er tilstrækkelig stor. Disse krav er typisk opfyldt for det primære udfald i et studie, men RR bruges også jævnligt til at undersøge risikoen for alvorlige bivirkninger og andre sjældne sekundære udfald, det vil sige i situationer, hvor et lavt antal prævalente tilfælde, potentielt 0, er forventet. I sådanne tilfælde vil mange inferensmetoder for RR enten ikke være anvendelige eller resultere i misvisende resultater. Som et eksempel rapporterede artiklen [3] konfidensintervallet (0; 0) for en relativ risiko, indikerende, at den undersøgte bivirkning var udelukket i eksponeringsgruppen, hvilket desværre skyldtes misvisende bestemmelse af konfidensintervallet med bootstrapping og ikke at udfaldet er en umulighed (vi kommenterede på den konkrete problemstilling i [10]).

I dette bidrag giver vi derfor et overblik over mulige tilgange til at håndtere sådanne situationer.

Metoder der ikke virker

Klassiske metoder til at bestemme konfidensintervaller for RR vil typisk udnytte at binomialfordelinger asymptotisk kan approksimeres af en normalfordeling. Dette betyder desværre at sådanne metoder ikke vil være brugbare i situationer med 0 observerede udfald, da approksimationen først vurderes at være acceptabel, når der er mindst 10 observationer i hver celle i 2×2 -tabellen [2].

Tilsvarende vil metoder, der bruger maksimum likelihood-tilgange (f.eks. i forbindelse med binomialregression), ikke give relevante konfidensintervaller, da observationen med 0 udfald vil ligge på randen af parameterrummet, og derfor ikke vil kunne resultere i retvisende inferens.

Endvidere vil bootstrapping ikke kunne anvendes, da de 0 observerede udfald vil indebære, at der vil være 0 udfald i alle de bootstrappede stikprøver, og RR-estimaten derfor vil være 0 i alle iterationer, resulterende i et ukorrekt konfidensinterval på $(0; 0)$, hvilket gik galt i det ovenfor nævnte studie [3].

Metoder der virker

Modificerede antal

En klassisk lavpraktisk tilgang er kunstigt at øge det observerede antal udfald fra 0 til en værdi større end 0. Der er forskellige forslag til hvilken konstant, der skal lægges til, og om denne skal lægges til alle celler i 2×2 -tabellen eller kun til den celle, der er 0 [6]. Desværre viser det sig, at de resulterende konfidensintervaller er følsomme overfor valg af konstant [4].

Vores anbefaling er, såfremt denne tilgang bruges, i stedet at flytte et ikke-udfald til at være et udfald i eksponeringsgruppen med 0 udfald, for at oprettholde den samlede stikprøvestørrelse, samt bibeholde et samlet udfald, der er kompatibel med den antagede sandsynlighedsfordeling.

Metoder for odds ratio

Modsat metoder for RR kan metoder der bestemmer eksakte (kombinatoriske) konfidensintervaller for odds ratioen (OR) typisk håndtere 0 observerede udfald i den ene gruppe. Der findes en håndfuld forskellige metoder til at bestemme disse eksakte konfidensintervaller, men den mest udbredte er forslaget fra Baptista og Pike, som også er tilgængelig i softwarepakker [1,5]. En udfordring ved denne tilgang er, at OR er et andet associationsmål end RR, dog viser det sig at OR er en acceptabel approksimation for RR, såfremt udfaldet er sjældent i begge eksponeringsgrupper [8].

Bayesianske metoder

Som en tredje tilgang, kan bayesiansk estimering for RR også bruges. Den mest oplagte tilgang er at modellere sandsynligheden for udfald separat i de to grupper der sammenlignes, og så bestemme posteriorfordelingen for RR ud fra posteriorfordelingerne for disse to proportioner. Denne procedure vil typisk uden videre kunne håndtere 0 observerede udfald i en (eller begge) grupper. Resultaterne vil dog afhænge af de valgte priorfordelinger for proportionerne, hvor en Betafordelt prior vil være det naturlige valg, da denne både er teoretisk velbegrundet (som konjugeret prior til en binomialfordeling) og er let at anvende i praksis [7].

Konklusion

Vi har redegjort for, at det er muligt at rapportere revisende konfidens- eller kredibilitetsintervaller for RR også ved 0 observerede udfald i eksponeringsgruppen. Vores opfordring er derfor, at sådanne intervaller rapporteres for at dokumentere (u)sikkerheden af estimererne, men at intervallerne bestemmes med en

metode, såsom dem vi foreslår ovenfor, der kan håndtere 0 observerede udfald, sådan at misvisende resultater undgås.

Referencer

1. BAPTISTA, J., AND PIKE, M. C. Algorithm as 115: Exact two-sided confidence limits for the odds ratio in a 2×2 table. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 26, 2 (1977), 214–220.
2. BLYTH, C. R., AND STILL, H. A. Binomial confidence intervals. *Journal of the American Statistical Association* 78, 381 (1983), 108–116.
3. DAS, M. K., ARORA, N. K., POLURU, R., TATE, J. E., GUPTA, B., SHARAN, A., AGGARWAL, M. K., HALDAR, P., PARASHAR, U. D., ZUBER, P. L. F., BONHOEFFER, J., RAY, A., WAKHLU, A., VYAS, B. R., IQBAL BHAT, J., GOSWAMI, J. K., MATHAI, J., K. K., BHARADIA, L., SANKHE, L., M K. A., MOHAN, N., JENA, P. K., SARANGI, R., SHAD, R., DEBBARMA, S. K., J. S., RATAN, S. K., SARKAR, S., KUMAR, V., MAURE, C. G., DUBEY, A. P., GUPTA, A., SAM, C. J., MUFTI, G. N., TRIVEDI, H., SHAD, J., LAHIRI, K., R. K., LUTHRA, M., BEHERA, N., P. P., G. R., KUMAR, R., SARKAR, R., A, S. K., SAHOO, S. K., GHOSH, S. K., MANE, S., DASH, A., CHAROO, B. A., TRIPATHY, B. B., G. R. P., S, H. K., K, J., SARKAR, N. R., ARUNACHALAM, P., MOHAPATRA, S. S. G., AND GARGE, S. Risk of intussusception after monovalent rotavirus vaccine (Rotavac) in Indian infants: A self-controlled case series analysis. *Vaccine* 39, 1 (01 2021), 78–84.
4. DEWEY, M. E. Collated responses from r-help on confidence intervals for risk ratios, 2006.
5. FAGERLAND, M. W. Exact and mid-p confidence intervals for the odds ratio. *STATA JOURNAL* 12, 3 (2012), 505–514.
6. GART, J. J. Alternative analyses of contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 28, 1 (1966), 164–179.
7. GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A., AND RUBIN, D. B. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2011.
8. MCNUTT, L. A., WU, C., XUE, X., AND HAFNER, J. P. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *Am J Epidemiol* 157, 10 (May 2003), 940–943.
9. MÖLLER, S., AND AHRENFELDT, L. J. Estimating Relative Risk when Observing Zero Events – Frequentist Inference and Bayesian Credibility Intervals. *IJERPH* 18, 11 (2021), 5527.
10. MÖLLER, S., AND NIELSEN, S. Letter to the editor. *Vaccine* (online first 2021).

Studying Preference of Pork Hedonic or Paired comparisons

Gorm Gabrielsen, Associate Professor emeritus, Copenhagen Business School

In cooperation with M.D.Aaslyng, S.Støjer, E.V.Olsen, D.V.Byrne, W.L.P.Bredie

The overall aim of the study was to investigate the consumer preference of pork in Denmark. A methodological study was imbedded into the investigation to compare the results appearing from the use of hedonic scales to the results appearing from the application of paired comparisons. A subset ($n = 56$) of the total set of consumers first tasted four samples of meat as paired comparison. The consumers were asked to mark on a line which sample they preferred. A mark right between the two sample means that they equally liked or disliked the samples. The closer the mark was to one of the sides the more was this sample preferred. All of the six possible pairs of four samples of meat were served to the consumers giving six servings. In each comparison, the preference was scored on a 15 cm visual analog scale. After a 10 minutes break the consumers scored the same four samples of meat on a hedonic 15 cm visual analog scale. The meat was served to the consumers in a sensory laboratory (ISO 8589).

Using the two methodologies results at the aggregated level basically in the same differences in preferences, but it seems that the consumers differentiates the samples more in the paired comparisons.

In addition, the method of paired comparisons makes the individual differences visible. This may a heterogeneity among consumers, which question the concept of "mean consumer".

For the consumer to rate a meat samples into a hedonic scale often requires use of preliminary sorting. The preliminary sorting may be related to individual values - mothers cooking - or it may be related to family values or to cultural values – sometimes named "meta preferences". The choice between the use of hedonic scale or paired comparison therefore involves the question of which kind of preferences is of interest in a specific situation.

Diskussion af designet for

'National Overvågning af Arbejdsmiljøet blandt Lønmodtagere' (NOA-L)

En spørgeskemaundersøgelse, hvor lønmodtagere inviteres til at besvare spørgsmål om deres arbejdsmiljø

cand. stat. Hans Bay

Arbejdstilsynet, august 2022

Indledning:

NOA-L er en del af Beskæftigelsesministeriets overvågning af udviklingen på arbejdsmiljøområdet. Overvågningen skal bidrage til grundlaget for vidensbaseret policyudvikling, myndigheds-udøvelse og indsatser på arbejdsmiljøområdet og dermed til et sikkert og sundt arbejdsmiljø. NOA-L er en spørgeskemaundersøgelse, hvor lønmodtagere inviteres til at besvare spørgsmål om deres arbejdsmiljø. Undersøgelsen forsøger at give et billede af arbejdsmiljøet i Danmark i første halvår 2021. Data i undersøgelsen 'National Overvågning af Arbejdsmiljøet blandt Lønmodtagere' er indsamlet i første halvår 2021 under usædvanlige forhold på grund af COVID-19-epidemien. Restriktioner og nedlukning af dele af samfundet kan have haft væsentlig indflydelse på personlige forhold, arbejdsforhold og arbejdsmiljøet.

NOA-L afløser den tidligere undersøgelse 'Arbejdsmiljø og Helbred'ⁱⁱ (AH), som blev gennemført af Det Nationale Forskningscenter for Arbejdsmiljø (NFA) hvert andet år i perioden 2012-2018.

Formålet med undersøgelsen er at belyse forekomsten og udviklingen i væsentlige negative og positive faktorer i arbejdsmiljøet samt bidrage til datagrundlaget for forskning i arbejdsmiljøet. Undersøgelsen skal også anvendes til at følge op på de prioriterede nationale mål for arbejdsmiljøindsatsen, der følger af trepartsaftalen indgået mellem regeringen og arbejdsmarkedets parter i december 2020ⁱⁱⁱ. NOA-L vil blive gentaget hvert andet år.

Datagrundlaget for NOA-L

Population:

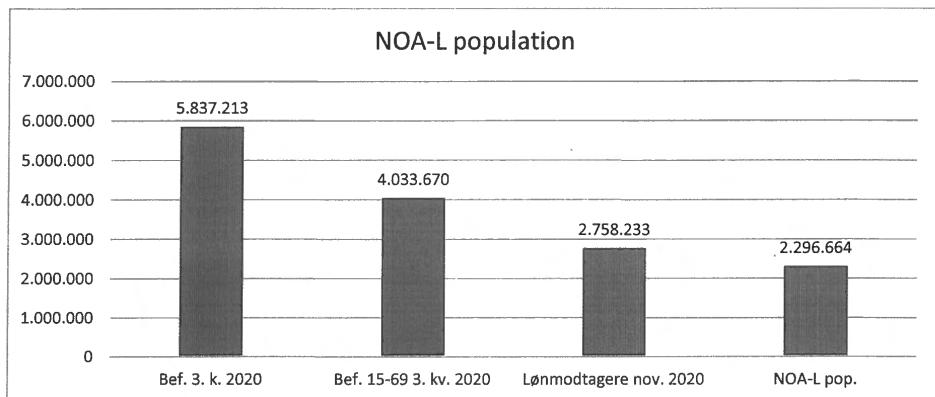
Populationen er her betegnelsen for de personer, der vil have mulighed for at blive udtrukket til undersøgelsen.

Pr. 1/1-2021 var der 2.296.664 personer i Danmark, som opfyldte følgende kriterier:

- 1) De var i alderen 15 til 69 år (opgjort pr. 1/1 2021)
- 2) De var i beskæftigelse i september, oktober og november 2020 og havde minimum 34 timers arbejde pr. måned
- 3) De var registreret i E-indkomst registeret.

I nedenstående figur er NOA-L-populationen sammenlignet med hele befolkningen, den del af befolkningen som udgøres af de 15 til 69 årige samt alle lønmodtagere.

Fig. 1. NOA-L population



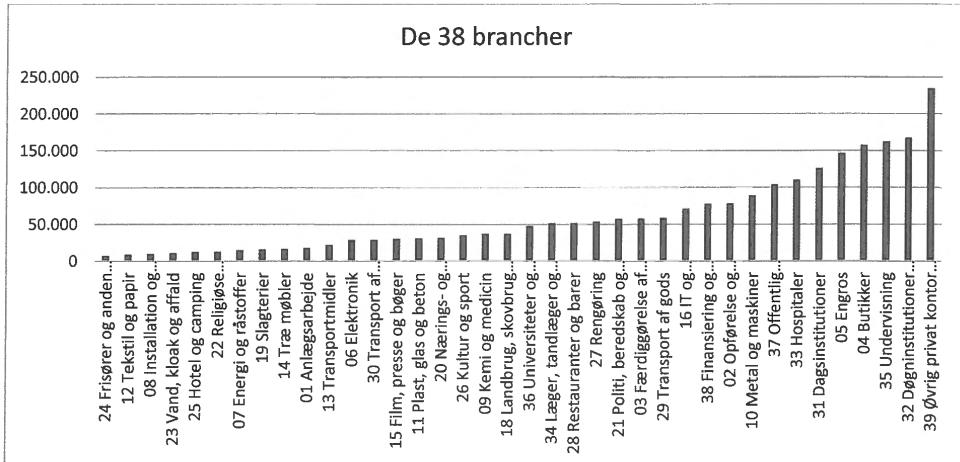
Stikprøve og stikprøvedesign.

Fra universet på 2.296.664 personer er der udtrukket 1.668 lønmodtagere simpelt tilfældigt inden for hver af 38 brancher. Stikprøven består derfor af i alt 63.391 respondenter, hvilket udgør 2,7% af hele populationen. Stikprøven er således trukket ved stratificering^{iv} (=disproportional allokering), hvor der er trukket lige mange lønmodtagere i hver branche, uanset branchens størrelse. Derved er der forskellige sandsynligheder for at blive udtrukket i forskellige brancher. I mindre brancher som 'Frisører og anden personlig pleje' er sandsynligheden for at blive udtrukket til at deltage større end i store brancher som 'Øvrige privat kontor og foreninger'.

Arbejdstilsynet opererer normalt med 36 brancher, men på grund af den disproportionale stikprøve blev branche 17 kontor opdelt i tre brancher: 37 Offentlig administration og arbejdsformidling, 38 Finansiering og forsikring og 39 Øvrig privat kontor og foreninger, for at sikre tilstrækkelig bredde i respondenterne.

Stratifikationsvariablen er de 38 brancher. I nedenstående figur er strataene gengivet. Den største branche er "øvrige privat kontor" med ca. 230.000 personer. Den mindste branche er "Frisører" med mindre end 10.000 personer.

Fig. 2. De 38 brancher fordelt efter antal lønmodtagere.



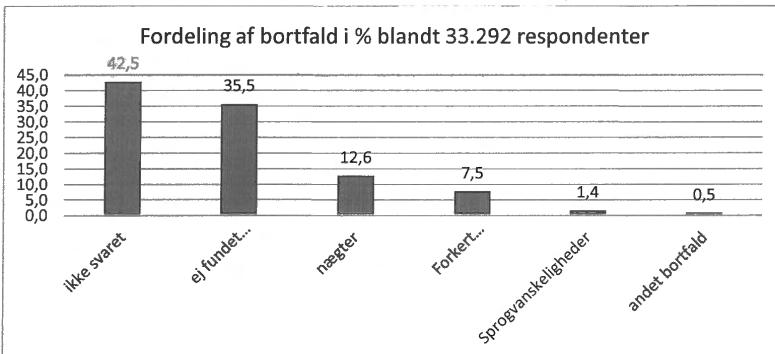
Opnåelse

Af de 63.391 kom der i alt 30.099 besvarelser retur, svarede til en opnåelsesprocent på 47,5%. Blandt disse er der 2.739 respondenter, som kun har svaret på en del af spørgsmålne, men som har svaret på så stor en del af spørgeskemaet, at de er taget med i den samlede opgørelse. Yderligere 897 besvarelser, havde svaret på så få spørgsmål, at de var delvist gennemført og derfor ikke indgår i opgørelserne.

Bortfald

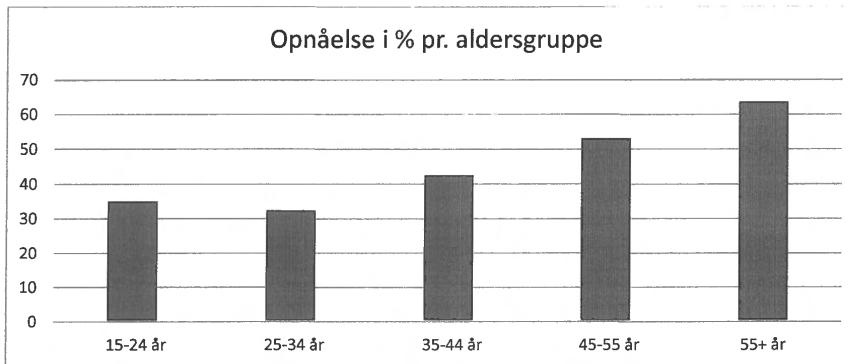
Af de 63.391 var der i alt 33.292 udtrukne personer, som ikke svarende på spørgeskemaet. Den største del af bortfaldet er 'Ikke svaret' (42,5 pct.), dvs. personer, som enten ikke har returneret spørgeskemaet, eller har returneret det med så få besvarelser, at det ikke kan indgå i opgørelsen (se nedenstående figur), mens yderligere 35,5 pct. ikke havde et telefonnummer der kunne findes.

Fig. 3. Fordeling af Bortfald



Bortfaldet afhænger af køn, etnicitet og alder. I nedenstående figur er bortfaldet som funktion af alder.

Fig. 4. Opnåelse i % efter aldersgruppe.



I mange større nationale undersøgelser vil opnåelsen stige med alderen. Dette gælder også for NOA-L. Så en diskussion af at veje (på plads) på basis af køn og alder vil næsten altid forekomme.

Vægte

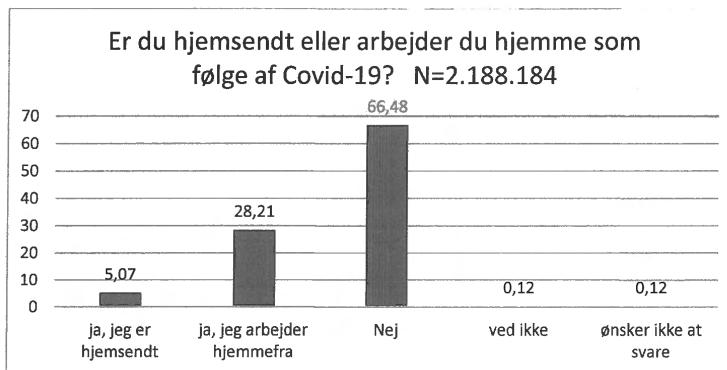
I og med at stikprøven er udtrukket disproportionalt (på basis af stratifikationen efter brancher) så vil dette alene kræve, at den indsamlede stikprøve udstyres med vægte. Det er Danmarks statistik (DST) der har foretaget dataindsamlingen, og det er også DST, der har beregnet vægtene. Vægtene er den vægt, som anvendes til at opregne den enkelte persons bidrag til den fulde population. De indsamlede data bruges som grundlag til en udregning af opregningsvægte. Vægtene udregnes ved hjælp af CLAN-proceduren, som er en samling SAS-makroer udviklet på Statistiska Centralbyrån (Sverige). For at finde frem til de baggrundsværdier, der har signifikant betydning for bortfaldet har DST foretaget logistisk regressionsanalyse. På baggrund af den analyse er vægtene beregnet ud fra køn, alder, uddannelse, herkomst og jobgruppe. Endelig er hele populationen vægtet på plads efter branche.

COVID-19

Dataindsamlingen har fundet sted i en periode, hvor Danmark var lukket ned.

Populationen er trukket ultimo 2020, hvor landet var så godt som fuldt åbent, mens spørgeskemaerne er sendt ud i en periode, hvor mange brancher var lukket. Dette kan delvis forklare, at 4,73 pct. af de adspurgte svarer, at de ikke er i arbejde da de modtager invitation til spørgeskemaundersøgelsen. Yderligere er der blevet spurgt til om 'man er hjemsendt eller arbejder hjemme som følge af Covid-19'. Her svarer 66 pct. af respondenterne, at de hverken arbejder hjemme eller er hjemsendt som følge af Covid-19 (Fig. 5). De respondenter, der svarer, at de er hjemsendt, bliver de bedt om at svare på spørgsmålene, som de husker deres arbejdsmiljø, inden de blev sendt hjem.

Fig. 5. Hjemsendelse som følge af Covid-19

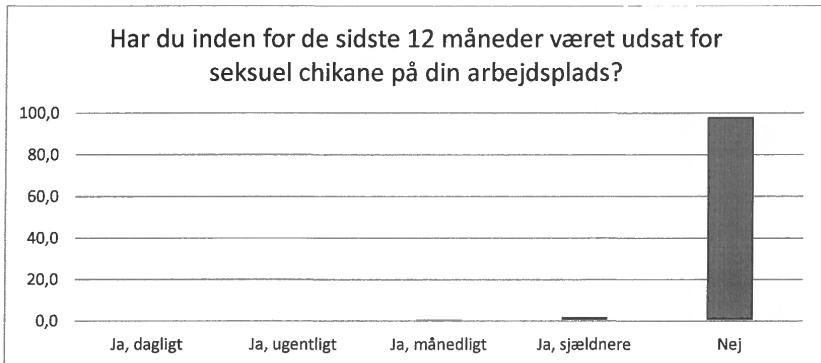


Det skal dog bemærkes, at selvom 66 pct. arbejder under nedlukningen, er det sandsynligvis under andre omstændigheder end i normale år. Dette kan have betydning for arbejdsmiljøet – Hvor nogen brancher havde mere travlt end normalt, fik andre brancher nye opgaver og endelig kan der være brancher, hvor der er færre arbejdsopgaver.

Krækende handlinger (seksuel chikane).

Et af spørgsmålene lød: Har du inden for de sidste 12 måneder været utsat for seksuel chikane på din arbejdsplads? I nedenstående figur er vis fordelingen af svarene.

Fig. 6. Seksuel chikane



Samlet svarer 2,4 % af de har været utsat for seksuel chikane inden for det sidste år.

Sammenligning over tid.

Tabel 1. Seksuel chikane i 4 undersøgelser.

undersøgelse	i %	Netto stikprøve	Univers	Opnåelses-procent	Emne	længde
AH-2016 ^v	3,7	Ca. 35.000	2.296.601	Ca. 50	Bredt om arbejdsmiljø	langt
AH-2018 ^{vi}	3,6	Ca. 30.000	2.283.198	Ca. 50	Bredt om arbejdsmiljø	langt
VIVE -2020 ^{vii}	2,8	Ca. 8.000	2.256.291	50	Krænkende handlinger	10 min.
NOA-L 2021	2,4	30.099	2.296.664	47,5	Bredt om arbejdsmiljø	langt
NFA ultimo 2021 ^{viii}	5,8	Ca. 10.000	2.322.161	27	Sexchikane	(kort)

Spørgsmålet var stort set det samme nemlig: Har du inden for de sidste 12 måneder været utsat for seksuel chikane på din arbejdsplads? VIVES spørgsmål lyder: Har du været utsat for seksuel chikane på dit arbejde inden for de sidste 12 måneder? (VIVE er Det Nationale Forsknings- og Analysecenter for Velfærd).

VIVES Spørgeskemaundersøgelse blev foretaget i perioden 2. oktober til 23. november 2020. Populationen i denne undersøgelse blev afgrenset til personer mellem 18 og 64 år, som inden for de sidste 3 måneder før undersøgelsen havde arbejdet mindst 35 timer om måneden, og som havde tjent mindst 3.000 kr. om måneden. De seneste 3 måneder dækker i denne undersøgelse over april, maj og juni 2020, hvor der var tilgængelige oplysninger om lønmodtagernes arbejdstid og indkomst. Der er også brugt stratifikation efter brancher.

NFAs undersøgelse er foretaget blandt personer, der den 1. jan 2021 var mellem 15 og 69 år, og som fra juni 2021 til september 2021 har været registreret som lønmodtager i e-indkomst med minimum 34 timer pr. måned. Her er der tilsvarende til NOA-L brugt stratifikation efter brancher. Spørgsmålet om seksuel chikane lyder: "Har du inden for de sidste 12 måneder været utsat for seksuel chikane på din arbejdsplads?"

Det store spørgsmål er: Kan man sammenligne disse tal over tid, ikke mindst når der har været COVID-19? Har eksempelvis COVID-19 bevirket at der er opstået et trendbrud, når man skal sammenligne målingerne over tid.

Forsøg på at klassificere trendbrud i surveys omhandlende arbejdsstyrken.

Ændringen (i dette tilfælde anskueliggjort ved seksuel chikane) fra 2018 til 2022 er en funktion af nedenstående elementer:

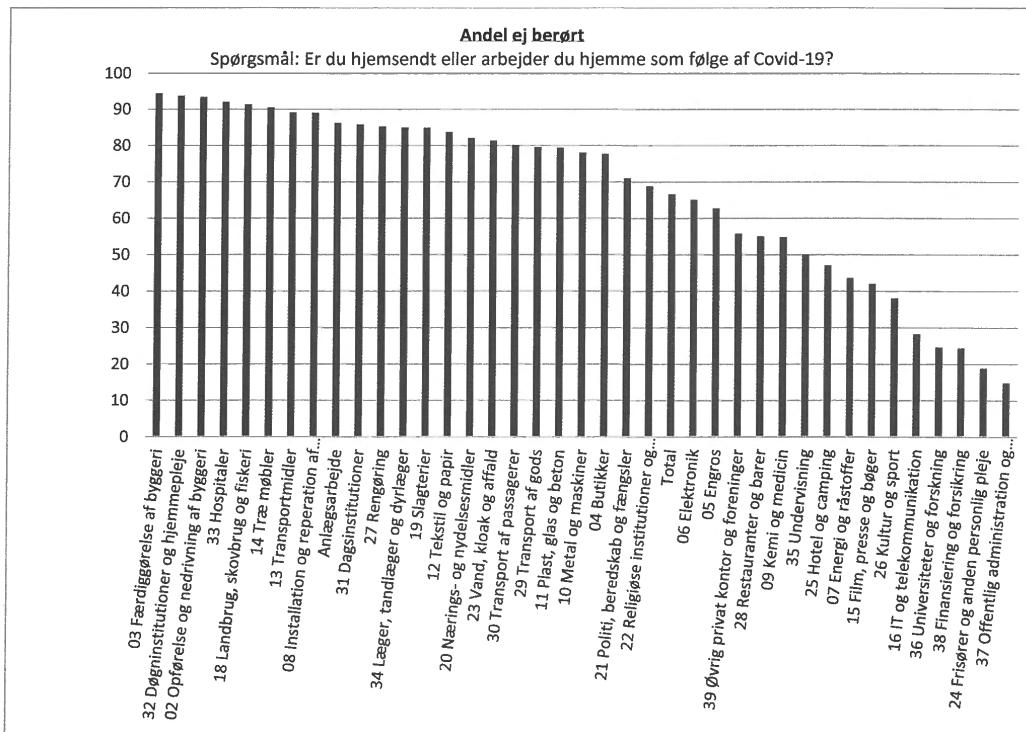
Trend brud	Årsag
Systematiske trendbrud.	Ændring af aldersgrænser. I 2018 var grænserne 18 til 65 år. I 2021 var grænserne sat til 15 og 69 år. Pensionsalderen bliver i disse år løbende løftet, dette har også betydning for størrelsen af arbejdsstyrken.
Konjunkturmæssige trendbrud	Arbejdsstyrken er også påvirket af konjunkturerne. Her medio 2022 har man eksempelvis den laveste arbejdsløshed i de seneste ti år.
Branche rotation	Et eksempel er: at mange fra restaurationsbranchen tog job som peder. Mange arbejdsmiljømål er meget afhængige af branchen.
Stød	Her tænkes på en pludselig opstået situation. COVID-19 er et godt eksempel på dette. Kunne også være at en vulkan gik i udbrud og dermed påvirkede luftfarten.
Design ændring	Markante eksempler på dette er når man gik fra besøgsinterview (F2F) til telefoninterview (CATI) og senere til postale og derefter til web-baseret interview. Der har også været diskussion af om en stratifikation (efter brancher) ville kunne bevirket et trendbrud. Det sidste er svært at måle, da man bl.a bruger vægte baseret på stratifikationen. Og at vægtene i principippet vejer stikprøven på plads.
Ændring i spørgsmål	Det er selvfølgelig klart, at ændringer i spørgsmålsformuleringerne uundgåeligt vanskeliggør en sammenligning.
Framing	Rækkefølgen af spørgsmål kan bevirket, at respondenterne får en specifik opfattelse af de senere spørgsmål. Og derfor vil ændring i rækkefølgen af spørgsmål kunne skabe en ændring over tid. Men Framing kan dække over mange facetter: herunder emnet/fokus for selve undersøgelsen.

	NFA's undersøgelse kunne være et eksempel herpå. "Der er dog tegn på, at bortfaldet ikke har været tilfældigt. Der ses i datamaterialet en højere svarprocent Det er dog muligt, at der i tillæg hertil har været en yderligere selektiv besvarelse i indeværende undersøgelse, hvor fokus er på uønsket seksuel opmærksomhed og chikane. Det er muligt, at personer, som har oplevet disse typer af adfærd, har været mere tilbøjelige til at svare end andre inviterede personer" fra NFA forskningsnotat.
Andre brud	Her er der frit slag. Men et lille eksempel kunne være at betydningen af ord ændrer sig, og at man ikke opdager dette før senere. Måske var det at have stress en "stør" ting for 40 år siden, mens det i dag er mere "naturligt".
Den "naturlige" udvikling	Det er dette man håber at måle (lidt urealistisk) ved at holde de andre faktorer uændret.

Illustration af COVID-19 indflydelse

I nedenstående figur ses, at der stor forskel på hvilke brancher, der er berørt af COVID-19 (i forbindelse med at være hjemsendt).

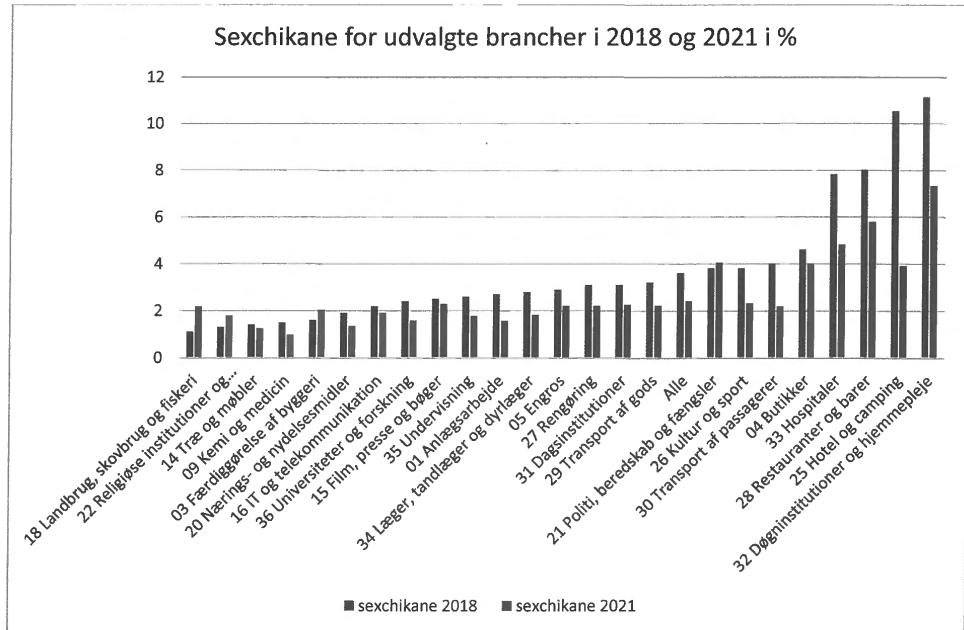
Fig. 7. Brancher fordelt efter hvor meget berørt de er af hjemsendelse.



Her skal bemærkes at brancherne: 'Restauranter og barer' samt 'Hotel og camping' ikke overraskende er ramt af hjemsendelse.

I den næste figur vises ændringen i seksuel chikane fra undersøgelsen i 2018 til 2021 for udvalgte brancher.

Fig. 8. Ændring i andelen af seksuel chikane fra 2018 til 2021.



Sammenstiller man de to grafer så er der grundlag for en hypotese der siger, at hjemsendelse mindsker seksuel chikane på arbejdspladserne.

En hypotese der siger, at COVID-19 restriktionerne (ud over hjemsendelse) også mindsker seksuel chikane på arbejdspladserne, er vanskeligere at vurdere ud fra disse to undersøgelser.

Litteratur:

-
- ⁱ <https://at.dk/arbejdsmiljoe-i-tal/national-overvaagning-af-arbejdsmiljoeet-blandt-loenmodtagere-2021/>
 - ⁱⁱ <https://nfa.dk/da/nyt/nyheder/2019/ahundersogelsen-er-en-vigtig-kilde-til-viden-om-arbejdsmiljøet>
 - ⁱⁱⁱ <https://at.dk/nyheder/2020/12/regeringen-og-arbejdsmarkedets-parter-er-enige-om-nye-maal-for-arbejdsmiljoeet/>
 - ^{iv} Om stratifikation og vægte se. Anders Milhøj: Stikprøveteori for økonomi og erhverv (bogen er udsolgt). Hans Bay: Vejning/vægtning af stikprøver. Dansk Sociologi december 2009. SURVEY: design, stikprøve, spørgeskema, analyse. (Hans Reitzel 2017).
 - ^v <https://at.dk/media/5994/danskernes-arbejdsmiljoe-2016.pdf>
 - ^{vi} <https://at.dk/arbejdsmiljoe-i-tal/arbejdsmiljoe-og-helbred-2012-2018/>
 - ^{vii} <https://www.vive.dk/da/udgivelser/en-undersogelse-af-omfanget-af-kraenkende-handlinger-og-konflikter-paa-det-danske-arbejdsmarked-16066/>
 - ^{viii} <https://at.dk/media/7574/initial-afrapportering.pdf>

Risikovillighed

En analyse af faktorer, der påvirker unges villighed til at påtage sig risiko

Af

Mogens Dilling-Hansen

Department of Economics and Business Economics

Aarhus Universitet

Mail: dilling@econ.au.dk

Resumé

Der kigges på begrebet risikovillighed i dette papir, og de mikroøkonomiske teoretiske analyser viser, at effekten af usikkerhed i høj grad afhænger af typen af usikkerhed. Den empiriske analyse undersøger 3.231 unge studerendes holdninger til usikkerhed gennem et tænkt spil med forskelle i gevinsternes størrelse. Det påvises, at 60% af respondenterne faktisk ønsker at deltage i et mere risikabelt spil end forventet, der findes ingen corona-effekter og der findes der markante kønsforskelle i en ung persons risikovillighed.

1 Indledning

I dette papir fokuseres der på begrebet risikovillighed. Den danske Ordbog forklarer ordet meget kort: Det drejer sig om personer, der er parate til at løbe en (økonomisk) risiko. Formålet med analysen er med andre ord at forklare handlinger forbundet med usikkerhed i en økonomisk model; men også at fokusere på personerne bag handlingerne.

Til formålet anvendes en model opstillet af K. J. Arrow, som blandt meget andet har formuleret modeller til beskrivelse af adfærd under usikkerhed. I modellen nyttemaksimerer effekten af usikkerhed bliver en funktion af personens risikovillighed

Opbygningen af analysen er som følger. Kapitel 2 indeholder en økonomisk model til forklaring af risikovillighed, og kapitel 3 beskrives de anvendte data i den empiriske del. Kapitel 4-6 analyser risikovilligheden og de tre opstillede hypoteser testes. Kapitel 7 runder af med en diskussion af de fundne resultater.

2 En økonomisk model for investeringer og risikovillighed

Arrow (1951) forbinder risikovilligheden til det forhold, at der ved en konkret økonomisk beslutning/handling ofte er forbundet usikkerhed til en række forhold, der er grundlæggende for den konkrete beslutning, men ikke nødvendigvis let at identificere. Til illustration af det forhold bliver konkrete udfald af en usikker hændelse klassificeret i tre kategorier - (i) hændelse alene bestemt at (ægte) usikkerhed, (ii) hændelse, som i sidste ende er opstået pga. usikkerhed og (iii) hændelse uden direkte relation til usikkerhed.

Den første type af hændelse (klassifikation 1) er inden for kategorien spil (gambling) og forsikringer, og fælles for disse handlinger er, at de er betinget af den forventede værdi af handlingen: Hvis den forventede værdi er positiv, så spilles der/tegnes der en forsikring. Den anden type (klassifikation 2) af hændelser eksemplificeres ved økonomiske aktiver, hvor værdien af aktivet afhænger af evt. usikkerhed på et givet marked: Det er usikkerheden, der påvirker lysten til at holde et økonomisk aktiv. Endelig er den tredje kategori af hændelser (klassifikation 3) baggrunden for eksistensen af (overnormal) profit: På grund af mangel på information om fremtidige hændelser, kan der skabes en overnaturlig profit. Denne form for usikkerhed har ikke direkte indflydelse på den enkeltes adfærd, medens de to første typer af usikkerhed påvirker adfærd mere direkte.

Arrow (1965) analyserer sammenhængen mellem usikkerhed, risikovillighed og investering i økonomske aktiver ved at opstille en 2-periode model til bestemmelse af formuen i ”final period”, Y , ud fra den initiale formue, A , plus afkastet af den risikable aktiv, X . Den stokastiske variabel X bestemmer dermed formuen i slut-perioden:

$$(1) Y = A + a X$$

Hvor

- A hhv. Y er formuen i de to perioder
- a er investeringerne i det risikable aktiv mellem de to perioder
- m er den del af formuen, som ikke er placeret i det risikable aktiv: $m = A - a$

Maksimeringsproblemet for den enkelte person er dermed en funktion af personens præferencer for at udsætte sig for en risiko mod en mulig gevinst, og med en traditionel mikrobaseret analyse vurderes disse præferencer ved nytte af ændringen i personens velfærd og beslutningsvariablen er a :

$$(2) W(a) = E(U(Y)) = E(U(A + aX))$$

Anvendelsen af en standard mikroøkonomisk nyttefunktion betyder, at maksimeringsproblemet (2) bliver et – næsten – triviet matematisk problem. Nyttefunktionen antages dels at være kontinuert differentielabel og dels at være risiko-avers, som betyder at grænsenyten af varen ”risikabel investering” er faldende, jo højere risiko. En nødvendig, men ikke tilstrækkelig betingelse for at der nyttemaksimeres er førsteordensbetingelsen, $W'(a_{opt}) \leq 0$.

Figur 1-3 er hentet fra Arrow (1965) til illustration af, hvordan en persons risikovillighed har indflydelse på beslutningen om at investere i et usikkert aktiv. Er den forventede værdi af afkastet negativt, $E(X) < 0$, så vil det ikke være optimalt at investere i et usikkert aktiv, og dermed vil der være nyttemaksimering for $a = 0$ (ingen investering i det usikre aktiv) – se figur 1. Figur 2 er derimod det realistiske scenario, hvor der for relativt små investeringer vil være en forventning om at ”...the positive expectations outweighs the risk” (Arrow (1965, p. 40) og dermed vil nytte-/velfærdfunktionen først være stigende hvorefter den vil falde igen grundet den generelle anta-

gelse om risikoaversion. Det sidste tilfælde, hvor $W'(A) \geq 0$, antages af Arrow ikke at være relevant; men situationen vil svare til, at en person investerer hele sin formue i et aktiv med en negativ forventning til afkastet.

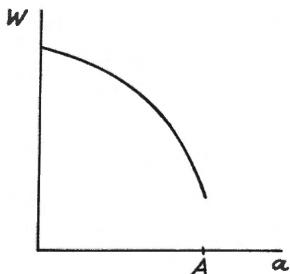


Figure 1

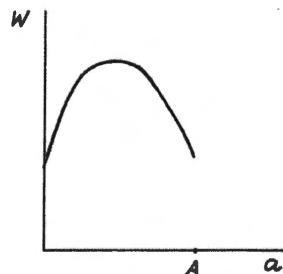


Figure 2

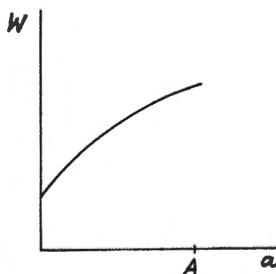


Figure 3

Kilde: Arrow (1965)

På baggrund af modellen er formålet med den empiriske model at analysere følgende to hypoteser:

Hypotese 1: *Der er forskel på lysten til at investere i et risikofyldt aktiv, givet den relative størrelse af investeringen er mindre*

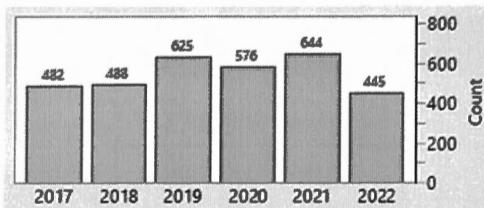
Hypotese 2: *Der er ikke forskel på risikovilligheden over tid (periode før og under covid-19)*

Hypotese 3: *Der er ikke forskel på risikovilligheden for forskellige persontyper*

3 Empirisk design og data

Data, der anvendes i analyserne, er selvrapporterede data i form af et årligt spørgeskema for erhvervsøkonomi-studerende, der i foråret i perioden 2017-2022 har accepteret at deltage i undersøgelsen.

Figur 4 Oversigt over fordelingen af brugbare svar for perioden 2017-2022



Noter. Der er samlet 3.260 brugbare respondenter i stikprøven. Alle responderne er bachelorstuderende studerende inden for erhvervsekonomi. Tidspunkt for indsamling af data er februar/marts måned, og derfor må det forventes, at observationerne for 2020-22 kan være påvirket af, at respondenten er påvirket af de generelle corona-restriktioner.

Spørgeskemaet indeholder en række baggrundsoplysninger for respondenten sammen med en række holdningsmæssige spørgsmål. For en oversigt over data, se tabel 1.

En række spørgsmål som fx alkoholforbruget har været anvendt som proxy (se SSI (2019)), men her er begrebet risikovillighed bestemt ud fra svaret på en konkret stillingtagen til hvilken form for lotteri, der foretrækkes. Det oplyses, at den samlede præmiesum er 10 mio. kroner og der er i alt 10.000 lodder (dvs sandsynlighed for gevinst bestemmes alene af antal gevinst). Der er 4 typer af lotteriet med hhv. 1, 10, 100 og 1.000 gevinst, således den mest risikovillige vil modtage 10 mio. kroner med sandsynligheden 1/10.000 og den mindst risikovillige vil modtage gevinsten på 10.000 kroner med sandsynligheden 1/10.

Figur 5 Spørgsmål til bestemmelse af risikovillighed

12. What type of lottery do you prefer?

- (1) 10.000 tickets: 1 prize (10.000.000 kr) and 9.999 blanks
- (2) 10.000 tickets: 10 prizes (1.000.000 kr) and 9.990 blanks
- (3) 10.000 tickets: 100 prizes (100.000 kr) and 9.900 blanks
- (4) 10.000 tickets: 1.000 prizes (10.000 kr) and 9.000 blanks

Tabel 1 Deskriptiv statistik for spørgeskemaundersøgelsen 2017-22

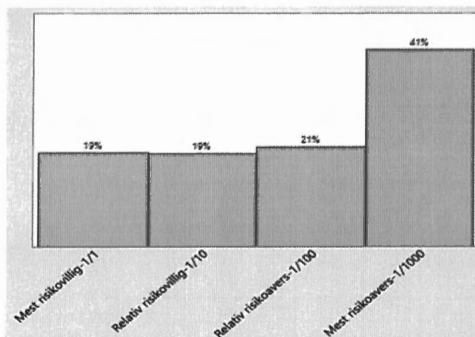
Variabel	Gennemsnit	Std. dev.
Køn – dummy (andel mænd)	0.57	-
Civilstand – dummy (andel enlige)	0.73	-
Alder	21.6	2.0
Alkohol (gns. ugentlig antal genstande)	5.8	6.1
Antal venner	11.6	20.3
Holdning (1-5) til karriere	3.9	0.9
Holdning (1-5) til miljø	3.8	1.0
Pris-bevidsthed ved køb (1-5)	3.1	1.1
Mærkevare-bevidsthed ved køb (1-5)	2.7	1.0
Idealisme (holdning til begreb)	3.1	0.8
Grad af risikoaversion (1: mest risikovillig)	2.8	1.2
Stikprøvestørrelse (n)	3.260	

4 Risikovillighed – hypoteze 1

Måling af risikovillighed er i bedste fald kun muligt på en ordinal skala, og det kan også ansføres, at den konkrete måling af graden af risikovillighed kun er gældende for klassifikation 1-typen af hændelser (se Arrow(1951) for uddybning af de fire klasser).

Ikke desto mindre er resultatet vist i figur 6 ret markant på flere måder. For det første er der stærk evidens for at den første hypoteze, som siger at jo mindre spillet fylder i det samlede budget, jo større er tilslutningen til spillet; men for det andet så påvises det (endnu en gang), at en ret stor andel af unge personer foretrækker det mest usandsynlige spil, hvor der kun er én vinder. På trods af viden om, at den forventede værdi af et spil er negativ, så udtrykker omkring 20% af stikprøvens respondenter, at de foretrækker spillet med størst usikkerhed for gevinst.

Figur 6 *Grad af risikovillighed*



Note. Graden af risikovillighed er bestemt ud fra spørgsmålet om hvilken type af lotteri (se figur 5), som respondenten foretrækker.

5 Risikovillighed og mulig effekt fra corona-nedlukning – hypotese 2

Tabel 2 viser risikovilligheden for hele stikprøven opdelt på de 6 år baseret på variablen ”Riskoaverse”. En afvisning af hypotese 2 skulle indebære skift i andelen af de mest risikoaverse respondenter og signifikante forskelle i konfidensintervallerne for den gennemførte 1-way Anova.

Tabel 2 *Grad af risikovillighed over tid, 2017-22*

	Andelen af mest risikoaverse	Gennemsnit / 1-way Anova 95% KI
2017	0.39	2.81 / [2.71 ; 2.92]
2018	0.44	2.88 / [2.78 ; 3.00]
2019	0.41	2.83 / [2.74 ; 2.93]
2020	0.44	2.90 / [2.80 ; 2.99]
2021	0.38	2.77 / [2.68 ; 2.86]
2022	0.39	2.77 / [2.66 ; 2.88]
Hele perioden	0.41	2.83

Note. Graden af risikovillighed er bestemt ud fra spørgsmålet om hvilken type af lotteri (se figur 5). Midterste søjle viser andelen af respondenter, som har svaret 4 på spørgsmålet om lotteri (mest risikoavers). Søjlen til højre viser 95% konfidensintervallet for de enkelte år. Sidste række viser gennemsnittet for alle år; det tilsvarende konfidensinterval for variablen er [2.79 ; 2.87].

Det ses af tabel 2, at der ikke er statistisk belæg for at konkludere, at hypotese 2 afvises. Året 2020 er givet et specielt år, fordi evt. corona-effekter ikke har haft tid til at slå igennem, og årene 2021-22 er i et vist omfang kendtegnet ved at andelen af mest risiko-averse respondenter er faldet og at det simple gennemsnit for Risk-averse er faldet til 2.77; men baseret på konfidensintervallerne fra analysen baseret på 1-way Anova er der ikke nogle signifikante afvigelser fra det overordnede gennemsnit på 2.83.

6 Risikovillighed og mulige typer af forskelle - hypotese 3

De mikroøkonomiske antagelser om fuld information, rationel adfærd og tilsvarende forudsætninger er med god grund blevet kritisert for at være urealistiske. Den første hypotese, der er opstillet i afsnit 2, viser, at det helt grundlæggende spørgsmål med at forklare, hvorfor personer spiller. Det er muligt, at adfærd i Arrows 2. kategori (investeringer i aktiver med usikkert afkast) kan forklares ved, at visse personer faktisk har mere viden end gennemsnittet, og på den baggrund er de i stand til at tjene penge på investeringerne; men det er stadig lidt uklart, hvorfor personer deltager i spil ... vel at mærke hvis forudsætningen om at det er den forventede gevinst, der er bestemmende for om der spilles eller ej.

Karlsson et al. (2019) har et friskt bud på, hvorfor nogle personer har et ”risikobetonet liv”; der identificeres via en DNA-kortlægning af mere end 1 mio. personer en række DNA-profiler, der findes ved personer, der er mere risikovillige. Der anvendes flere mål for risikovillighed (fartsyndere, seksualliv, alkohol- og rygevaner), men på trods af anstrengelserne konkluderes det dog, at en den væsentligst årsag til forskelle i syn på risiko må findes ved at kigge på socio-økonomiske forskelle.

Analyserne i denne fremstilling er baseret på forklaring af det opstillede mål for risikovillighed, og i tabel 3 ses resultaterne for de estimerede modeller. Den afhængige variabel er risikovillighed, dvs. positive koefficienter betyder højere risikovillighed, og fordi der ikke kan påvises forskelle i adfærd i perioden, så er årgangseffekter fjernet. Det bemærkes

Ud fra en Arrow-synspunkt er der ikke belæg for at forvente signifikante modeller; men der findes en meget klar kønseffekt, hvor mænd er signifikant mere risikovillige end kvinder. Antallet af observationer er stort og derfor er resultaterne også robuste over for ændringer i specifikation af

modellen. De øvrige effekter i tabel 3 understøtter til en vis grad forventninger fra Karlsson et al. (2019) specielt med en positiv sammenhæng mellem risikovillighed og alkoholforbrug.

Tabel 3 Risikovillighed og personlige karakteristika, regressionsanalyse, 2017-2022

	<i>Fuld model</i>	<i>Reduceret model</i>
Intercept	1.45** (0.27)	1.52** (0.26)
Køn – 1: mand, 0:kvinde	0.30** (0.04)	0.30** (0.04)
Civilstand – 1:enlig, 0:gift	-0.02 (0.05)	-
Alder	0.02* (0.01)	0.02* (0.01)
Alder ²	- 0.004* (0.002)	-
Alkoholforbrug	0.01** (0.00)	0.01** (0.00)
Antal venner	-0.001 (0.001)	-
Syn på karriere	0.06** (0.02)	0.06** (0.02)
Syn på klima og miljø	-0.07** (0.02)	-0.07** (0.02)
Pris-bevidsthed ved køb	-0.09** (0.02)	-0.09** (0.02)
Brand-bevidsthed ved køb	0.08** (0.02)	0.09** (0.02)
Grad af idealisme	0.05 (0.02)	-
R ² -adj	0.044	0.044
Observations	3.231	3.231

Noter. Den estimerede parabel for alder har toppunkt for alder lig 27.4 år.

Stikprøven er dannet ud fra populationen af erhvervsøkonomistuderende ved Aarhus Universitet i perioden 2017-22, og derfor er respondenternes alder tæt på gennemsnittet og spredningen er relativt lille. Der påvises i tabel en signifikant (dog kun på 5% niveau) positiv sammenhæng mellem alder og risikovillighed, og den ikke-lineære version af modellen viser, at risikovilligheden vokser op til godt 27 år, hvorefter den falder – det bemærkes at der er meget få responder i

gruppen af personer over 27 år, så der vil være behov for andre data, hvis dette resultat skal valideres.

Tabel 4 Sandsynlighed for risikovillighed og personlige karakteristika

Kvinde (gennemsnitlig)	0.567	-
Mand (gennemsnitlig)	0.622	+5.5%
Alder +5 år (kvinde)	0.600	+3.3%
Alkohol +5 genstande (kvinde)	0,590	+2.3%
Karrierefokus +1 (kvinde)	0.591	+2.4%
Miljøfokus +1 (kvinde)	0,539	-2.8%
Prisbevidsthed ved køb +1 (kvinde)	0,529	-3.8%
Mærkevarebevidsthed ved køb +1 (kvinde)	0.610	4.3%

Note. Risikovillige personer er defineret som personer, der IKKE foretrækker spillet med flest (små) gevinster. Modellen er estimeret i en logistisk model, og der er benyttet gennemsnitlige værdier for de forklarende variable. Midterste sæje er sandsynligheder for at være risikovillig, og de beregnede stigninger i sandsynligheder (sidste sæje) er målt i forhold til den kvindelige gennemsnitsperson.

Betydningen af de socio-økonomiske personlige forhold for risikovilligheden er vist i tabel 4. Resultaterne er baseret på en logistisk regressionsmodel, og den første sandsynlighed er beregnet for gennemsnitsværdier for alle variable, på nær køn, hvor de 56.7% angiver sandsynligheden for at en kvindelig gennemsnitsperson er villig til at deltage i et af spillene med færre store gevinster. Sidste sæje viser de marginale effekter ved at ændre på de forklarende variable, og der ses at den største effekt er kønseffekten, hvor sandsynligheden stiger med 5.5%, hvis der er tale om en mandlig respondent.

Effekterne ligger alle i intervallet til 2-5%, og det er interessant at finde negative effekter for respondenter, der har særligt fokus på samfundsmæssige forhold. Den fundne negative effekt for prisbevidste forbrugere er måske vanskeligere at fortolke; men det kunne være en effekt af, at denne gruppe af personer har gennemskuet, at den forventede gevinst er negativ for alle spil.

7 Afrunding

Usikkerhed har indflydelse på personers lyst til at investere i usikre aktiver. Arrows bidrag er blandt andet at vise i en mikroøkonomisk model, at i en lang række af tilfælde er det ikke forventet, at en person deltager i spil ol.

Analyserne i dette papir er baseret på godt 3.000 svar fra erhvervsøkonomistuderende i periode 2017-22, og der påvises store forskelle i unges præferencer mht. deltagelse i spil. En ret stor andel (godt 40%) mener er ikke risikovillige og den andel er stabil over hele analyseperioden. Den sidste del af analysen påviste store variationer i risikovilligheden, og specielt er det interessant at der påvises markante forskelle (>5%) mellem mænd og kvinder.

De fundne forskelle er baseret på risikovillighed for hændelser klassificeret af Arrow i gruppen af spil, forsikringer mm. Om forskellen også holder for personers adfærd inden for gruppen af investeringsprodukter er naturligvis en anden sag.

Referencer

- Arrow, K. J. (1951), “Alternative Approaches to the Theory of Choice in Risk-Taking Situations”, *Econometrica*, vol. 19, no. 4: 404-437.
- Arrow, K. J. (1965), “Aspects of the Theory of Risk-Bearing”, Yrjö Jahnsson Lectures, *The Academic Book Store*, Helsinki, 1965.
- Karlsson L., R., Biroli, P., Kong, E. et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature Genetics* 51, 245–257 (2019). <https://doi.org/10.1038/s41588-018-0309-3>
- Keller & Gaciu (2020): *Statistics for Management and Economics*, 2nd ed. Cengage.
- Statens Serum Institut SSI (2019), “Forskere finder sammenhæng mellem gener og risikovillighed”, pressemeldelse om SSI’s bidrag til Karlsson et al (2019), <https://www.ssi.dk/aktuelt/nyheder/2019/forskere-finder-sammenhang-mellem-gener-og-risikovillighed>

Imputering af gennemsnit med regressionsmodeller

- Et mål for kvaliteten af imputeringsmatricer fra et datasæt til et andet

Peter Linde, statistisk konsulent. Peter@Brede.dk

Introduktion

Registeroplysning fra en population kan flettes på en survey (dataindsamling) med en entydig personidentifikation, og på den måde udvide en survey med nye variabler, respondenterne ikke er spurgt om. Med de nye registervariabler kan der gennemføres statistiske analyser. Nøglen til at udvide en survey med registervariabler og danne et nye datasæt er personidentifikationen.

Imputerede gennemsnit fra en survey til et register er den modsatte proces. Registrer måler fx indlæggelse, pensionering eller beskæftigelse, men normalt ikke belastninger i arbejdslivet, adfærd, helbred eller personlige forhold. Da en survey normalt har en lille stikprøve i forhold til populationen i registeret, blæses stikprøven på denne måde op. Spørgsmålet er, hvordan man kan imputere fx belastninger i arbejdslivet målt ved en survey til et register. Man har her ikke en personidentifikation, der kan bruges til at flette stikprøvens svar på alle personerne i registeret. Registeret indeholder måske stikprøven til surveyen, men også alle andre i populationen. Eksponeringsmatricer er gennemsnit baseret på en regressionsmodel, typisk med et begrænset antal hovedvirkninger. Selv med et begrænset antal hovedvirkninger kan der være mange kombinationer. Med fx 247 jobfunktioner, to køn og 47 aldersgrupper giver det 23.218 kombinationer i eksponeringsmatricen, fx mht. graden af stress.

Hvis man har valget mellem at imputere fra en stikprøvebaseret survey til et fuldt dækkende register eller flette registeroplysninger på den stikprøvebaserede survey og analysere med vægte, der opregner til populationen, er det sidste at foretrække. Denne analyse vil ikke have modelbaseret bias fra imputeringen, da datasættet indeholder den fulde kompleksitet af respondenternes svar uden restriktioner. Denne artikel er derfor mest relevant, hvis man ikke har dette valg mellem at imputere eller flette datasæt.

Der findes også andre imputeringsmetoder, der inddrager flere faktorer og bruger multiple imputeringer med de direkte svar fra de personer i surveyen til de individer i registeret, der minder om personen i surveyen, og som kan estimere usikkerheden ved im-

puteringen vha. de gentagende imputeringer. Der er derfor en række analytiske fordele med denne tilgang.

Når der er tale om en modelbaseret estimation/imputering er der en række antagelser. Hvis man fx vil imputere et stress-mål, betyder regressionsmodellen med fx job, alder og køn som hovedvirkninger mindst to ting. Det antages at de tre faktorer virker som hovedvirkninger. Dvs. at effekten af fx alder er den samme for alle jobgrupper for begge køn, eller at forskellen mellem mænd og kvinders stress er den samme i alle jobfunktioner med samme alder. Hvis der for nogle jobgrupper faktisk ikke er forskel mellem mænd og kvinder, får disse jobgrupper alligevel imputeret middelforskellen på mænd og kvinder. Og modsat hvis der er stor forskel mellem mænd og kvinder. En anden betydning af regressionsmodellen er, at alle andre faktorer lades ude af betragtning og neutraliseres – givet faktorerne i modellen. Så hvis der en signifikant betydning af fx arbejdes karakter, helbred eller arbejdsforhold i surveydatasættet vil imputeringen i registeret betyde, at disse forskelle helt forsvinder givet job, køn og alder. Og at disse signifikante faktorer kun virker indirekte gennem deres korrelation med faktorerne i modellen, når køn, alder og jobtype ikke indgår i tabellerne. Det siger derfor sig selv, at masse-imputering med eksponeringsmatricer fra en survey til et register af fx stress ikke kan bruges til kausale statistiske analyser udover imputeringsmodellen. Ellers havde man opfundet en metode til vha. én survey, til at kunne imputere svar, man ikke har spurt om til uendeligt mange datasæt – en form for gratis evighedsmaskine til generering af nye datasæt. Men kan imputeringen bruges som en indikator eller en form for instrumentvariabel? Det er fokus i denne artikel. Det giver værdi at udnytte data optimalt, men det forudsætter man deler viden om imputeringens kvalitet, da der ellers drages forkerte konklusioner.

Kvalitetsmål

Nedenfor er en række kvalitetsmål for en række kvalitetsdimensioner i regressionsimputeringen, der alle kan måles på en procentskala fra 0 til 100%. 0% betyder kvaliteten af delmålet er totalt fraværende, og 100% at delmålet har maksimal kvalitet. Ingen imputering er bedre end sit svageste led, og en pragmatisk tilgang for at danne et fælles kvalitetsmål er, at gange de seks foreslæde kvalitetsmål med hinanden.

1) Graden af forklaret variation

Regresionsmodellen med fx 3 hovedvirkninger har en grad af forklaret variation mht. variablen, der estimeres, fx stress. Graden af forklaret variation er helt essentielt ved imputeringen. En imputering med en forklaret variation på 60%, har helt grundlæggende en anden kvalitet end en på 30%. Et umiddelbart ønske til denne kvalitetsdimension kunne være mindst 50%. Hvis graden af forklaret variation er tæt på 100%, betyder det umiddelbart, at de brugte baggrundsfaktorer og modellen alene forklarer nærmest alt, og at alle andre faktorers betydningen skyldes deres korrelation med baggrundsfaktorerne ved imputeringen. Det er dog kun umiddelbart, da der kan være konfunderende virkninger.

2) Opnåelsen i surveyen der imputeres fra

Høj opnåelse er udtryk for at man har fået mange af de udvalgte motiveret til at svare. En høj opnåelse er typisk udtryk for, man har fået svar fra flere, der er i kanten af arbejdsmarkedet, lever mere alene eller har udfordringer med helbredet. De ressourcestærke respondenter har man normalt lettere ved at få til at svare. En lav opnåelse tegner derfor normalt et for rosenrødt billede af virkeligheden.

3) Oplyste værdier i imputeringen

Hvis køn og alder bruges vil det være to variabler, der næsten altid er oplyst. Men der kan være udfordringer med responsvariablen, fx stress, eller jobgruppen, hvis den bruges. Jobgruppen oplyses af arbejdsgiveren ved indberetning af løn og har ofte en del uoplyste værdier. Oplysningen af jobgruppen har i mange tilfælde ikke i sig selv nogen betydning for lønnen, der indberettes, og valideres ikke på samme måde som lønnen. Både lønmodtageren og arbejdsgiveren kender lønnen, men lønmodtageren kender ikke nødvendigvis den jobgruppe arbejdsgiveren har indberettet. Uoplyste jobkoder vil selvfølgelig gøre imputeringen fra et datasæt (survey) til et andet datasæt (registeret) mere usikker. Generelt har andelen af uoplyste jobkoder været stigende, og kan for nogle brancher nærme sig 10%.

4) Oplyste kombinationer

I registeret vil der være kombinationer i registeret, der ikke findes i den survey, der er imputeres fra. Specielt hvis der er tale om en lille stikprøve med stor usikkerhed.

Her kan der ikke laves modelkontrol. Stikprøvestørrelsen i surveyen i forhold til størrelsesordenen af registeret har betydning for analysen. Andelen af oplyste kombinationer i registeret, der også findes i surveyen, er en indikator herfor. Stikprøvestørrelsen i surveyen bør inddrages i analysen som en varianskomponent. Det behandles senere.

5) Designet af stikprøven

Mange stikprøver har vægte pga. bortfald og stratificering. Der kan estimeres en designeffekt, der måler stikprøven i forhold den klassiske stikprøve med simpel tilfældig udvælgelse, som de fleste statistiske analyser antager. Hvis designeffekten fx er 1,1 betyder det, at stikprøven har en varians, der er 10% større end variansen ved en simpel tilfældig stikprøve. Ved at tage den inverse af designeffekten kan måle betydningen af designet. Hvis designeffekter fx er 1,1, er kvalitetsmålet herfor $1/1,1=91\%$. Hvis stikprøven er simpel tilfældig uden vægte, er designeffekten 1, og dette delkvalitetsmål er 100%.

6) Berkson biasen

Imputeringen af gennemsnit kan give en Berkson bias ved analyser i registeret, fordi det er middelværdien og ikke det oprindelige udfaldsrum, der bruges til imputeringen. Hvis udfaldsrummet i surveyen er, om man har stress eller ej, kodet som en 0/1 variabel, vil imputeringen blive et kontinuert tal mellem 0 og 1 svarende til sandsynligheden for at blive stresset i den pågældende kombination af køn, alder og job. Dette påvirker en OLS regression. Hvis der er lige mange observationer i alle kombinationer af baggrundsfaktorerne, har estimationen af hældningen i en OLS regression begrænset bias. Datasættet er da ortogonalt mht. alle baggrundsfaktorer. Det er dog en ret vidtgående antagelse, da der sjældent er ens fordeling af fx køn i alle jobgrupper, og nogle jobgrupper er meget store i forhold til andre. Forudsætningen om lige mange observationer kan bruges til et kvalitetsmål af Berkson biasen. Man kan selvfølgelig også prøve at regne den direkte ud. En indikator for Berkson biasen får man, hvis man med den konkret stikprøvestørrelse beregner et idealkrav til hvor mange observationer, der skal være i hver gruppe (kombination). Forholdet mellem idealkravet og den faktiske stikprøve sættes i anden potens i hver gruppe. Der summes over alle kombinationer og der beregnes et gennemsnit af dette

kvadrattal. Dette gennemsnit tages der derefter kvartronen af. Det vil altid være større end 1, og 1 er udtryk for kravet om lige mange observationer i alle kombinationer er opfyldt. Det inverse heraf kan bruges som kvalitetsmål. Hvis kvadratgennemsnittet fx er 2, er kvadratroden heraf $\sqrt{2} = 1,41$ og kvalitetsmålet $1/\sqrt{2} = 0,707 = 70,7\%$.

Det samlede kvalitetsmål

Imputeringen kan bruges som en indikator for det, der ikke er spurgt om, og man derfor ikke ved noget reel om i registeret. Man skal ikke forvente høje kvalitetsmål ved imputeringen med regressionsestimater, hvilket også reflekterer de usikkerheder, der er ved imputeringen.

Den forklarede variation kan fx være 80%, delmål 2 om opnåelsen 60%, delmål 3 om oplyste faktorer 95%, delmål 4 om oplyste kombinationer 90%, delmål 5 om designefekter fx 91% og delmål 6 om Berkson biasen 71%. Ganget med hinanden giver det et samlet kvalitetsmål på 27%. Man kan altid diskutere hvor højt et kvalitetsmål, man vil bruge, før man bruger imputeringen til artikler om mulige tendenser, men 25% virker ikke unaturligt. Delmål 1 kan forøges med en bedre regressionsmodel, bedre imputeringsmetoder der inddrager flere forhold eller Machine Learning (ML) metoder. Det er klart, at en model med 3 hovedvirkninger er ret simpel, og derfor kan have en ret lav grad af forklaret variation og har svært ved at holde i praksis i alle undergrupper. Bedre imputeringsmetoder vil et stykke af vejen kunne rette herpå. Hvis det er en 0/1 variabel der imputeres, kan det imputerede gennemsnit bruges til at danne 0/1 variabler. Hvis gennemsnittet fx er 0,6 kan man danne 4 observationer med udfaldet 0 og 6 med udfaldet 1, og give hver af disse 10 observationer vægten 0,1. Det vil et langt stykke af vejen håndtere Berkson biasen, hvis man bruge de programmer i fx SAS (surveyprocedurerne) og STATA (svy), der analyserer korrekt med vægte.

Forudsætninger og dokumentation

Man kan mod ovenstående forsøg på at danne et kvalitetsindeks indeks for imputeringen indvende, at der ses bort fra helt grundlæggende forhold, der skal være opfyldt. For det første om de to datasæt overhovedet kan sammenlignes, og der er metodisk og empirisk er belæg herfor. Det må dokumenteres konkret. For det andet om klassifice-

ringen af fx jobgrupper er ens i de to datasæt. Hvis koden om job kommer fra den samme nationale registrering af DISCO-08, er der grund til at antagelsen holder. Men hvis det er to forskellige lande eller registerdatasættet er baseret på DISCO-08 og surveyen er baseret på egen registrering, er der udfordringer med antagelsen. Egen rapporteringen af job kan godt være af høj kvalitet, men det er afgørende for imputeringen, at der er brugt samme metode, og den er den samme som registerets. Et andet forhold, der skal være opfyldt er vægtningen af imputeringen. Hvis stikprøvedesignets repræsentativitet er baseret på vægte fra fx stratifikation og skævt bortfald, skal disse vægte bruges til imputeringen, ligesom de skal i analyser i det oprindelige datasæt. Endeligt er der spørgsmålet om stikprøvestørrelsen i den oprindelige survey. Hvis der laves signifikans vurderinger i det store register, skal der inddrages en varianskomponent fra surveyen, der reflekterer stikprøvestørrelsen her. Populationsstørrelsen i registeret N og antal opnåede svar i surveyen n kan bruges som en indikator for varians-forøgelsen, fx $1/(1+N/n)$.

Helt grundlæggende kan masseimputeringen med en simpel regressionsmodel ikke bruges til statistiske analyser af kausale sammenhænge i registeret, men kun som en indikator. Og i det lys kan en simpel indikator, som ovenfor, være et brugbart mål for kvaliteten af indikatoren.

Kilder

Der er del artikler og eksempler, hvis man søger på nettet. Flere peger på, at flere analyser er nødvendige. Her en oversigtartikel fra 2022.

https://www.researchgate.net/publication/339933395_Developing_a_Job_Exposure_Matrix_of_Work_Organization_Hazards_in_the_United_States_A_Review_on_Methodological_Issues_and_Research_Protocol

https://en.wikipedia.org/wiki/Job-exposure_matrix

Google searches linked to Apple stock volatility

Niels Buus Lassen^a

^a Centre for Business Data Analytics, Copenhagen Business School, Denmark

ARTICLE INFO

Keywords: Investor behavior, Google searches, Stock markets, Investor sophistication, Decision making

ABSTRACT

The recent studies on social media that link news data to volatility show a Twitter buzz up is typically linked to higher volatility, while a general news media buzz is linked to lower volatility in the following month. This article demonstrates that Google searches influence Apple stock volatility in either on a weekly basis by analyzing the behavior of private and professional investors in relation to Google searches and how this behavior links to Apple stock volatility. To this end, this study employs the logic of sales modeling and, thus, contributes to the theoretical construction of the novel “investor journey model” by mapping Google searches onto investor behavior, which is an under-researched field in the literature. Subsequently, the paper summarizes the main findings in this field and outlines future challenges in this research.

1. Introduction

Twitter data have been included in several models and shown to have predictive power for both stock price indexes and specific stock price movements (see, e.g., Bollen & Mao 2010; Jiao, Veiga, & Walther 2016; Li, van Dalen, & van Rees 2018.).

A logical explanation for the predictive power of social media data in terms of financial market behavior is the size of the big data from social media chats and also Google searches about the respective stocks. Stocks with large amount of social media data, are popular stocks people like to talk about on social media and also do Google searches about.

Text mining can identify patterns in the big data from the social media and Google searches. Statistical and machine learning methods can be tested to model the human behavior on social media and Google searches to financial market behaviors.

This approach is comparable to the sales of products and services, where big data from social media can be used to predict sales. The sales models that build on social media data work well if social media data are big enough or, in other words, if customers like to talk about a product or service on social media. Examples are H&M, Nike, and Apple (see, e.g., Lassen et al. 2017; Boldt et al. 2016). These are popular social media topics that produce big enough data to have predictive power for their sales.

All chat text on social media about products and services can be categorized into one of the phases in the Customer Infinity Model (Figure 1). These phases of customer behavior can then be modeled based on sales and provide logical explanation for why social media data can predict sales, if the data are big enough (see, e.g., Asur & Huberman 2010; Lassen et al. 2014).

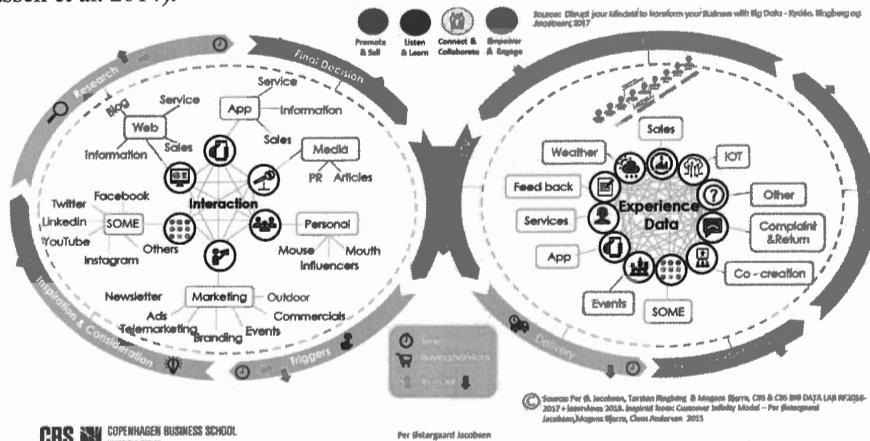


Fig. 1 Customer Infinity Model

Source: Per Ø. Jacobsen, Torsten Ringberg, & Mogens Bjerre, CBS & CBS BIG DATA LAB RF 2016-2017 + Interviews 2018. Inspired from: Customer Infinity Model – Per Østergaard Jacobsen, Mogens Bjerre, Claus Andersen 2015

The logic for stock prices is comparable, based on which this article proposed the “investor journey model,” which posits that all social media text or web searches can be categorized into one of the phases of this model, which can in turn be linked to Apple stock volatility, as well as other stocks with big enough data on social media and from web searches. This paper focuses on both the private and professional investor behavior related to Google searches.

More than 90% of all global web searches use Google; specifically, approximately 63% on the search engine, 23% on Google Images, 4% on YouTube, and 1% on Google Maps (BusinessInsider.com 2018). When amateur and professional investors search for information about the stocks they are interested in, Google searches provide relevant stock information and form the majority of all web searches. The model proposed in this article has been tested on more than 60 Google searches before identifying good proxies for the amateur and professional investor behavior based on Google searches for Apple stocks.

The research questions this article tackles as follows:

RQ1: *Can Google search data predict stock price volatility?*

RQ2: *Which Google searches are creating ups and downs in Apple stock price volatility?*

RQ3: *Can the identified ups and downs in Apple stock price volatility be linked to household and professional investor activity?*

The contribution of this article is identifying new patterns for investor behavior on web searches, and how these patterns link to ups and downs instock volatility theoretically explained through the proposed investor journey model, which relies on insights from related queriesand topics in Google Trends from more than 60 Apple-related Google searches. Apple is the example used in this article, and model may be applicable for similar big tech stocks with high volume google search data.

2. Literature review

One of the most notable articles modeling stock price volatility using Google Trends, is Preis et al. (2013), which mentions: “We suggest that *Google Trends* data and stock market data may reflect two subsequent stages in the decision making process of investors.”

This article suggests that Apple stock-related Google Trends data follow an investor decision making journey, which affects Apple stock volatility. The proposed investor journey model is detailed in section 3.

Jiao, Veiga, and Walther (2016) find that a buzz up in coverage by traditional news media predicts subsequent decreases in volatility and turnover, while a buzz up in coverage by Twitter predicts increases in the subsequent volatility and turnover. However, they do not explain why the buzz in traditional newsmedia and Twitter coverage have these different effects on stock volatility.

Greenwich Associates published a report in 2015 based on interviews with 256 asset owners from more than 250 institutional investor organizations, which shows that institutional investors use Twitter in a very limited manner in their decision-making process, compared to LinkedIn, for example. Among the 256 interviewed institutional investors, LinkedIn was used by more than half and often played an important role in investor decision making. The interviewed institutional investors recognized the value of the Twitter news feed in seeking opinions or commentary on market events, but considered LinkedIn feeds to be better targeted, as they reflected their professional ties more closely.

This article does not include LinkedIn data, but only Apple-related Googlesearches due to their free availability through Google Trends. Specifically, LinkedIn data are difficult to access and expensive. Twitter data were also not considered due to their cost for this article.

Institutional and professional investors use information processed from Twitter by analytical companies such as Dataminr. For example, TheGlobeAndMail.com (2018) states: “Dr. Mohanram cautions that individual investors are not likely to be able to correctly replicate the conditions of the study to benefit from crowd-sourced opinion. It requires a certain amount of data crunching ability and sophistication [to] analyze a large sample of tweets, categorize them and aggregate them all in real time.”

Private investors are using more Twitter unprocessed information in their decision-making processes such as by following investor gurus and searching for stock related information on Twitter. For example, MarketWatch.com’s (2018) “Finance Twitter: The 50 most importantpeople for investors to follow” is one of many articles recommending private investors who to follow on Twitter to get smart insights on investing. Based on such recommendations, it becomes logical that the abundance of financial gurus on

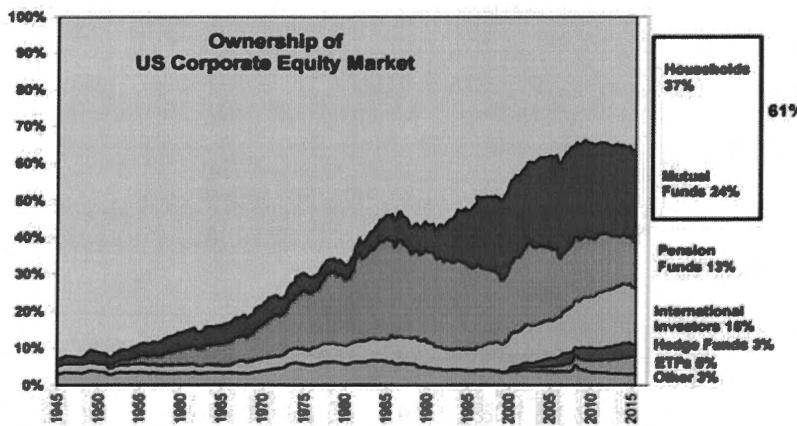
Twitter is creating noise and increases stock volatility. Searching for stock related information on Twitter typically yields several credible news sources along with sources that require vetting. In short, the many news sources on Twitter create noise and increase stock volatility. Namely, private investors are more exposed to the risk of rumors, old news being perceived as new news, speculations, and manipulating info to drive stock prices up or down. All this noise can lead to irrational investor behavior by creating higher volatility when there is a buzz up in Twitter info for a stock.

Therefore, the logic of Twitter usage is that private investors are the main actors in the increased stock volatility after a buzz up in the Twitter info about a stock.

Traditional news media info about stocks are typically used by professional and institutional investors. Examples of reliable traditional news media channels are Financial Times, Bloomberg, Wall Street Journal, Reuters, CNBC, Forbes, and MarketWatch. They are associated with rational investor behavior and lower volatility after a news buzz up related to a specific stock.

For the Apple stock, around 60–70% of Apple stock trading is conducted by professional and institutional investors. For details, please refer to

<https://finance.yahoo.com/quote/AAPL/holders/>.



Source: Federal Reserve and Goldman Sachs Global Investment Research.

Fig. 2 Ownership of the US corporate equity market: USD 36 trillion as of Q2 2016 and includes USD 7 trillion of foreign equity holdings.

Based on Fig. 2, US households own approximately one third of the US corporate equity market, meaning that the household investor activity in Apple stocks is estimated to be one third on average.

Many models for stock investor decision making are based on the prospect theory developed by Kahneman and Tversky (1979), which is a descriptive model of decision making under risk (see, e.g., Wakker & Levy 2015). Prospect theory has been criticized and several alternatives have been suggested (see, e.g., Nwogugu 2005; Levy et al. 2002).

Prospect theory assumes decision making has two phases: editing and valuation. The editing phase refers to investors scanning the information and forming their beliefs on eventual outcomes. This phase is heavily dependent on psychological biases. The second phase is the valuation phase, where agents value these eventual outcomes based on the beliefs in the first phase. This phase depends more on the risk preferences of investors. These two decision-making phases are distinct, albeit closely dependent on each other.

The model for stock investor decision making developed in this article is based on the empirical observation of Google searches related to the Apple stock (stock symbol: AAPL). The model does contain some editing and valuation by investors but focuses on explaining how different types of investor use information before, during, and after the Apple Quarterly Reports and iPhone models are released.

3. The Investor Journey Model

The “investor journey model” is a conceptual model developed in this study for the analysis of investor behavior based on web searches and is applied to Apple stocks in this paper. Specifically, it has been developed based on related queries and topics on Google Trends from more than 60 Apple-related Google searches. It is thus the main analysis framework in this study, as described in Figs. 3 and 4.

	Household Investors	Professional & Institutional Investors
Phase 1	Would I like to buy or sell the Apple stock? Or should I buy or sell Google, Amazon, Facebook, Microsoft, Nvidia or IBM instead?	Should I add or decrease Apple stocks in my portfolio? The Apple stock is here in a setup of portfolio diversification & risk profiles.
Phase 2	Which stock or stocks do I proceed with? Which rumor news will affect this?	What is my analysis of Apple sales & results, before the Apple Quarterly Report? Any reliable Apple rumor news indicating level of sales before the Apple Quarterly Report?
Every end of Phase 3a Jan, Apr, July & Oct	Apple Quarterly Report release. Above or under my expectations? Time to buy or sell Apple stock?	Apple Quarterly Report release. Above or under my analysis? Should I add or decrease Apple stocks in my portfolio?
Every Sep. Phase 3b	Would I like to buy or sell the Apple stock, before the Sep launch of new iPhone? Will the iPhone launch drive stock up or down?	What is my analysis of iPhone launch, before the Sep launch of new iPhone? What is the level of rumor news compared to earlier years? Will the iPhone launch drive stock up or down?

Fig. 3 The Investor Journey

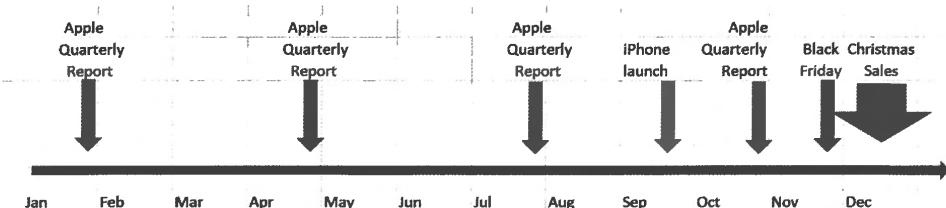


Fig. 4 Timeline for the investor journey

I divide the Apple stock-related Google searches into two groups—household and professional investors—based on the following hypotheses.

Hypothesis 1: The Google searches for stock symbols AAPL, AMZN, and IBM are assumed to be mostly made by household investors, as professional investors are likely familiar these stock symbols and do not need to Google search them. Google searches for these stock symbols lead to SeekingAlpha.com and StockTwits.com, which have large groups of household investors among their readers.

Hypothesis 2: The Google searches for Apple rumor news are assumed to be largely done by professional investors, as the Apple rumor news sites use high vetting levels and due diligence for the sources.

4. Methodology

The analysis is based on the multiple regression modeling of more than 60 Apple-related Google searches as input variables and investor behavior in the form of Apple stock volatility as the dependent variable.

For the Apple and rumor news Google searches, professional investors largely consider the vetted news and rumor sites, while household investors largely Google the stock symbols of Apple and other big tech companies. These patterns were identified based on the related topics and queries for each Google search, but also on the assumption of the tendency that household investors drive volatility up and professional investors drive it down. This argumentation of the patterns for household and professional investors will be further elaborated upon in the following sections.

The statistical software used is Oxmetrics 8.10, which mainly focuses on using the Autometrics functionality. Autometrics is a part of the PcGive module of Oxmetrics, being the automatic econometric model selection procedure that is available in PcGive. It is based on regression modeling under the general unrestricted model (GUM) framework. In Autometrics, the variable and model selection criteria are based upon the unique method developed by David Hendry and Jürgen Doornik, which performs well on gauge and potency. Gauge is the retention rate of irrelevant variables in the selected model and akin to size, because it accounts for the wrongly selected variables. Potency is the retention rate of the relevant variables in the selected model. It is also akin to size because it accounts for the variables that have been correctly selected (see Hendry et al. 2014).

The chosen target size for the dataset used in this study is 1%, which is the t-probability threshold for choosing and eliminating input variables. The 1% level was chosen because of the 60+ Google searches selected as input variables in 1–4 time lags each; therefore, the large amount of input variables could be cut down to a reasonable number. After selecting among more than 60 Google searches with a target size of 1%, a target size of 5% was also tested when the input variables were reduced to a group of 10 predictors.

Another variable selection method that comes from machine learning is the least absolute shrinkage and selection operator (LASSO) method. It is a type of linear regression that uses shrinkage, where data values are shrunk towards a central point, such as the mean. LASSO adds the “absolute value of magnitude” of the coefficient as penalty term to the loss function, which will be minimized. This is called L1 regularization.

This regression works especially well for many input variables and multicollinearity and

can limit the input variables significantly. The input variables field is cut down by the described L1 regularization.

The analysis will examine if Apple-related Google searches are good proxies for investor behavior regarding the Apple stock. Both household and professional investor are reflected in Google searches on the investorjourney to buy or sell Apple stocks.

5. Data

5.1 Google searches during 2015–2020

The Google search data were collected from Google Trends (<https://trends.google.com/trends/>). Specifically weekly Google search data were collected from April 2015 to April 2020, as this is the longest period available on Google Trends with weekly Google search data.

It is possible to get obtain Google searches on Google Trends, but they are only available up to the 90 prior days. Weekly Google search data are available on Google Trends for the past 12 months or 5 years. For longer periods, starting from 2004, monthly date are available. The selected data were evaluated to be the most suitable dataset for modeling Applestock prices and volatility because it was the longest and most recent time period available with weekly data at the time of this study.

The Google search data are given in indexes from 0 to 100, as positive integers, available on Google Trends. An index 100 for one or more weeks would be the highest weekly search volume for the entire 5-year period. In this article, more than 60 Apple-related Google searches were extracted for the 5-year period and tested for their relationship to stock price and volatility. These searches were found by exploring Google Trends for the Apple stock symbol (AAPL) and products such as iPhone, iPad, MacBook iOS, or MacOS. Google Trends includes both related topics and queries, based on which I found stocks related to the Apple stock, which led to the idea of developing the investor journey model. Fig. 5 shows one data extract.

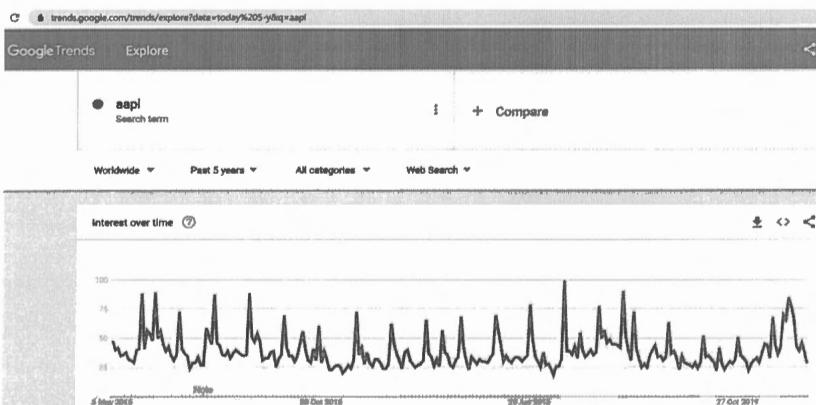


Fig. 5 Google search data for “AAPL” for the period 2015–2020

<https://trends.google.com/trends/explore?date=today%205-y&q=AAPL>

The Google searches can be extracted in sets of up to five, but I instead considered one search at a time, as for more than five searches extracting sets will create problems. That is, because the five searches will be indexed in a group of five, a second set will not be indexed against the first set. As such, unless there is an overall baseline, the highest of all Google searches will be identified for each set. However, the baseline can change during the research to include all the dataset. As such, the most practical approach is to extract Google searches one at a time and, in the end, the searches can be indexed together in one set of five or two sets of 10 for 5–10 input variables.

For example, for the Google search of the Apple stock symbol (AAPL) shown in Fig. 5, all weeks in the 5-year period are indexed around the datapoint with index 100 in week 31 of 2018. For up to 5 searches in the same Google Trends query, all weeks would be indexed around the highest search with index 100 in a given week. The final dataset from Google Trends included sets of five Google searches and all searches were indexed to the highest search index, “IBM,” in the 5-year period.

Variables	N	Mean	Median	St. dev.	Min	Max
AvgWeeklyClose	260 weeks	164.5	158.1	54.3	91.9	323.6
Weekly Volume, number of million shares	260 weeks	170.8	156.4	72.7	32.5	500.4
Weekly volatility	260 weeks	3.20%	2.64%	2.24%	0.51%	19.17%
First diff Log(avg- Wee kClose)	260 weeks	0.13%	0.24%	1.69%	-13.46%	12.76%

Table 1 List of weekly financial variables

Table 1 shows that the weekly volatility includes an extreme outlier of 19.17%. This was due to the COVID-19 pandemic from February 24 to April 12, 2020, where the weekly volatility ranged between 9% and 19.2%, peaking at 19.2% in week 11—March 9–15, 2020. In the same time window, the average trading volume was 300 million shares per week and stock price varied from USD 212 to USD 304.

Table 2 presents the descriptive statistics for the Google search data. As previously mentioned, all Google searches for 5 years have been downloaded from Google Trends as 260 weekly observations. Specifically, 62 Apple related Google searches were tested for modeling the Apple stock volatility, among which nine were significant and were chosen for further modeling. The two most important input variables, the Google searches for “AAPL” and “AMZN”, selected by SPSS LASSO and Autometrics are marked with **bold**. The Google searches for “MacRumors” and “Apple rumors” are also marked with **bold**, as they were additionally selected by Autometrics when the target size was changed from 1% to 5%. The target size in Autometrics is the t-probability threshold for choosing and eliminating input variables. There are two groups for the searches:

1. News and rumors searching/vetting for Apple stock
2. Apple and other related big tech stock searches

The Google search “Apple rumors” marked with green is the only variable considered with an overweight of professional investors and negative coefficient. All other Google searches are considered to have an overweight of household investors.

Google searches	N	Tested time lags	Mean	Median	St. dev.	Min	Max
First section, rumor, and news searches							
Apple Rumors	260 weeks	1–4 weeks	0.9	1.0	0.4	0.5	3.0
9to5mac	260 weeks	1–4 weeks	1.2	1.0	0.6	0.5	5.0
TheVerge	260 weeks	1–4 weeks	1.0	1.0	0.2	0.5	2.0
MacRumors	260 weeks	1–4 weeks	2.7	2.0	1.5	1.0	12.0
AppleInsider	260 weeks	1–4 weeks	0.7	0.5	0.3	0.5	2.0
Second section, Apple, and related big tech stock searches							
AAPL	260 weeks	1–4 weeks	20.2	18.0	7.6	9.0	51.0
AMZN	260 weeks	1–4 weeks	14.9	13.0	9.7	2.0	48.0
IBM	260 weeks	1–4 weeks	72.2	72.0	10.9	38.0	100.0

Table 2 Descriptive statistics for the Google searches

From Table 2, the search for “IBM” has the highest index number 100, all other Google searches being indexed to it. Had this search not been included, the search for AAPL would have been the main index, meaning all other Google searches would have been indexed after it, since it has the second highest index of 51. Excluding the Google search on IBM, would also have increased the index numbers for Google searches on MacRumors, Apple rumors, and AMZN, which could have changed their significance.

In Autometrics, the estimation sample cannot start before week 19 in 2015 because of the tested time lags for 1–4 weeks in the dataset. Therefore, the Autometrics estimation is conducted from week 19 in 2015 to week 42 in 201, covering 88% of the dataset. The last 26 weeks of the dataset from week 43 in 2019 to week 16 in 2020 are chosen as hold-out data for the forecast evaluation.

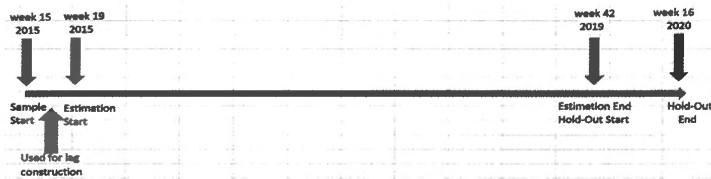


Fig. 6 Timeline for the dataset and visualization of the train/test split.

Fig. 7 shows time plots of the two tested dependent variables, namely the weekly volatility and first difference log(avg close), which is the stock price return. It also shows the time plots of the two most significant regressors, the Google searches for AAPL and AMZN and the highest Google search index in the 5-year period for IBM.

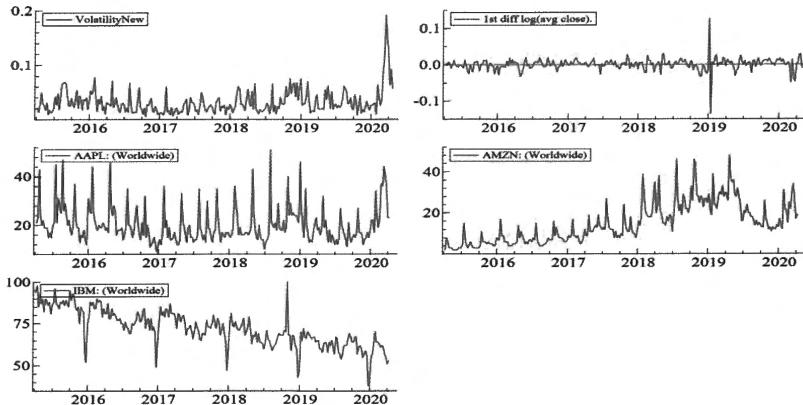


Fig. 7 Timeplots of the main variables

The selected Google searches developed as follows:

AAPL: Quarterly peaks around the ends of January, April, July, and October in each year, when Apple is releasing its quarterly reports to Nasdaq and on their investor site.

AMZN: Quarterly peaks around the ends of January, April, July, and October in each year, when Amazon is releasing its quarterly reports to Nasdaq and on their investor site.

IBM: Quarterly down peaks around the end of December. An explanation could be that IBM is not linked to the Christmas season, and the search for more Christmas-linked stocks overtakes the December searches. The downward pattern in the Google searches follows the decline in stock prices during 2015–2020. The main role of the Google searches for “IBM” is that it is the highest Google search index in the 5-year period for 10 out of 60 Google searches that were most significant for the weekly Apple stock volatility. Therefore, all last 10 Google searches in the last dataset tested are indexed after the IBM searches. The declining trend for IBM searches during 2015–2020 follows the declining trend in both IBM stock price and IBM’s position in machine learning and AI, where IBM is not among the leaders.

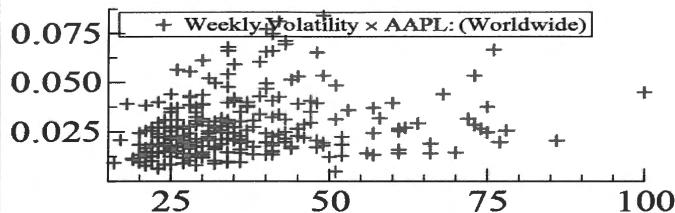


Fig. 8 x-y plot of weekly volatility x AAPL

Y axis is Weekly Volatility and X axis is AAPL Google searches. The x-y plot shows patterns of higher volatility for higher AAPL Google search indexes.

5.2 Apple announcement data

Apple releases its quarterly reports at the end of the month after a quarter closes, that is, approximately 1 month after the quarter closes, to stock holders and the media through the investor portal at Apple.com, <https://investor.apple.com/investor-relations/default.aspx>

There was a test of event variables for the iPhone launch in September; the quarterly Reports every end of January, April, July, October; and the Black Friday and Christmas sales. These event variables did not show any effect on either weekly volatility or stock price returns. These event variables were defined as 0/1 variables, taking 1 in the week the event occurred, and 0 otherwise.

There are patterns in some Apple-related Google searches before, during, and after the quarterly reports, which are considered as quarterly regular spikes in these Google searches. Refer to Fig. 7 for an example of these patterns.

Apart from the quarterly reports from Apple, the single most important event for Apple is the yearly iPhone launch in September. At the iPhone launch 2015, the weekly Apple stock volatility increased, but from 2016 onwards the iPhone launch has been a mean event based on weekly volatility. That also tells a story of the iPhone hype wearing off. The iPhone hype topped in 2012, and has been decreasing since then. Refer to the Google searches for iPhone from 2007-2021:

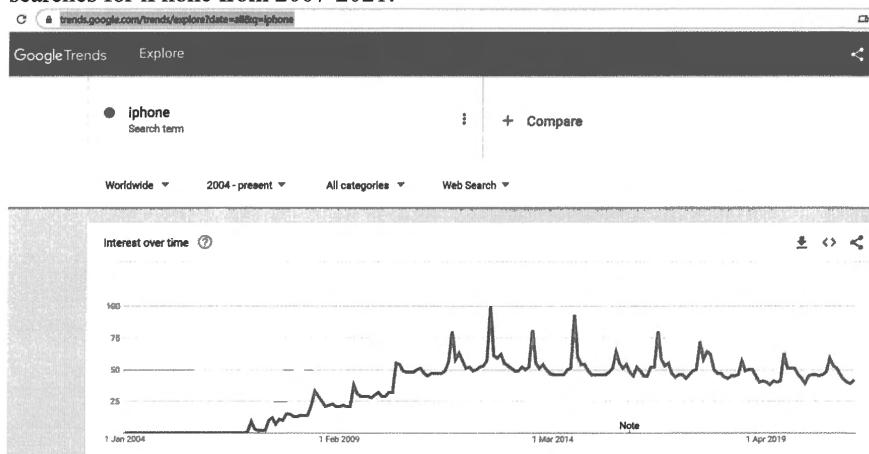


Fig. 8, iPhone Google searches, 2007-2021. <https://trends.google.com/trends/explore?date=all&q=iphone>

The first iPhone was launched in June 29, 2007 and, from 2012, the main iPhone launch has always been in September. There are patterns in some Apple-related Google searches before, during, and after the September iPhone launch, which are seen as yearly peaks in these searches. Refer to Fig. 7 for an example of these patterns.

There are also some announcement data on other Apple products, but the most important product announcement for Apple and Apple stock is the yearly iPhone launch.

5.3 Weekly return data

Nasdaq.com provides free data download for their listed stocks going back 10 years at <https://www.nasdaq.com/symbol/aapl/historical>.

These data provide daily info for the selected period on Apple stock price, namely open, high, low, close, and volume. These daily data from Nasdaq.com were the basis for calculating weekly Apple stock price volatility, based on formula (5).

These daily data were also used to calculate the average weekly stock price and weekly volatility. The weekly average stock price was calculated based on the average of all the daily closing prices. Daily closing prices were also used to calculate the daily changes, forming the basis for calculating the weekly volatility. As the Nasdaq data only have daily close, high, low values for stock prices—with no daily average—it was chosen to use the close for the weekly calculations of the financial variables. These weekly financial variables were needed in the model together with the weekly Google searches.

5.4 Formulas

P_t is the weekly average stock price in week t, calculated as:

$$(1) P_t = (P_{1t} + P_{2t} + P_{3t} + P_{4t} + P_{5t})/5,$$

where P_{it} is the closing stock price for Apple on day i in week t.

Approximately 80% of the trading weeks, have 5 trading days—Monday to Friday—based on the above formula.

The remaining 20% of the trading weeks have 4 trading days, the formula being:

$$(2) P_t = (P_{1t} + P_{2t} + P_{3t} + P_{4t})/4.$$

Very few trading weeks have 3 trading days, with the following formula:

$$(3) P_t = (P_{1t} + P_{2t} + P_{3t})/3.$$

The stock price return in this article is expressed as a first difference log variable, in line with the ARCH models. The use of stock price log returns has advantages over the arithmetic return (see, e.g., Hudson & Gregoriou 2010). The first difference stock price return is expressed as:

$$(4) \Delta \log(P_t) = \log(P_t) - \log(P_{t-1}).$$

The best explanatory variable for the weekly stock price return is the weekly stock price return from the previous week and two weeks before. Prior to volatility, I also investigated modeling stock price return. The results are available upon request. The

historical stock price return only creates a R^2 of 10% for the training data and 6% for the hold-out data. With the adding of the best three Google searches, R^2 increases to 17% for the training data and 10% for the hold-out data. The model for stock price return is not strong, the focus of this study being thus on modeling the weekly volatility for the Apple stock.

For the weekly average volatility, the model is much stronger. I follow Christiansen et al. (2012) and Paye (2012), and define weekly volatility as:

$$(5) \quad \sigma_t = \sqrt{r_1 t^2 + r_2 t^2 + r_3 t^2 + r_4 t^2 + r_5 t^2},$$

where the r_s are the five daily changes in stock price for a week with 5 trading days. For 80% of the weeks in the dataset with five trading days, there are five r_s in the above formula. For 20% of the weeks in the dataset with 4 trading days, there are four r_s in the formula. For the few weeks with 3 trading days, there are three r_s in the formula. A classical weekly volatility formula is the standard deviation of five daily stock price changes, which is a variance. The above formula does not subtract the mean from each daily change, before considering each of the five daily returns in a week. The above formula is also not dividing with $N - 1$ before applying the square root. Therefore, compared to the classical volatility formula, the new formula can be interpreted as a measure of stock price fluctuations in a given week without using the weekly mean for daily changes.

All variables are now defined, so the final regression model for weekly volatility can be defined as:

$$(6) \quad \sigma_t = \alpha_0 + \alpha_1 \sigma_{t-1} + \dots + \alpha_p \sigma_{t-p} + \sum_{i=1}^N \beta_i X_{it} + \varepsilon_t,$$

where X is the variable constructed from the Google search data and also the event variables for iPhone launch, Black Friday and Christmas sales, and quarterly reports. The event dummies are defined as 0/1 variables, which take 1 in the week the event occurred, and 0 otherwise.

1. Results and Discussion

The initial test runs on 62 Google searches not indexed against each other, were mostly used to identify the most relevant Google searches. After the most relevant 10 Google searches were indexed against each other, both IBM SPSS LASSO and Autometrics picked only two relevant Google searches, that is, "AAPL" and "AMZN," which are the stock symbols for Apple and Amazon, respectively.

Oxmetrics 8.10 was used to model the 62 Google searches as input variables, with weekly volatility as the dependent variable. Using the automatic model selection function in Oxmetrics 8.10, called Autometrics, the 62 Google searches were tested with time lags from 1 to 4 weeks, and 51 weekly seasonal dummies were also included in the modelling. Event dummies for Apple quarterly reports, iPhone releases, and Black Friday and Christmas sales were also tested.

The final model output in Table 1 is from the automatic model selection, single-equation dynamic modeling using Autometrics.

Software:	OxMetrics 8.10						IBM SPSS Statistics 26					
Variable selection method:	Autometrics						Lasso					
Variables selected	Coefficient	Std.Error	t-value	t-prob	Part.R^2	Coefficient	Std.Error	t-value	t-prob	Part.R^2		
t-1 AAPL: (Worldwide)	0.00147	0.0890	16.5	0.000	0.5552	0.00132	0.0594	22.2	0.000	0.6901		
t-4 AMZN: (Worldwide)	0.00037	0.0763	4.91	0.000	0.0995	0.00027	0.0715	3.71	0.000	0.0585		
t-4 macrumors: (Worldwide)	0.00183	0.0008	2.39	0.018	0.0255							
t-4 apple rumors: (Worldwide)	-0.00659	0.0027	-2.45	0.015	0.0268							

Diagnostics tests:		Diagnostics tests:	
AR 1-7 test:	F(7,211) = 2.6487 [0.0121]*	AR 1-7 test:	F(7,215) = 4.2173 [0.0002]**
ARCH 1-7 test:	F(7,210) = 4.4896 [0.0001]**	ARCH 1-7 test:	F(7,210) = 3.0202 [0.0048]**
Normality test:	Chi^2(2) = 9.5215 [0.0086]**	Normality test:	Chi^2(2) = 10.519 [0.0052]**
Hetero test:	F(10,213) = 3.4856 [0.0003]**	Hetero test:	F(4,219) = 7.9119 [0.0000]**
Hetero-X test:	F(25,198) = 3.1901 [0.0000]**	Hetero-X test:	F(5,218) = 6.7054 [0.0000]**
RESET23 test:	F(2,216) = 6.7985 [0.0014]**	RESET23 test:	F(2,220) = 5.8947 [0.0032]**

Table 1. Model Output from Autometrics & Lasso

6.1 Diagnostic tests

The AR 1-7 test is a standard test of autocorrelation up to degree 7. It tests the joint hypothesis that ε^t is uncorrelated with ε^{t-j} , for any choice of j, against the alternative that ε^t is correlated with ε^{t-j} . The null hypothesis of no autocorrelation between residuals can be rejected for a P-value of 1.21% and significance level of 2.5%. Therefore, there is formal evidence of little autocorrelation between residuals. At a significance level of 1%, the null hypothesis is accepted, which makes it a borderline scenario for this test. In the LASSO model with just two predictors, there is clear rejection of the null hypothesis and formal evidence of autocorrelation between residuals.

The ARCH 1-7 test is a standard ARCH test of the null hypothesis of no ARCH effect, that is, if the squared standardized residuals do not exhibit autocorrelation. The null hypothesis of no ARCH effect is rejected for P-values of 0.0001 and 0.0048. Therefore, there is formal evidence of the ARCH effect in the model.

Normality test. The null hypothesis of normality is rejected at a significance level of 1%, with P-values of 0.86% and 0.52%. Hence, there is no formal evidence of normality for this model.

Hetero and hetero-X tests. The null hypothesis of homoscedasticity can be rejected for P-values of 0–0.03% at the 1% significance level. Hence, there is formal evidence of heteroscedasticity in this model.

RESET123 test. The regression specification error test has a null hypothesis of no squared and cubic terms in the regression model. The null hypothesis can be rejected for P-values of 0.14% and 0.32% at the 1% significance level. Hence, there is formal evidence of mis-specification of the regression model from this test.

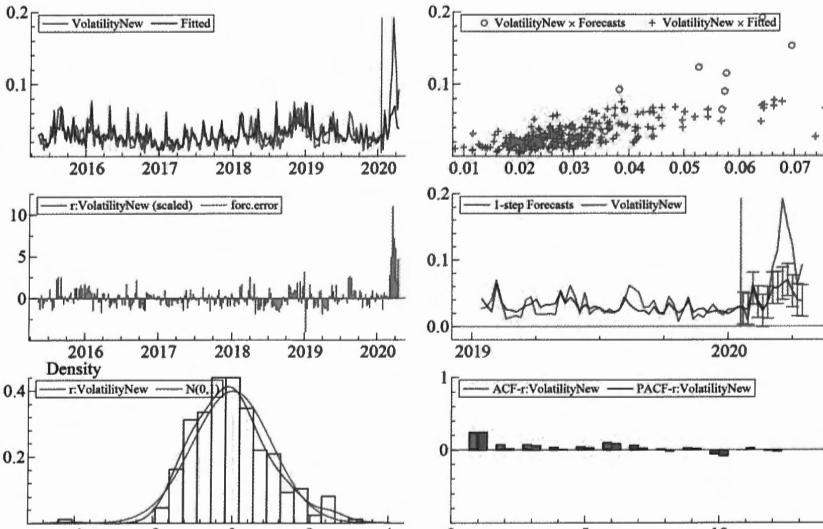


Fig. 9 Model Output from Autometrics

The residual plot shows a random pattern, suggesting a linear model would fit well the data. Residuals are also normally distributed, again suggesting the linear regression model is fitting the dependent variable well. The ACF and PACF plots show no autocorrelation. The COVID-19 peak in March 2020 is an outlier, but the model is not capturing this, as there were no COVID-19 data for training the model.

2 Forecasting evaluation

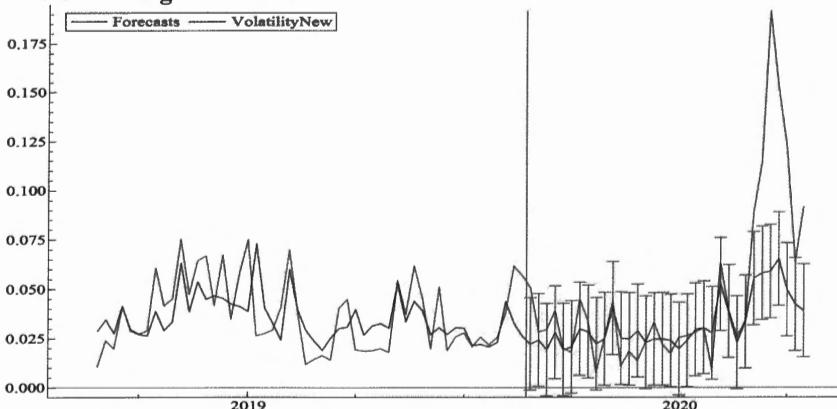


Fig. 10 Forecast graph, Y axis is Weekly volatility. 34 weeks out-of-sample

In Fig. 10, the blue line is the out-of-sample forecast from September 2019 to April 2020 for the last 34 weeks of the dataset. The estimation sample is from April 2015 to August 2019. The green band around the blue line is a 95% band marking +/- 2 forecast standard errors.

From late February 2020 to April 2020, the weekly volatility for the Apple stock took a hit due to the COVID-19 pandemic period, and the forecast model fail to forecast this peak. This is probably due to the COVID-19 pandemic not being captured by the Google searches in this model. The Oxmetrics output for the above 34 weeks dynamic forecast out-of-sample is in Appendix 1, including the COVID-19 pandemic period.

To test how the forecasting performed when excluding the COVID-19 pandemic period, a forecast scenario similar to the setting was tested for a 26-week forecast ending forecast in February 2020 before the pandemic hit the stock market. The difference in forecasting periods for the 34 weeks including the COVID-19 pandemic period and 26 weeks excluding it ensures identical training datasets from week 19 in 2015 to week 34 in 2019 under both forecasting scenarios.

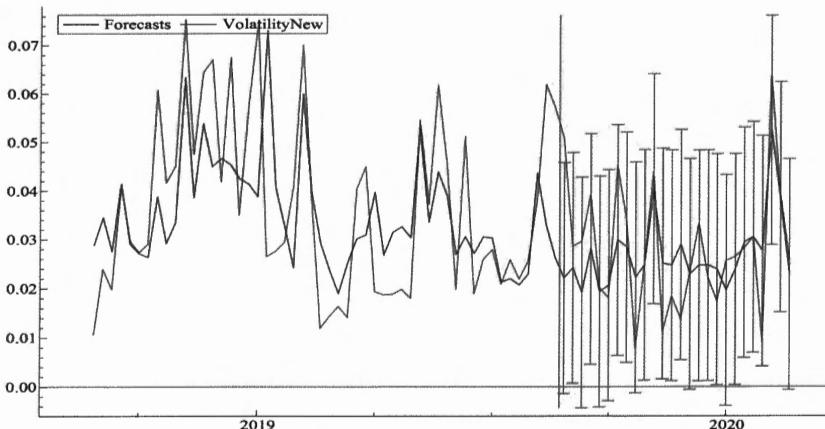


Fig. 11 Forecast graph, Y axis is Weekly volatility. 26 weeks out-of-sample

The Oxmetrics output for the above 26 weeks out-of-sample dynamic forecast is shown in Appendix 2. The Oxmetrics outputs for the two out-of-sample dynamic forecasts without and with the COVID-19 pandemic period are summarized in Table 2.

Excluding the COVID-19 pandemic period	Including the COVID-19 pandemic period
One-step (ex post) forecast analysis 2019 (35)–2020 (8)	One-step (ex post) forecast analysis 2019 (35)–2020 (16)
Training dataset 2015 (19)–2019 (34)	Training dataset 2015 (19)–2019 (34)
Parameter constancy forecast tests:	Parameter constancy forecast tests:
Forecast Chi^2(26) = 19.244 [0.8259]	Forecast Chi^2(34) = 291.67 [0.0000] **
Chow F(26,218) = 0.73626 [0.8217]	Chow F(34,218) = 7.5341 [0.0000] **
CUSUM t(25) = 0.4978 [0.6230]	CUSUM t(33) = 6.580 [0.0000]**
RMSE = 0.010157	RMSE = 0.034577
MAPE = 38.524	MAPE = 40.205
mean(Error)= 0.001264	mean(Error)= 0.014414
SD(Error)= 0.010078	SD(Error)= 0.031434

Table 2. Autometrics Output excluding & including COVID-19.

Comparing the forecasts without and with the pandemic period, the former performs relatively well in terms of Chi^2, Chow, CUSUM, RMSE, MAPE, and mean(Error).

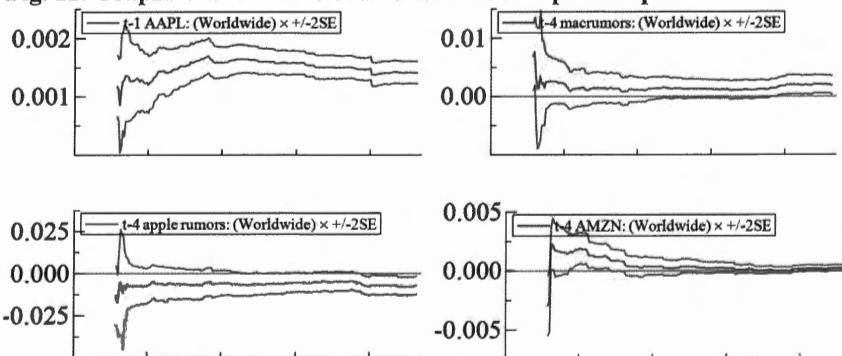
All forecasting was conducted using the model with four predictors selected by Autometrics. For comparing the two used variable selection methods in this article—Autometrics and LASSO—Table 3 presents both methods and their forecasting KPIs.

Software:	OxMetrics 8.10	IBM SPSS Statistics 26
Variable selection method:	Autometrics	Lasso
Variables selected:	t-1 AAPL, t-4 AMZN, t-4 macrumors & t-4 apple rumors	t-1 AAPL & t-4 AMZN
Forecasting excluding the COVID-19 pandemic period		
One-step (ex post) forecast analysis 2019 (35)–2020 (8)		
Training dataset 2015 (19)–2019 (34)		
Parameter constancy forecast tests:		
Forecast Chi^2(26) = 18.646 [0.8510]		
Chow F(26,218) = 0.71147 [0.8482]		
CUSUM t(25) = 0.6245 [0.5380] (zero forecast innovation mean)		
Forecasting including the COVID-19 pandemic period		
One-step (ex post) forecast analysis 2019 (35)–2020 (8)		
Training dataset 2015 (19)–2019 (34)		
Parameter constancy forecast tests:		
Forecast Chi^2(34) = 288.54 [0.0000]**		
Chow F(34,218) = 7.4183 [0.0000]**		
CUSUM t(33) = 6.659 [0.0000]** (zero forecast innovation mean)		
Forecasting excluding the COVID-19 pandemic period		
One-step (ex post) forecast analysis 2019 (35)–2020 (8)		
Training dataset 2015 (19)–2019 (34)		
Parameter constancy forecast tests:		
Forecast Chi^2(26) = 19.189 [0.8284]		
Chow F(26,222) = 0.73701 [0.8210]		
CUSUM t(25) = 0.3433 [0.7342]		
Forecasting including the COVID-19 pandemic period		
One-step (ex post) forecast analysis 2019 (35)–2020 (8)		
Training dataset 2015 (19)–2019 (34)		
Parameter constancy forecast tests:		
Forecast Chi^2(34) = 278.85 [0.0000]**		
Chow F(34,222) = 7.6773 [0.0000]**		
CUSUM t(33) = 6.536 [0.0000]**		

Table 3. Forecasting output from Autometrics & SPSS

7.1 Autometrics recursive graphs

Fig. 12. Graphs of the coefficients for the most important predictors



In Fig. 12, the Google searches for AAPL have the most significant pattern. The coefficient on the Google search AMZN t - 4 is close enough to zero so that it is not worth analyzing it. However, the IBM SPSS LASSO has a much higher partial R² on 8% for this variable compared to only 2% in Autometrics. The coefficients for the Google searches for MacRumors t - 4 and Apple rumors t - 4 are larger than forAAPL, but these Google searches are also relative small compared to the AAPL searches

7.2 IBM SPSS 26 output

Model Summary								
Multiple R	R Square	Adjusted R Square	Regularization "R Square" (1-Error)	Apparent Prediction Error	Expected Prediction Error			N ^b
.725	.525	.514	.475	.525	.560	.077	.256	
Penalty .420 Dependent Variable: VolatilityNew Predictors: t-1 AAPL: (Worldwide) t-1 9to5mac: (Worldwide) t-1 theVerge: (Worldwide) t-1 macrumors: (Worldwide) t-1 IBM: (Worldwide) t-1 appleinsider: (Worldwide) t-1 AMZN: (Worldwide) t-1 apple rumors: (Worldwide) t-1 SNNLF: (Worldwide) t-2 AAPL: (Worldwide) t-2 9to5mac: (Worldwide) t-2 theVerge: (Worldwide) t-2 macrumors: (Worldwide) t-2 IBM: (Worldwide) t-2 appleinsider: (Worldwide) t-2 AMZN: (Worldwide) t-2 apple rumors: (Worldwide) t-2 SNNLF: (Worldwide) t-3 AAPL: (Worldwide) t-3 9to5mac: (Worldwide) t-3 theVerge: (Worldwide) t-3 macrumors: (Worldwide) t-3 IBM: (Worldwide) t-3 appleinsider: (Worldwide) t-3 AMZN: (Worldwide) t-3 apple rumors: (Worldwide) t-3 SNNLF: (Worldwide) t-4 AAPL: (Worldwide) t-4 9to5mac: (Worldwide) t-4 theVerge: (Worldwide) t-4 macrumors: (Worldwide) t-4 IBM: (Worldwide) t-4 appleinsider: (Worldwide) t-4 AMZN: (Worldwide) t-4 apple rumors: (Worldwide) t-4 SNNLF: (Worldwide)								
a. .632 Bootstrap estimate (50 bootstrap samples). b. If N is smaller than the number of active (training) cases, this is due to excluding cases from estimation of the expected prediction error for reason(s) explained in the warnings table.								

Table 4: Model summary from SPSS Lasso

The LASSO model in IBM SPSS 26 has an R^2 on 52.5%, which is very comparable to the one in Autometrics, around 50%. The above lambdian model yielded 0.42, which is optimized after the 19% bootstrap test set (50 bootstrap samples out of the 260 weekly observations).

Final model

The automatic model selection in Autometrics results in two significant Google searches linked to the weekly volatility of the Apple stock and no effects from the seasonal and event dummies or historical values of volatility. The significant Google searches are AAPL t - 1 and AMZN t - 4.

The Google search “AAPL” is time lagged 1 week, being the most significant search linked Apple stock volatility, with a partial R^2 of 49% in Autometrics and 50% in SPSS 26 LASSO.

The Google search AAPL at t - 1, which is the Google search for the Applestock symbol on Nasdaq, is driving up volatility 1 week after thesearches. This is because it is assumed the Google searches for AAPL are mainly done by private investors, under the assumption that professional investors have a lower need to Google search the Apple stock symbol AAPL. These assumptions cannot be proven but are logical.

To verify this conjecture, I interviewed Henrik Ekman, Independent Investment Consultant, former Head of Equities at Maj Invest, on January 28, 2021. He confirmed that professional portfolio analysts are using Google searches to find indications of sales going up and down for the stocks they are analyzing. They also use Google searches in general for information gathering for the stocks currently in their portfolio. While I did not find out any specifics on the difference between private and professional investors in terms of Google searches for stock symbols, this confirmed the professional investors’ general use of Google searches, as shown in this articles Investor Journey Model.

The LASSO algorithm was run in IBM SPSS 26 on the exact same dataset, with weekly volatility as the dependent variable and two significant input variables. AAPL was time lagged 1 week as the most significant input variable, similar to the Autometrics method, with a similar partial R² of 50%. For the second significant input variable, LASSO also chose AMZN t - 4, with a partial R² of 8%. In Autometrics, AMZN t - 4 had a similar partial R².

Autometrics was also tested with a target size of 5% instead of the 1% target size for all other tests. The target size is the t-probability threshold for choosing and eliminating input variables.

The target size for 5% also results in the choice of the MacRumors t - 4 Google search with a partial R² of 2.7% and a positive coefficient and Apple rumors t - 4 with a partial R² of 3.1% and negative coefficient. Given these relative small partial R², further analysis is not necessary. The positive coefficient on MacRumors indicates more private investors conduct these Google searches, while the negative coefficient on Apple rumors indicates more professional investors rely on these Google searches. MacRumors.com has good reputation for vetting Apple rumors, but the Google searches for Apple rumors lead also to MacRumors.com, 9to5mac.com, AppleInsider.com, and TheVerge.com; this wider mix of Apple rumor news sites are used more by professional investors compared to just MacRumors.com. However, given the small partial R² of around 3% for these Apple rumor Google searches, this analysis should of course be interpreted cautiously.

Conclusions

More than 60 Apple-related Google searches were tested in this study as predictors for weekly Apple stock volatility. Under the framework of the newly proposed investor journey model, I analyzed and explained why a buzz in some Apple-related Google searches will dampen the weekly Apple stock volatility and why a buzz in the other searches will increase the weekly volatility for the Apple stock.

A buzz up in the Google search for “AAPL” will increase the Apple stock price volatility in the following week. This is the most significant pattern, since the partial R² for the Google search “AAPL” is 44% in Autometrics and 50% in SPSS’s LASSO.

A buzz up in Google search “AMZN” will also increase Apple stock price volatility after 4 weeks. However, the effect is small compared to AAPL, as the partial R² for this Google search is just 2% in Autometrics and 8% in SPSS LASSO. With the small partial R² of 2–8% for these AMZN Google searches, the results of this analysis should be interpreted with caution.

When the target size changed from 1% to 5% in Autometrics, which is the t-probability threshold for choosing and eliminating input variables, MacRumors t - 4 and Apple rumors t - 4 also become significant, but their partial R² are 2.7% and 3.1%, respectively, so their effect is quite small. MacRumors has a positive coefficient and will increase the Apple stock price volatility after 4 weeks. Apple rumors has a negative coefficient and will decrease the Apple stock price volatility after 4 weeks. The explanation could be that an increase in Google searches 4 weeks before this information is available will result in investors buying and selling more. However, it could also be a random pattern,

considering the small partial R² of around 3% for these Apple rumor news Google searches.

The most predictive Google search for Apple stock volatility was AAPL t-1. The Google searches for AAPL stock symbol are mostly done by private investors, which explains why their buzz up increases volatility.

When the COVID-19 pandemic disrupted the stock market during February–April 2020, both private and professional investors panicked, which is why and the proposed model could not capture the disruption, as the model was not trained with data from this period.

Further research ideas would be to model the Amazon stock based on Google searches, as the AMZN Google searches show a better visual correlation with the Amazon stock price compared to the Apple stock and AAPL Google searches. Perhaps the stock price return for the Amazon stock has a better potential for being modeled and predicted with Google searches. However, there is the need to test whether the Amazon stock price volatility modeling based on Google searches is stronger than that for Apple in this article. That could require an additional model.

The novel investor journey model presented in this article will thus enable further analyses linking big social data to investor behavior on the financial markets.

References

- Asur, S., Huberman, B. A., 2010. Predicting the future with social media. 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), IEEE, pp. 492–499.
- Boldt, L. C., Vinayagamoorthy, V., Winder, F., Schnittger, M., Ekran, M., Mukkamala, R. R., Lassen, N. B., Flesch, B., Hussain, A., Vatrapu, R., 2016. Forecasting Nike's sales using Facebook data. 2016 IEEE International Conference on Big Data (Big Data), pp. 2447–2456.
- Bollen, J., Mao, H., Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of Computer Science*, 2(1), pp. 1–8.
- BusinessInsider.com, April 2018. How Google retains more than 90% of market share". <http://uk.businessinsider.com/how-google-retains-more-than-90-of-market-share-2018-4?ref=US&IR=T>
- Buus Lassen, N., la Cour, L., Vatrapu, R., 2017. Predictive analytics with social media data. In Sloan, L. and Quan-Haase, A. (eds). *The SAGE Handbook of Social Media Research Methods*, Chapter 20, pp. 328–341, Sage.
- Christiansen, C., Schmeling, M., Schrimpf, A., 2012. A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics*, 27(6), pp. 956–977.
- Diebold, F., Mariano, R., 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), pp. 253–263. doi:10.2307/1392185
- Federal Reserve Board, 2016. Triennial Survey of Consumer Finance (SCF), <https://www.federalreserve.gov/econres/files/BulletinCharts.pdf>
- Greenwich Associates, 2015. Social media influencing investment decisions at global institutions.

<https://www.greenwich.com/press-release/social-media-influencing-investment-decisions-global-institutions>

Hendry, D., Castle, J., Doornik, J., Johansen, S., Pretis, F., 2014. Model selection with big data. Oxmetrics Conference, September 2014.

<http://www.timberlake.co.uk/media/wysiwyg/pdf/David%20F.%20Hendry.pdf>

Hendry, D. F., Doornik, J.A., 2014. Empirical model discovery and theory evaluation: Automatic model selection methods in Econometrics.

Hudson, R., Gregoriou, A., 2010. Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns. Available at SSRN: <https://ssrn.com/abstract=1549328> or <http://dx.doi.org/10.2139/ssrn.1549328>

Jacobsen, P. Ø., Ringberg, T., Bjerre M., CBS & CBS BIG DATA LAB RF2016- 2017 + interviews 2018 inspired from: Customer Infinity Model – Per Østergaard Jacobsen, Mogens Bjerre, Claus Andersen 2015<https://cbs-executive.dk/wp-content/uploads/2019/05/Ten-Deadly-Marketing-Sins-30-april-2019-report-and-conclusions.pdf>

Jiao, P., Veiga, A., Walther, A., September 25, 2018. Social media, news media and the stock market. Available at: <https://ssrn.com/abstract=2755933> or <http://dx.doi.org/10.2139/ssrn.2755933>

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, 47, pp. 263–291.

Lassen, N. B., Madsen, R. and Vatrapu, R., 2014,. Predicting iPhone sales from iPhone tweets. In 2014 IEEE 18th International Enterprise Distributed Object Computing Conference, IEEE, pp. 81–90.

Levy, M., Levy, H., 2002. Prospect theory: Much ado about nothing? *Management Science, INFORMS*, 48(10), pp. 1334–1349. DOI: 10.1287/mnsc.48.10.1334.276

Levy, H., 2015. *Stochastic Dominance: Investment Decision Making under Uncertainty*. Springer.

Li, T., van Dalen, J., van Rees, P. J., 2018. More than just noise? Examining the information content of stock microblogs on financial markets. *Journal of InformationTechnology*, 33(1), pp. 50–69.

MarketWatch.com, 2018. Finance Twitter: The 50 most important people for investors to follow. <https://www.marketwatch.com/story/finance-twitter-the-50-most-important-people-for-investors-to-follow-2018-12-13>

Nwogugu, M., 2005. Towards multi-factor models of decision making and risk: A critique of Prospect Theory and related approaches, part I. *Journal of Risk Finance*, 6(2), pp. 150–162. <https://doi.org/10.1108/15265940510585815>

Paye, B.S., 2012. “Déjà vol”: Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, 106(3),pp. 527–546.

Preis, T., Moat, H. S., Stanley, H. E., 2013. Quantifying trading behavior in financial markets using Google Trends. *Nature, Scientific Reports*, 3, p. 1684. DOI:10.1038/srep01684 (2013).

TheGlobeAndMail.com, 2018. How Twitter can help institutional investors market better trading. <https://www.theglobeandmail.com/business/careers/business-education/article-how-twitter-can-help-institutional-investors-make-better-trading/>

Wakker, P. P., 2010. *Prospect Theory: For Risk and Ambiguity*. Cambridge University Press.

Appendix 1. Forecast output from Oxmetrics, September 2019–April 2020, including the COVID-19 pandemic period.

1-step (ex post) forecast analysis 2019(35) - 2020(16)							
Parameter constancy forecast tests:							
Forecast	Chi^2(34)	=	291.67	[0.0000]**			
Chow	F(34, 218)	=	7.5341	[0.0000]**			
CUSUM	t(33)	=	6.580	[0.0000]**	(zero forecast innovation mean)		
Dynamic (ex ante) forecasts for VolatilityNew (SE based on error variance only)							
Horizon	Forecast	SE	Actual	Error	t-value	-2SE	+2SE
2019(35)	0.0221757	0.01181	0.0510110	0.028835	2.443	-0.0014354	0.045787
2019(36)	0.0242622	0.01181	0.0286650	0.0044028	0.373	0.00065115	0.047873
2019(37)	0.0192232	0.01181	0.0297024	0.010479	0.888	-0.0043879	0.042834
2019(38)	0.0281562	0.01181	0.0393877	0.011231	0.951	0.0045452	0.051767
2019(39)	0.0194324	0.01181	0.0202125	0.00078013	0.066	-0.0041787	0.043043
2019(40)	0.0207155	0.01181	0.0181770	-0.0025385	-0.215	-0.0028956	0.044327
2019(41)	0.0299563	0.01181	0.0452540	0.015298	1.296	0.0063453	0.053567
2019(42)	0.0284997	0.01181	0.0341143	0.0056147	0.476	0.0048886	0.052111
2019(43)	0.0223038	0.01181	0.00787080	-0.014433	-1.223	-0.0013072	0.045915
2019(44)	0.0248636	0.01181	0.0253074	0.00044384	0.038	0.0012525	0.048475
2019(45)	0.0405084	0.01181	0.0441805	0.0036721	0.311	0.016897	0.064119
2019(46)	0.0251504	0.01181	0.0111957	-0.013955	-1.182	0.0015393	0.048761
2019(47)	0.0247585	0.01181	0.0185567	-0.0062018	-0.525	0.0011474	0.048370
2019(48)	0.0290163	0.01181	0.0138208	-0.015195	-1.287	0.0054052	0.052627
2019(49)	0.0229785	0.01181	0.0235305	0.00055208	0.047	-0.00063259	0.046590
2019(50)	0.0247348	0.01181	0.0334344	0.0086596	0.737	0.0011237	0.048346
2019(51)	0.0247689	0.01181	0.0222302	-0.0025387	-0.215	0.0011579	0.048380
2019(52)	0.0239338	0.01181	0.0175467	-0.0063871	-0.541	0.00032276	0.047545
2020(1)	0.0197210	0.01181	0.0257094	0.0059884	0.507	-0.0038901	0.043332
2020(2)	0.0239680	0.01181	0.0265275	0.0025596	0.217	0.00035692	0.047579
2020(3)	0.0294778	0.01181	0.0282960	-0.0011817	-0.100	0.0058667	0.053089
2020(4)	0.0305566	0.01181	0.0306042	4.76248-05	0.004	0.0069455	0.054168
2020(5)	0.0277304	0.01181	0.00949557	-0.018235	-1.545	0.0041193	0.051341
2020(6)	0.0525129	0.01181	0.0638056	0.011293	0.957	0.028982	0.076124
2020(7)	0.0388666	0.01181	0.0394267	0.00062008	0.053	0.015196	0.062418
2020(8)	0.0229329	0.01181	0.0259554	0.0030225	0.256	-0.00067814	0.046544
2020(9)	0.0334589	0.01181	0.0340971	0.00063827	0.054	0.0098478	0.057070
2020(10)	0.0554885	0.01181	0.0898427	0.033554	2.842	0.031877	0.079100
2020(11)	0.0581828	0.01181	0.114265	0.056082	4.750	0.034572	0.081794
2020(12)	0.0590218	0.01181	0.191717	0.13270	11.240	0.035411	0.082633
2020(13)	0.0653553	0.01181	0.152223	0.086868	7.358	0.041744	0.088966
2020(14)	0.0498154	0.01181	0.122598	0.072782	6.165	0.026204	0.073426
2020(15)	0.0423422	0.01181	0.0638132	0.021471	1.819	0.018731	0.065953
2020(16)	0.0391658	0.01181	0.0919330	0.052767	4.470	0.015555	0.062777
mean(Error)	=	0.014404	RMSE	=	0.034577		
SD(Error)	=	0.031434	MAPE	=	40.205		

Appendix 2. Forecast output from Oxmetrics, September 2019–February2020, excluding the COVID-19 pandemic period.

1-step (ex post) forecast analysis 2019(35) - 2020(8)							
Parameter constancy forecast tests:							
Forecast	Chi^2(26) =	19.244	[0.8259]				
Chow	F(26,218) =	0.73626	[0.8217]				
CUSUM	t(25) =	0.4978	[0.6230]	(zero forecast innovation mean)			
Dynamic (ex ante) forecasts for VolatilityNew (SE based on error variance only)							
Horizon	Forecast	SE	Actual	Error	t-value	-2SE	+2SE
2019(35)	0.0221757	0.01181	0.0510110	0.028835	2.443	-0.0014354	0.045787
2019(36)	0.0242622	0.01181	0.0286650	0.0044028	0.373	0.00065115	0.047873
2019(37)	0.0192232	0.01181	0.0297024	0.010479	0.888	-0.0043879	0.042834
2019(38)	0.0281562	0.01181	0.0393877	0.011231	0.951	0.0045452	0.051767
2019(39)	0.0194324	0.01181	0.0202125	0.00078013	0.066	-0.0041787	0.043043
2019(40)	0.0207155	0.01181	0.0181770	-0.0025385	-0.215	-0.0028956	0.044327
2019(41)	0.0299563	0.01181	0.0452540	0.015298	1.296	0.0063453	0.053567
2019(42)	0.0284997	0.01181	0.0341143	0.0056147	0.476	0.0048886	0.052111
2019(43)	0.0223038	0.01181	0.00787080	-0.014433	-1.223	-0.0013072	0.045915
2019(44)	0.0248636	0.01181	0.0253074	0.00044384	0.038	0.0012525	0.048475
2019(45)	0.0405084	0.01181	0.0441805	0.0036721	0.311	0.016897	0.064119
2019(46)	0.0251504	0.01181	0.0111957	-0.013955	-1.182	0.0015393	0.048761
2019(47)	0.0247585	0.01181	0.0185567	-0.0062018	-0.525	0.0011474	0.048370
2019(48)	0.0290163	0.01181	0.0138208	-0.015195	-1.287	0.0054052	0.052627
2019(49)	0.0229785	0.01181	0.0235305	0.00055208	0.047	-0.00063259	0.046590
2019(50)	0.0247348	0.01181	0.0334344	0.0086996	0.737	0.0011237	0.048346
2019(51)	0.0247689	0.01181	0.0222302	-0.0025387	-0.215	0.0011579	0.048388
2019(52)	0.0239338	0.01181	0.0175467	-0.0063871	-0.541	0.00032276	0.047545
2020(1)	0.0197210	0.01181	0.0257094	0.0059884	0.587	-0.0038901	0.043332
2020(2)	0.0239680	0.01181	0.0265275	0.0025596	0.217	0.00035692	0.047579
2020(3)	0.0294778	0.01181	0.0282960	-0.0011817	-0.180	0.0058667	0.053089
2020(4)	0.0305566	0.01181	0.0306042	4.7624e-05	0.004	0.0069455	0.054168
2020(5)	0.0277304	0.01181	0.00949557	-0.018235	-1.545	0.0041193	0.051341
2020(6)	0.0525129	0.01181	0.0638056	0.011293	0.957	0.028902	0.076124
2020(7)	0.0388066	0.01181	0.0394267	0.00062008	0.053	0.015196	0.062418
2020(8)	0.0229329	0.01181	0.0259554	0.0030225	0.256	-0.00067814	0.046544
mean(Error)	=	0.0012644	RMSE =	0.010157			
SD(Error)	=	0.010078	MAPE =	38.524			

Vil Polen opleve et EU-exit (Polexit)?

Agnieszka Szmydynska

Økonomisk Institut, Københavns Universitet

August 2022

Abstract

The aim of this paper is to examine variables of importance for the membership of the European Union, using data from the European Social Survey (ESS), in the case of Italy and Poland for the years 2016 and 2018, as compared to Germany. A logistic regression on the question of the EU-membership exhibits a hit-ratio between 65% and 90% depending on country and ESS-round. A supplementary multiple linear regression model of trust in the European-Parliament is conducted both as an indicator of the EU-membership variable and to allow the respondents to express their opinion through a broader range of options. Despite significant country differences, immigration is found to be of particular importance for the membership of the EU. A significant difference between 2016 and 2018 seems attributable to the refugee crisis of 2015, which reinforced a surge of Eurosceptic parties. The findings further suggest great importance of trust in national institutions and authorities, especially regarding trust in the European Parliament.

1 Indledning

Den Europæiske Union har i årtier sikret fred, stabilitet og velstand for sine medlemslande ved at styrke den europæiske politiske og økonomiske integration, udenrigspolitikken og ved at fremme et samarbejde med hensyn til immigration, asyl og retlige anliggender.

Det forekommer som en generel tendens i Europa, at euroskeptiske partier vinder indpas blandt vælgerne ved at udnytte kriser og midlertidige udfordringer, som medvirker til en nedbrydning af det positive syn på det europæiske samarbejde. Trods værende hovedmodtageren af EU-støtte og indehaveren af placeringen som det mest pro-EU europæiske land i ESS' runde 9 synes Polen paradoksal siden 2016 at være på vej i en retning, der modarbejder de demokratiske ændringer. Udviklingen tilskrives valgsejren af det højrepopulistiske parti "Lov og retfærdighed" (PiS) i 2015. Den italienske befolkning har imidlertid allerede siden 00'erne oplevet en stigende euroskepticisme primært drevet af den kontroversielle Dublin-forordning og Berlusconi-regeringens politik i 2001-05, hvor Italiens økonomiske udfordringer blev projekteret over på euroen. Ikke desto mindre støtter majoriteten af den italienske befolkning et fortsat EU-medlemskab.

I indeværende indlæg søges det at afdække hvilke faktorer, der påvirker holdningen til EU-medlemskabet, suppleret med en analyse af tillid til Europa-Parlamentet. Undersøgelsen udføres for Polen og Italien i 2016 og 2018, holdt op imod Tyskland som referenceland. For at modellere holdningen til medlemskabet benyttes data fra European Social Survey i en logistisk regressionsmodel, der betragter en binær responsvariabel. Tilliden til Europa-Parlamentet analyseres gennem en multipel lineær regressionsmodel.

Undersøgelsen udviser at immigration, målt ved faktorer, der udtrykker henholdsvis holdningen til immigranter bosættelse i og effekt på landet, var af størst signifikans og negativt korreleret med holdningen til EU. I Polen kan korrrelationen tilskrives en overdreven frygt for immigration i forbindelse med flygtningekrisen, skabt af PiS under valget i 2015. Den kontroversielle fordeling af flygtninge i forbindelse med Dublin-forordningen, udnyttet af EU-skeptiske partier, kan forklare den negative korrelation med immigrationsfaktorerne for Italien. Tillid til nationale institutioner og myndigheder findes at være af næststørst signifikans for medlemskabet og størst for tillid til Europa-Parlamentet. En positiv korrelation med tillidsvariable ses i forbindelse med støtte til demokratiet, som er en kerneværdi i EU, hvor den negative korrelation med institutionsfaktoren i Polen kan tilskrives PiS's udfasning af demokratiet.

Afsnit 2, 3 og 4 præsenterer den teoretiske ramme og data. Afsnit 5 analyserer den empiriske evidens for EU-medlemskabet og tilliden til Europe Parlamentet. Afsnit 6 diskuterer konsekvenserne for landene og EU, mens afsnit 7 konkluderer.

2 Metode

Den multiple lineære regressionsmodel benyttes til at undersøge, hvordan og hvilke variable, der påvirker tilliden til Europa-Parlamentet. Data fra runde 8 (2016) og runde 9 (2018) anvendes til at identificere en eventuel udvikling i holdninger indenfor landene. Den multiple lineære regressionsmodel kan udtrykkes som:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k + u, \quad (1)$$

hvor β_0, \dots, β_k igen er de estimerede parametre og u er fejleddet. Den multiple lineære regression benyttes både som indikator for EU-medlemskabet og med den hensigt at give respondenterne en bredere mulighed for at udtrykke sig.

Den logistiske regressionsmodel vil blive anvendt til at undersøge den binære variabel "Forblive i EU", i Italien, Polen og Tyskland på baggrund af udvalgte informationsrige variable. Den logistiske regression kan opskrives som følger:

$$\hat{p} = P(y = 1|x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)}, \quad (2)$$

hvor $P(Y = 1|x)$ er sandsynligheden for at stemme "ja" til at "Forblive i EU". Ved anvendelse af logit-funktionen fås:

$$\log\left(\frac{\hat{p}}{1 - \hat{p}}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k, \quad (3)$$

hvor $\log\left(\frac{\hat{p}}{1 - \hat{p}}\right)$ er log af odds. På trods af ligheden med ligningen for den multiple lineære regression, kan man ikke fortolke estimatorne på samme vis. Effekten af en én-enheds ændring af x varierer alt efter, hvilken værdi den uafhængige variabel antager, da størrelsen af effekten ligeledes påvirkes af de resterende uafhængige variables værdier.

3 Data - European Social Survey (ESS)

ESS-data er baseret på ansigt-til-ansigt interview udført på tilfældigt udvalgte stikprøver bestående af respondenter på 15 år og op efter i landenes private husstande, uagtet af statsborgerskab, nationalitet og sprog. (ESS, 2021c) Stikprøverne blev stratificeret, og klynger udvalgtes geografisk fra landenes respektive folkeregistre. (Lynn, 2021) ESS 9-data vedrører perioden 29. august 2018 - 20. marts 2019 med

stikprøvestørrelser på 2.745 for Italien, 1.500 for Polen og 2.358 for Tyskland. (ESS, 2021b) Runde 8 omfatter perioden 23. august 2016 - 26. marts 2017. Stikprøvestørrelserne befinner sig på 2.626 for Italien, 1.694 for Polen og 2.852 for Tyskland. (ESS, 2021a)

Under antagelse af, at processen hvormed de manglende observationer genereres er MAR (Missing At Random), og da andelen af non-response ikke er betydelig, foretages multipel imputation med MCMC (Markov Chain Monte Carlo) for optimering af statistisk information. For at tage højde for forskellige udvælgelsessandsynligheder, stikprøvesfejl, dækningsfejl og non-response bias, anvendes analysevægten, der blandt andet bruger et kernesæt af demografiske faktorer (køn, alder og uddannelse), som forbinder med tilbøjeligheden til at svare. χ^2 -tests implicerer, at de vægtede stikprøver generelt er repræsentative på basis af køn fordelt på alder. Imidlertid identificeres betydelige forskelle mellem de betragtede landes stikprøver og populationer med hensyn til valgstemmer fordelt på partier.

3.1 Faktorenanalyse

For en mere efficien udnyttelse af data foretages en faktorenanalyse, som søger underliggende uboserverbare variable, der er reflekteret i de observerede variable på baggrund af deres indbyrdes korrelation. De resulterende faktorer er en linearkombination af optimalt vægtede observerede variable. Varimax rotation antager ortogonale faktorer (ikke korrelerede), og anvendes for at sikre lettere fortolkelige faktorer, der kan anvendes som prædiktorer i den statistiske analyse (O'Rourke and Hatcher, 2014).

Betydningen af den enkelte variabel afspejles af korrelationen mellem variablen og faktoren. Dette taget i betragtning søges en optimal og simpel struktur, således at enhver variabel loader højt på én, og udelukkende én, faktor. Minimumsværdien for "Factor Loading" sættes til 0,4. Tabel 3.1 præsenterer faktormønstret for de endelige modeller for Italien, mens et ensartet mønster gælder for Polen og Tyskland.

Kun fortolkelige faktorer med en egenværdi på mere end 1 er inkluderet (Kaiser's kriterie) (O'Rourke and Hatcher, 2014). I analysen af faktorerne tages udgangspunkt i runde 8 for Italien, men konklusionerne kan udvides til de betragtede runder og lande.

De to første faktorer dannes på basis af otte spørgsmål, hvor faktor 1 forklarer 61% af variationen mens faktor 2 forklarer en relativt mindre del, 12%. Givet mønstret af variablene i faktoren beskriver første faktor holdningen til hvilke immigranter der "bør" tillades i landet, hvis nogen, mens faktor 2 beskriver, hvordan immigration menes at have påvirket landet. Jo højere værdi faktor 1 antager, desto mere negativt stemt er respondenten overfor at tillade immigranter at bosætte sig i landet. Derimod indikerer en højere værdi af faktor 2 at respondenten udviser en positiv holdning til immigranternes indflydelse på landet. Retningen af faktorerne udledes ved først at inddæle de normalfordelte faktorer i to grupper, henholdsvis over og under 0, og efterfølgende holde det sammen med de variable, der indgår i faktoren. Retningen fortolkes ved at se på, hvordan svarmulighederne indenfor variablen fordeler sig i forhold til faktoren.

Faktoren "Vurdering af den offentlige sektor" er baseret på høje loadings af holdningen til tilstanden af uddannelse og sundhedsvæsenet i landet. Faktoren forklarer selvstændigt 62% af variationen i fire observerede variable. En højere værdi af faktor 3 implicerer en bedre vurdering af den offentlig sektor.

Tabel 3.1: Faktoranalyse, Italien

Faktor nr.	Faktornavn og Spørgsmål	Faktor loading:	R8	R9
1. Immigrantfordeling				
1.	Tillad immigranter af samme race/etnisk gruppe som majoritet	86*	86*	
1.	Tillad immigranter af forskellig race/etnisk gruppe fra majoritet	89*	90*	
1.	Tillad immigranter fra fattigere lande uden for EU	86*	87*	
1.	Immigranter gør land et bedre sted at leve	-44*	-49*	
1.	Immigration dårlig eller god for landets økonomi	-52*	-57*	
1.	Landets kulturliv undermineret eller beriget af immigranter	-46*	-57*	
1.	EU unificering skal gå længere eller gået for langt	-	-45*	
2. Immigrationseffekt				
2.	Immigranter gør land et bedre sted at leve	73*	69*	
2.	Immigration dårlig eller god for landets økonomi	70*	65*	
2.	Landets kulturliv undermineret eller beriget af immigranter	71*	66*	
2.	EU unificering skal gå længere eller er gået for langt	72*	63*	
2.	Følelsesmæssig tilknytning til EU	72*	79*	
3. Vurdering af offentlig sektor				
3.	Udannelsestilstand i dag	83*	81*	
3.	Sundhedsvæsenets tilstand i dag	76*	88*	
4. Tillid til de nationale institutioner				
4.	Tilfreds med demokratiet virken	67*	59*	
4.	Tillid til politikere	93*	92*	
4.	Tillid til landets parlament	84*	78*	
4.	Tillid til politiske partier	92*	91*	
4.	Tillid til retssystemet	49*	-	
5. Tillid til de nationale myndigheder				
5.	Tillid til retssystemet	71*	80*	
5.	Tillid til politiet	93*	90*	
6. Diskrimination overfor LGBT				
6.	LGBT kan frit leve livet, som de vil	86*	87*	
6.	LGBT par har ret til at adoptere børn	68*	75*	
6.	Føler skam, hvis tæt familieemedlem er LGBT	-76*	-80*	

Kilde: egne beregninger med udgangspunkt i ESS-data fra runde 8 (2016) og 9 (2018).

Anm.: beregninger er udført ved brug af analysevejgent, anweight, og faktorer er normalfordelte $N(0,1)$

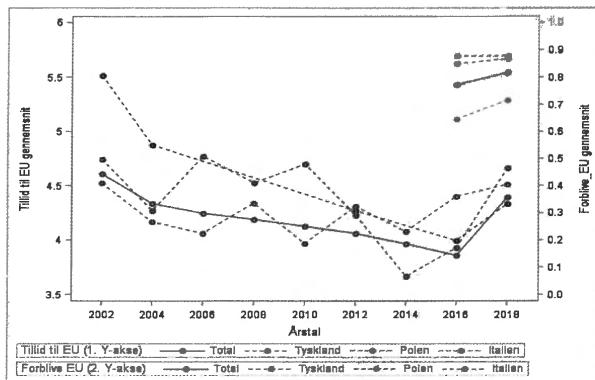
Faktorerne "Tillid til nationale institutioner" og "Tillid til nationale myndigheder" blev dannet i én faktoranalyse på baggrund af seks spørgsmål. Tilsammen forklarer faktorene 79% af variationen, hvoraf første faktor forklarer 63% og den anden faktor forklarer 16%. For begge faktorer gælder, at en højere værdi indikerer større tillid. Den sidste faktor beskriver holdningen til LGBT-miljøet og kaldes "Diskrimination, LGBT". Faktoren forklarer 59% af variationen og en lav værdi betyder at respondenten er mere støttende overfor LGBT-rettigheder.

4 Empiri

Den engelske befolkning udviste i 2014 det laveste niveau af tillid til Europa-Parlamentet blandt alle medlemslande. Alligevel forekom det som en overraskelse, da Storbritannien i 2016 vedtog at forlade EU (Fitzgerald and Sibley, 2015). Figur 1 implicerer at den gennemsnitlige tillid i Europa har været dalende frem til 2016 efterfulgt af en skarp stigning til 2018. Stigningen kan være et udtryk for en voksende bekymring om Brexits følger og en større bevidsthed om fordelene ved EU-medlemskabet grundet de vanskelige forhandlinger (Guillot and Corlett, 2018). Denne udvikling observeres trods fremgang i tilknytningen til

euroskeptiske partier (Wike et al., 2019).

Figur 1: Udvikling i holdningen til EU-medlemskab og tillid til Europa-Parlamentet



Kilde: egne beregninger med udgangspunkt i ESS data runde 8 og 9.

Anm.: beregninger udført med analysevægten, anweight. Total er gennemsnit for alle EU-lande.

Holdningen til at forblive i EU ses at være lavest i Italien og under gennemsnittet, mens Polen ligger højest efterfulgt af Tyskland. Det fremgår, at landene ikke adskiller sig betydeligt fra hinanden i forbindelse med tillid til Europa-Parlamentet og holdning til EU-medlemskabet i 2018. Imidlertid ligger Polen og Italien isoleret i den lave ende af skalaen, når tillidsvariable, set i tabel 4.1, betragtes. Tillid argumenteres som værende en nøglekomponent for økonomisk vækst og en større tillid virker som en indikator for demokratiets robusthed i et samfund. Det lave niveau af tillid i Polen og Italien giver et indblik i, hvordan landene adskiller sig fra eksempelvis Tyskland eller skandinaviske lande. (OECD, 2013)

Tabel 4.1: Tillidsvariable for flere lande.

	Tillidskapital, psykologisk		Tillid til politiske partier		Tillid til politiet		Tillid til retssystemet	
	2016	2018	2016	2018	2016	2018	2016	2018
Norge	6,63	6,55	5,43	5,33	7,38	7,57	7,25	7,37
Finland	6,56	6,69	4,83	5,00	8,20	8,09	7,08	7,09
Sverige	6,25	6,31	4,80	5,03	6,73	7,00	6,19	6,49
Holland	6,07	6,15	5,14	5,34	6,68	6,99	6,24	6,56
Estland	5,68	5,54	3,62	3,78	6,86	7,02	5,88	5,96
England	5,62	5,55	3,87	3,62	6,67	6,67	5,95	5,84
Tyskland	5,58	5,59	4,15	3,95	7,10	7,08	6,05	6,02
Belgien	5,22	5,27	4,00	4,22	6,36	6,54	5,26	5,33
Tjekkiet	5,06	5,10	3,56	3,60	5,80	5,92	5,17	5,20
Spanien	4,92	5,02	2,43	2,51	6,56	6,73	3,92	4,75
Slovenien	4,87	4,85	2,47	2,68	6,04	6,15	3,67	3,92
Portugal	4,58	4,49	2,67	2,82	6,06	6,22	4,05	4,22
Italiem	4,51	4,62	2,29	2,94	6,13	6,64	4,53	5,34
Polen	4,29	4,32	2,44	3,15	5,80	5,72	3,97	4,29

Kilde: egne beregninger med udgangspunkt i ESS-data runde 8 (2016) og 9 (2018).

Anm.: *** reflekterer et signifikansniveau på 0,001, ** på 0,01 og * på 0,05. Standardafvigelse er angivet i parentes. Beregninger er udført med analysevægten, anweight.

Estimationen af ”Tillid til Europa-Parlamentet” blev udført gennem en multipel lineær regression ved at anvende simpel OLS. Gennem en undersøgelse af antagelserne MLR 1-6, udførtes en modelkontrol. På grund af sandsynligheden for udeladte variable, målefejl og risikoen for model-misspecifikation, primært vedrørende endogenitet, bør resultaterne i tabel 4.2 behandles med omhu.

Tabel 4.2: Estimationsresultater - Tillid til Europa-Parlamentet

	Polen		Italien		Tyskland	
	2016	2018	2016	2018	2016	2018
Intercept	4,512*** (0,240)	5,864*** (0,270)	4,453*** (0,159)	5,190*** (0,180)	5,086*** (0,165)	5,298*** (0,165)
Immigrationseffekt	0,797*** (0,076)	1,179*** (0,085)	0,511*** (0,059)	0,803*** (0,059)	0,421*** (0,058)	0,787*** (0,069)
Tillidskapital, psykologisk	0,076** (0,029)	-0,033 (0,031)	-0,042* (0,019)	-0,047* (0,021)	-0,014 (0,021)	-0,019 (0,022)
Tillid til nationale institutioner	0,953*** (0,077)	1,107*** (0,092)	2,114*** (0,058)	1,833*** (0,056)	2,198*** (0,062)	2,461*** (0,071)
Tillid til nationale myndigheder	0,977*** (0,074)	0,718*** (0,087)	1,271*** (0,052)	1,067*** (0,054)	0,799*** (0,053)	0,859*** (0,063)
Vurdering af offentlig sektor	0,372*** (0,073)	0,149 (0,085)	0,069 (0,054)	-0,044 (0,056)	0,238*** (0,054)	0,182** (0,064)
Diskrimination, LGBT	-0,335*** (0,078)	-0,446*** (0,090)	-0,127* (0,053)	-0,126* (0,057)	-0,059 (0,059)	-0,162* (0,063)
Placering på højre/venstre-skala	-0,187*** (0,023)	-0,198*** (0,026)	0,006 (0,014)	-0,080*** (0,017)	0,007 (0,017)	-0,005 (0,017)
Køn (Reference = Mand)	0,295** (0,102)	-0,028 (0,111)	0,075 (0,069)	0,137 (0,073)	0,551*** (0,063)	0,260*** (0,066)
Alder	-0,001 (0,003)	0,002 (0,003)	-0,007*** (0,002)	-0,006** (0,002)	-0,019*** (0,002)	-0,016*** (0,002)
<i>R</i> ² -værdi	34,3%	29,9%	57,1%	50,6%	53,9%	59,4%
Observationer	1,558	1,399	2,489	2,559	2,746	2,273

Kilde: egne beregninger med udgangspunkt i ESS-data runde 8 (2016) og 9 (2018).

Anm.: *** reflekterer et signifikansniveau på 0,001, ** på 0,01 og * på 0,05. Standardafvigelse er angivet i parentes. Beregninger er udført med analysevægten, anweight.

I tabel 4.3 præsenteres parameterestimaterne for den logistiske regressionsmodel, estimeret ved brug af MLE. Som modelkontrol blev ROC-kurven og modellernes hit-ratio undersøgt. Afhængigt af ESS-runde og land havde ROC-kurverne et areal under kurven mellem 0,75 og 0,85 og modellerne en hit-ratio mellem 65% og 90%, hvor cutoff værdien blev valgt ved en maksimering af produktet mellem specificiteten og sensitiviteten.

Tabel 4.3: Estimationsresultater - Medlemskab af EU

	Polen		Italien		Tyskland	
	2016	2018	2016	2018	2016	2018
Intercept	2,800*** (0,240)	3,690*** (0,266)	1,447*** (0,141)	1,697*** (0,156)	1,985*** (0,197)	3,009*** (0,207)
Immigrantfordeling	-0,244*** (0,090)	-0,426*** (0,106)	-0,473*** (0,063)	-0,569*** (0,061)	-0,413*** (0,073)	-0,988*** (0,088)
Immigrationseffekt	1,043*** (0,093)	1,134*** (0,103)	1,355*** (0,075)	1,490*** (0,072)	1,589*** (0,081)	2,152*** (0,103)
Tillidskapital, psykologisk	-0,051 (0,035)	-0,212*** (0,036)	-0,061*** (0,023)	-0,080** (0,023)	0,116*** (0,028)	-0,035 (0,030)
Tillid til nationale institutioner	-0,111 (0,091)	-0,392*** (0,106)	0,146* (0,069)	-0,071 (0,063)	0,622*** (0,084)	0,917*** (0,101)
Tillid til nationale myndigheder	-0,130 (0,086)	0,484*** (0,103)	0,308*** (0,060)	0,275*** (0,060)	0,233*** (0,067)	0,337*** (0,078)
Vurdering af offentlig sektor	0,393*** (0,092)	-0,093 (0,101)	0,204*** (0,060)	0,390*** (0,062)	-0,020 (0,070)	-0,164 (0,089)
Diskrimination, LGBT	-0,401*** (0,095)	-0,270* (0,114)	-0,259*** (0,063)	-0,197*** (0,061)	-0,074 (0,071)	-0,215*** (0,081)
Placering på højre/venstre-skala	-0,061* (0,028)	-0,102*** (0,031)	-0,110*** (0,016)	-0,064*** (0,018)	-0,095*** (0,024)	-0,060* (0,024)
Stemt i sidste valg (reference = Nej)	0,128 (0,070)	0,595*** (0,067)	0,311*** (0,046)	0,318*** (0,050)	0,224*** (0,051)	0,318*** (0,053)
AUC-værdi	75,3	80,5	81,1	81,3	85,5	89,5
Observationer	1.475	1.284	1.913	2.161	2.426	2.077

Kilde: egne beregninger med udgangspunkt i ESS-data runde 8 (2016) og 9 (2018).

Anm.: *** reflekterer et signifikansniveau på 0,001, ** på 0,01 og * på 0,05. Standardafvigelse er angivet i parentes. Beregninger er udført med analysevegten, anweight.

5 Analyse

I det følgende præsenteres udvalgte analyselementer. Asyl- og immigrationspolitikken forekommer som et af de største stridspunkter mellem EU-medlemslande, konsistent med immigrationseffektens høje Wald χ^2 -værdi, og sætter det europeiske sammenhold på prøve. Der er en statistisk signifikant forskel på indvirkningen af immigrationseffekt på holdningen til at forblive i EU mellem runde 8 og runde 9 for Tyskland. På tværs af lande er forskellen signifikant i både 2016 og 2018.

Udviklingen ses at være i overensstemmelse med Polens brud på EU-loven vedr. omplacering af immigranter under flygtningekrisens begyndelse i 2015. Sammen med post-kommunistiske Ungarn og Tjekkiet, nægtede Polen at tage imod tildelte flygtninge, under henvisning til sikkerhedsmæssige foranstaltninger, sammenholdt med den allerede høje tilstrømning af immigranter fra Ukraine. Polen betvivlede det juridiske grundlag for en fast kvote-tildeling af 120.000 flygtninge oprindeligt ankommet til Italien og Grækenland, til hvert EU-land. Dette på trods af EU-domstolens udtalelse om solidaritetsprincippet, der til tider indebærer en byrdefordeling. (ECJ, 2020) Polen var som resultat begrænset berørt af flygtningekrisen i 2015.

EU modtog i 2015 over 1,3 mio. asyl-ansøgninger og 1,26 mio. i 2016, størstedelen fra Mellemøsten, hvorfra

Tyskland modtog majoriteten (476.510 og 745.160) (Eurostat, 2021a). Denne hændelse er markeret ved den tyske forbundskansler Angela Merkels kendte frase "Wir schaffen das", hvor Tyskland påtog sig en stor byrde på vegne af andre Europæiske lande. (Cohen, 2015) Denne ansvarstagning har landet dog siden hen, til en vis grad, fortrudt, hvilket fremgik af øget grænseovervågning og anvendelse af Dublin-forordningen, hvilken fremsiger at det land, som asylansøgeren initialet ankommer til har ansvar for krav om beskyttelse. (Barigazzi, 2019)

Dublin-forordningen er til ugunst for grænselandene Italien og Grækenland, da majoriteten af flygtninge ankommer med både via Middelhavet. Dette indikerer, at mens 2015 og 2016 var præget af en primær immigrationskrise var den efterfølgende periode op til 2018 karakteriseret ved en sekundær immigration, der begik sig på en omfordeling og tilbagesendelse af de modtagne immigranter mellem medlemslandene. (Villa, 2019) Det forekommer da intuitivt, at der er en statistisk signifikant forskel mellem 2016 og 2018 i Tyskland, da 2015 og 2016 var karakteriseret ved en stor tilstrømning af immigranter, mens Italien løbende har modtaget immigranter, både direkte og via en "returnering" fra Tyskland og andre medlemslande. Dertil karakterieres forskellen på tværs af lande kort sagt ved, at Polen ikke har modtaget et betydeligt antal, mens både Italien og Tyskland tog mange ind, dog i forskellig grad.

Påvirkningen af immigrationseffekt på tilliden til Europa-Parlamentet er statistisk signifikant mellem begge runder for hvert land, men på tværs af lande udelukkende mellem Polen og henholdsvis Tyskland og Italien. Denne tendens kan tilskrives Polens store afvigelse fra Tyskland og Italien i forbindelse med indtagelse af immigranter. En signifikant forskel mellem runderne kan skyldes den store migranttilstrømning i netop 2015-2016 modsat 2017-2018, hvor medlemslandene, især Tyskland, Ungarn og Sverige, var utilfredse med EU's håndtering og beskyttelse af EU's grænser udadtil. Utilfredsheden førte til landenes varsling om komplet grænslukning, såfremt endnu en immigrationskrise skulle ramme uden klare omfordelingsregler. (Barigazzi, 2019) Som respons forudsagde Centralbanken et fuldstændigt kollaps af den frie mobilitet mellem landegrænser i EU og estimerede et fald i BNP på 8% i Tyskland og 7% i Italien som resultat på handelsbarrierer. (Draghi, 2019)

De endelige modeller viser at holdningen til immigranter bosættelse i landet er negativt korreleret med "Forblive i EU". En udbredt frygt for øget terrorisme, risiko for udfasning af kultur og traditioner og bekymring om flygtninge som økonomisk byrde kan muligvis lægges til grund. (Wike et al., 2016) I Tyskland dominerede en bekymring om integrationen af immigranter grundet en utilstrækkelig infrastruktur, herunder boliger og sprogkurser, og ineffektiv asylbehandling som resultat af den uforventede, massive tilstrømning. Den tyske befolkning var oprørt over den uretfærdige byrde relativt til mange andre EU-lande. (Keita og Dempster, 2020) I Italien kan den negative korrelation muligvis findes i italienernes utilfredshed med Dublin-forordningen, hvilken i årevis er forsøgt revideret for en mere lige omfordeling af immigranter på tværs af medlemslandene. (Villa, 2019) Som resultat frygtes den skattemæssige byrde på landet, løndumping og andre udfordringer på arbejdsmarkedet. (Simionescu, 2020)

Davidov and Meuleman (2012) viser at menneskelige værdier er bestemmende for holdninger til immigranter, hvor en mere positiv indstilling til immigranter kan findes i mennesker, der prioriterer kosmopolitiske værdier. I modsætning hertil opstår mere negative holdninger blandt individer, der vægter konformitets- og traditionsværdier højest. Davidov et al. (2014) argumenterer at kontekstuelle forhold er tilbøjelige til at påvirke intensiteten, hvormed disse værdier, der er relevante for holdninger til immigration, aktiveres. Mennesker, der lever under hård konkurrence om knappe ressourcer, herunder arbejde og sociale ydelser, er mere tilbøjelige til at opfatte immigranter som en trussel. (Billiet et al., 2014) I denne sammenhæng kan det post-kommunistiske traume forestilles at forklare den relativt fjendtlige tilgang til immigranter i

Polen på trods af marginale migrantstrømme til landet. Dette vedrører især immigranter fra kulturer og religioner diametralt forskellige fra den polske. Som bemærket af Sztompka (2000) modtog Polen komunismens fald med den største entusiasme. Ikke desto mindre, førte en systemændring til traumatogene oplevelser såsom arbejdsløshed, ændring i social position, fattigdom og en stigning i kriminalitet. Det post-kommunistiske traume har en indflydelse på landets adaptive kompetence, herunder til forandringer som modtagelse af immigranter.

Imidlertid udviser modellerne ligeledes, at tilfredshed med immigranteres indflydelse på landet er positivt korreleret med sandsynligheden for at forblive i EU og tildid til Europa-Parlamentet. De positive korrelationer kan ses at være forbundet med den hastige aldring af Europa og mangel på arbejdskraft grundet emigration i Polen og Italien. Siden Polen fik adgang til arbejdsmarkedet i andre EU-lande i 2004 oplevede landet en emigration af unge erhvervsdygtige individer på over 2,59 mio. frem til 2018. (Eurostat, 2021b) Til sammenligning har Italien, allerede karakteriseret ved en usædvanlig dårlig demografi, i perioden 2008-2019 oplevet et tab på human kapital på omrent 2 mio. uddannede unge. (Romei, 2019) Det blev i 2016 estimeret, at Eurozonens forsørgerbyrde, målt ved antallet af personer på 65 år og derover som procent af befolkningen i den erhvervsdygtige alder, vil stige fra lidt over 30% i 2016 til omrent 52% i 2070. For Euroområdet som helhed vil nettoimmigration delvist opveje faldet i den erhvervsaktive befolkningssandel. (Nerlich and Schroth, 2018) For at facilitere økonomisk vækst og erstatte pensionerede arbejdere i Tyskland estimerede Detlef Scheele, formanden for den statslige arbejdsmarkedsstyrelse, et årligt behov på 400.000 faglærte immigranter (Hagelüken, 2021).

Ikke desto mindre har bekymringer om immigration, forstørket af flygtningekrisen i 2015, sammenholdt med EU's udvidelse mod øst, finanskrisen i 2008 og Brexit folkeafstemningen i 2016 resulteret i en øget euroskepticisme, der har banet vej for nationalistiske og populistiske partier. (Henley, 2020) Udviklingen i euroskepticismen er relevant i henhold til variablen "placering på højre/venstre skala", på baggrund af den negative korrelation med "Forblive i EU" og "Tillid til Europa-Parlamentet". Denne tendens forekommer i overensstemmelse med den voksende popularitet af Alternative für Deutschland, AfD, der er et tysk yderst til højre, anti-immigrant parti etableret i 2013. AfD vandt i sit første Forbundsdagsvalg i 2017, 94 ud af 709 mandater ved primært at føre kampeagne på en anti-immigrationsplatform og endte med en tredjeplads. Samtidigt gik Merkels parti CDU/CSU ned med 65 mandater til 246. (Bundeswahlleiter, 2017) Det forekommer som en tendens, at mainstream partier kendtegnet som værende omkring midten, fastholder sin politiske position mens "udfordrerpartier" forsøger at differentiere sig ved strategisk at appellere til vælgernes bekymringer og herved opnå en valgmæssig fordel. (Vasilopoulou, 2018)

Samme udvikling observeres i både Italien og Polen. Det polske parti "Den Borgerlige Platform" (PO), ved magten i 2007-2015, accepterede at modtage flygtninge som en del af EU's omplaceringsordning. Det højreekstremistiske parti "Lov og Retfærdighed" (PiS), dengang i opposition, argumenterede dog imod at acceptere mellemøstlige flygtninge. Flygtningekrisen blev udnyttet af PiS under kampanjerne til præsident- og parlamentsvalget i 2015, der resulterede i en sejr til PiS. (Kinowska-Mazaraki, 2021) Den største trussel for Italiens fortsatte EU integration var Matteo Salvini højreekstreme nationalistiske og populistiske parti Lega (tidligere Lega Nord). Salvini's politiske strategi er at skabe en "os versus dem" retorik, hvor modstanderen er udlændinge og immigranter, der beskyldes for Italiens stigende arbejdsløshed i perioden 2008-2018 mens EU og euroen ansvarliggøres for den stigende statsgæld. (Mcginnis, 2021) I Italien er det ikke kun Lega, der udgør en trussel mod fortsat deltagelse i EU. 5-stjerne bevægelsen (M5S) oprettet af komikeren Beppe Grillo i 2009 placerer sig hverken til højre eller venstre på skalaen. Partiet ser sig selv som et mellemmed mellem befolkningen og politikerne, og forsøger at reducere regeringens og andre autoritære institutioners magt i landet, herunder EU.

I modsætning til tilfældet i Tyskland og Italien, sejrede det polske euroskeptiske parti Lov og Retfærdighed (PiS) i både præsident- og parlamentsvalget i 2015. PiS's sejr har været et vendepunkt i Polens nyere historie, hvor landet udgør et velstående demokrati og stolt EU-medlemsland. Partiet påstod at landets transformationsfase efter kommunismen har været styret af en korrupt pseudo-elite, der ikke støttede de traditionelle polske værdier forbundet med den katolske kirke, men har vestlige liberale idéer som kernesæt. Ifølge PiS er den katolske kirke essensen af "polskhed", der blev fremstillet som at være blevet forrådt af eliten. (Markowski, 2019) PiS har siden 2016 gradvist opløst magtens tredeling primært via en udskiftning af hele kadren i hovedpositioner, herunder lovligt valgte dommere i forfatningsdomstolen, hvilket implicerede at enhver kontrol over regeringens handlinger blev fjernet og ændringer kunne gennemføres uden forfatningsflertal. (Sadurski, 2019) Partiet har opnået majoriteten af vælgernes tillid, på trods af en langsom fratagelse af deres frihed, blandt andet ved at give familier "500 +", der er en månedlig støtte til den traditionelle kernefamilie på 500 PLN for hvert barn. PiS giver uforholdsmæssige økonomiske tilskud til støttende organisationer, herunder primært til ejeren af det ekstremkonervative katolske radio-netværk Fader Tadeusz Rydzyk, som i perioden 2016-2019 modtog omkring 43,5 mio. euro i statsstøtte i bytte for at sprede PiS's politik i kirken og opfordre til at stemme på PiS. (Kinowska-Mazaraki, 2021) Her er det relevant at nævne, at omtrent 90% af den polske befolkning anser sig som katolikker. (EC, 2021a) Desuden førte PiS en frygtdrevet kampagne der forstærkede farens og fjendtligheden overfor især muslimske immigranter og fremstillede sig selv som landets vogtere og beskyttere. (Narkowicz, 2018) Udnyttelsen af Smolensk-katastrofen fra 2010, hvor landets daværende præsident og elite omkom i et flystyrt, hjalp ligeledes PiS med at holde vælgerne forenet. En national tragedie, der er rodfæstet meget dybt i landets befolkning blev brugt som en politisk kampagne. (Shotter and Foy, 2020)

Partiet PiS har siden 2015 haft vælgerflertal, hvorved det må betyde at en stor del af befolkningen, der stemmer, deler partiets anti-europæiske syn, hvilket delvist kan forklare den negative korrelation mellem tillid til nationale institutioner og ønsket om at forblive i EU. Partiets politik har siden valgsejren udviklet sig i en mere højreekstremistisk retning, hvilket forklarer forekomsten af en signifikant forskel mellem runderne og det signifikant mere negative parameterestimat i 2018. I Italien forekommer forholdet dog kun signifikant på et 5%-niveau i runde 8, hvilket muligvis kan tilskrives en manglende følelse af stabilitet og tilhørsforhold. Den manglende stabilitet kan muligvis tilskrives de hyppige udskiftninger i regeringen (65 i perioden 1946-2018) baseret på interne stridigheder i koalitionerne, men ligeledes korruptionsskandalen i 90'erne, hvor højststående politikeres korruption blev afdækket af "Operation Rene Hænder" og politiske partier blev opløst. Korruptionsskandalen mærkede det italienske, politiske system og italienerens holdning til politikerne for evigt. (Kvittingen, 2021)

Robert D. Putnam argumenterer at social kapital, der i store træk defineres som den samlede mængde af tilgængelig tillid, er afgørende for regeringens præstation i de enkelte lande. (Claridge, 2015) Lande med et tilstrækkeligt niveau af social kapital formår i højere grad at overvinde problemer via en kollektiv handling og forfolgelse af fælles mål, herunder fælles konsensus for vacciner. Ud fra dette perspektiv er det af afgørende betydning at udlede, hvorfor tillid synes at trives bedre i nogle samfund end i andre. ESS (2009) ser en sammenhæng mellem tillid mellem mennesker og tillid til nationale institutioner og myndigheder. Når individer stoler på hinanden forekommer det mindre risikabelt og mere givende at deltage i sociale anliggender, der hjælper med at opbygge og fremme demokratiske institutioner. Der opstår en villighed til at opgive individuelle, umiddelbare fordele for langsigtede, samfundsmæssige gevinsten i troen på at andre ikke udnytter muligheden ved opportunistisk adfærd. Med de lave niveauer af tillidsvariablene set for Italien og Polen bundet i kontekstuelle faktorer forekommer det som en udfordring at opnå samme samfundssind som set i Tyskland.

Politisk aktivitet, målt ved stemmedeltagelse, viste sig at være positivt korreleret med "Forbliv i EU". Fri og retfærdige valg er et grundlæggende princip i ethvert demokratisk samfund, hvorved er sundt og robust demokrati frit tillader borgere at udtrykke sine synspunkter og have en indvirkning på sin fremtid. Udøvelse af sin stemmerettighed indikerer da en tro på og tilfredshed med landets demokrati og tænkes at være positivt korreleret med et ønske om fortsat EU-medlemskab, da den demokratiske rettighed ligger i hjertet af Den Europæiske Union. (EC, 2021b) Medlemskabet af EU fungerer ligeledes som et sikkerhedsnet for medlemslandenes borgere, da de er bevidste om EU's intervention i tilfælde af en overskridelse af rettigheder og demokratiske værdier.

Ikke desto mindre finder ESS (2009) at bedre uddannede og velstillede er mere tilbøjelige til at stemme, til dels da de er opsatte på at bibeholde status quo, er mere bevidste om væsentligheden af demokratisk deltagelse og er mere politisk aktive. Dette implicerer at individer, der har tillid til og udviser tilfredshed med den siddende regering i en højere grad stemmer. Denne tendens kan forestilles at blive forstærket i både Italien og Polen, da en relativt stor andel af populationen mener at det politiske system ikke tillader dem at have en indvirkning på, hvad regeringen foretager sig, hvilket resulterer i en endnu lavere stemmedeltagelse, jf. tabel 5.1. I Polen kan kommunismens rødder forestilles at spille ind via en dominerende følelse af magtesløshed resulterende fra en længerevarende undertrykkelse i et kollektivt samfund. Den polske befolkning har siden 1989 været præget af valgpassivitet, hvilket fremgår af en relativt begrænset valgdeltagelse, hvor omrent halvdelen af de stemmeberettigede deltager. Imidlertid observeres en fremgang, omend langsom da en systemændring ikke kan forventes at fjerne den formede mentalitet automatisk. (Kinowska-Mazaraki, 2021)

Tabel 5.1: Fordeling af variablen "*Det politiske system tillader folket at have indflydelse på, hvad regeringen foretager sig*" på lande og runder

Politisk indflydelse	Polen		Italien		Tyskland	
	2016	2018	2016	2018	2016	2018
Overhovedet ikke	30,37%	21,81%	48,78%	37,66%	14,31%	16,40%
Meget lidt	39,97%	38,17%	41,32%	47,51%	36,44%	36,11%
En smule	24,86%	32,08%	8,78%	13,30%	37,40%	34,72%
En del	3,96%	6,50%	1,00%	1,38%	10,86%	11,88%
Rigtig meget	0,84%	1,43%	0,12%	0,15%	1,00%	0,89%

Kilde: egne beregninger med udgangspunkt i ESS-data fra runde 8 og 9.

Anm.: beregninger er udført ved brug af analysevejten, anweight.

Den politiske ustabilitet, der har præget Italien i de seneste to årtier, tænkes at forklare den store andel, der ikke føler sig hørt. I Italien eksisterer et stort antal små partier og hyppige koalitionsudskiftninger i regeringen. Den store ustabilitet er opstået som et resultat af ændringen til det proportionelle stemmesystem i 1993, som havde til mål at gøre det politiske system mere volatilt, da stabiliteten hidtil havde været præget af længerevarende mandater i form af Mussolinis diktatur, og regeringerne først af Craxi og Berlusconi. Det har siden vist sig, at det nye system har mere politisk ustabilitet end initialet forestillet. Siden 1994 er der blevet afholdt 4 folkeafstemninger om valgsystemet, 3 ud af de 10 mest volatile valg i vest-Europa siden 1945 har været i Italien og siden 1989 har der været 3 tyske kanslere, 5 franske præsident og 7 engelske premierministre, hvor der til sammenligning har været 13 italienske premierministre. Relevant at nævne er den italienske "tradition" til at lade en ekspert udefra, en teknokrat, styre landet, når økonomien for alvor svigter, som var tilfældet i 2011 med Monti-regeringen. Alt sammen faktorer,

der medfører en vurdering af det italienske politiske system som værende meget ustabilt, i forhold til at Italien er den 4. største økonomi i EU.

En manglende deltagelse i demokratiet, der observeres i både Italien og, udover den kommunistiske arv, i Polen, kan da ligeledes være tegn på utilfredshed med og mistillid til den politiske virkelighed. Denne tendens hænger sammen med det relativt lave niveau af tillid til politiske partier jf. tabel 4.1.

6 Diskussion

Da ESS er et tværsnitstudie er det i analysen kun muligt at etablere signifikante forhold mellem de uafhængige variable og ønsket om et fortsat EU-medlemskab, samt tillid til EP. Et kausalt forhold kan derfor ikke bevises med foreliggende data.

Indværende datasæt indikerer et mønster konsistent for Italien, Polen og Tyskland. De betragtede lande udtrykker bestemt en højere støtte for et fortsat EU-medlemskab end et exit, og i 2016 var der en større befolkningsandel, der, i tilfælde af en folkeafstemning om medlemskabet, ville stemme forlad end tilfældet var i 2018, jf. tabel 6.1. Forskellen mellem årene bemærkes at være størst for Italien med et stigning i støtten på 7 pct. point, efterfulgt af Tyskland med 2 pct. point og mindst for Polen. Imidlertid befandt Polen og Tyskland i udgangspunktet sig på et markant højere niveau.

Tabel 6.1: Fordeling af variablen "*Såfremt der i morgen var en folkeafstemning om EU-medlemskabet, ville du stemme forbliv eller forlad?*" på lande i 2016 og 2018.

EU-medlemskab	Polen		Italien		Tyskland	
	2016	2018	2016	2018	2016	2018
Forlad	12,35%	12,33%	35,44%	28,43%	15,07%	13,27%
Forbliv	87,65%	87,67%	64,56%	71,57%	84,93%	86,73%

Kilde: egne beregninger med udgangspunkt i ESS-data fra runde 8 (2016) og 9 (2018).

Anm.: beregninger er udført ved brug af analysevejgen, anweight.

Mens landene på mange områder er vidt forskellige udviser modellerne en fælles konsensus om de primære problemstillinger og fordele, der påvirker borgernes holdning til medlemskabet og tilliden til Europa-Parlamentet. (EP, 2018) Modellerne indikerer, at den absolut største udfordring for medlemskabet er immigration. Sammenholdt med flygtningekrisen i 2015 og 2016, hvor flygtningetilstrømningen var på sit højeste, kan den positive ændring i holdningen til medlemskabet muligvis forklares af det aftagende antal flygtninge, hvilket giver belæg for, at håndteringen af immigrationsspørgsmålet burde være blandt de øverste problemstillinger på dagsordenen.

Borgeres meningsdannelse afhænger af adskillige faktorer, hvorved de estimerede effekter ikke implicerer, at hvert land direkte berøres af immigration. Denne observation forekommer relevant i Polens tilfælde, hvor landet kun modtog 0,8% af det samlede antal asylansøgninger i EU i 2015. (Eurostat, 2021a) Ikke desto mindre, blev masseimmigration hovedelementet i valgkampagnerne i 2015 og skabte et øget had og diskrimination overfor etniske minoriteter blandt den polske befolkning, hvilket resulterede i en stigning i hadforbrydelser. Politikken om immigranter som en "imaginær fjende" førte til en opfattelse af at andelen af immigranter udgør mere end 10% af den samlede befolkning, mens det reelle tal var under 1%. En strategisk udnyttelse af frygt ved at forbinde truslen om fare med det man har mest kært, herunder familie, vækker så dybe følelser at rationalitet skubbes til siden. (Kinowska-Mazaraki, 2021)

Italien derimod, er et land der er direkte eksponeret overfor immigranter, da mange flygtninges første

møde med Europa var Italien, taget landets geografiske position midt i Middelhavet i betragtning. Trods intentioner om vidererejse, oftest med håb om at nå Tyskland, og Italiens egen videresendelse af flygtninge, blev mange sendt retur til Italien ved brug af Dublin-forordningen. Flygtningestrømme til Italien gennem Middelhavet var en forudsigelig begivenhed, men alligevel lykkedes det ikke for Italien efter gentagne forsøg siden starten af 90'erne at få emnet på dagsordenen i den Europæiske Kommission. I mange år var landet et af foregangslandene for et bredt EU-samarbejde og blandt de lande i EU, der var mest positivt stemt overfor immigranter. Men efter landet har været nødsaget til at modtage ukontrollerede bølger af immigration uden hjælp fra EU, har der været en gængs følelse af forrådthed i Italien. Samarbejdet mellem den italienske regering og EU har længe været præget af mistillid. Siden Berlusconi-regeringen i starten af det nye årtusinde beskyldte EU for nationale problemer, har det været et gennemgående mønster i de italienske regeringer, både regeringer skeptiske og ikke-skeptiske overfor EU, at bruge EU som syndebuk for interne problemer, og flygtningekrisen i 2015 har bestemt ikke bidraget positivt til det anspændte forhold. (Balfour and Robustelli, 2019)

Immigration er da særskilt en gennemtrængende udfordring, men giver sig ligeledes til udtryk i andre faktorer, herunder tillidskapital og placeringen på den politiske skala, via en fremgang i partier med anti-europæisk agenda. En stigende tendens til utilfredshed, især i et stort og EU-medstiftende land, vil sandsynligvis udvides til andre politiske spørøgsmål og medlemslande, hvorfor en intervention ikke kun ville være i Italiens, men hele Unionens interesse.

I tabel 4.1 fremgår det at Polen og Italien udviser nogle af de laveste niveauer af tillid blandt Europas lande. Den lave sociale kapital kan blandt andet ses på, hvordan tidlige regeringer har forsøgt at gøre op med nogle af de udfordringer landene står over for, men som konsekvens mistede styret. I Italien forsøgte Mario Monti som teknokrat at indføre reformer, der havde til mål at gøre landets finansielle situation mere langsigtet, og på trods af at Monti oprindeligt var velmodtaget, var han nødsaget til at trække sig halvandet år senere. Den daværende polske regering med PO som spidsparti forsøgte ligeledes tilbage under flygtningekrisen at få Polen til at tage imod kvoteflygtningene, men også her gav befolkningen sig tydeligt til udtryk, ved at stemme PiS ind ved valget i 2015, der garanterede at ingen flygtninge ville komme til landet, hvis de dannede regering. Begge tilfælde viser, hvordan lav social kapital, som set ved blandt andet lav tillid til myndigheder og institutioner, gør det svært for befolkningen i Polen og Italien at ofre sig på kort sigt for at opnå langsigtede fordele. Samme forventes at være tilfældet ved reformer, der kræver en ofring af øjeblikkelige fordele for "the common good" i fremtiden.

7 Konklusion

I de senere år har Europa oplevet en fremgang i højreekstremistiske partier, der i stigende grad udviser euroskeptisk diskurs. En øget spænding mellem gevinsterne af økonomisk integration og landenes politiske samarbejde er sædvanligvis drevet af troen på, at der er en iboende afvejning mellem EU-medlemskab og landenes særskilte evne til at udøve suverænitet.

I indeværende inddrag var det til hensigt at undersøge holdningen til EU-medlemskabet og identificere forhold af særlig indflydelse, i Italien og Polen, holdt op imod Tyskland. Undersøgelsen blev suppleret med en analyse af Tillid til Europa-Parlamentet både som en indikator for medlemskabet og for at observere eventuelle forskelle ved at lade respondenterne have større mulighed for at udtrykke sig. Data udviste, at omrent 85% af befolkningen i Polen og Tyskland ville stemme "Ja" ved en folkeafstemning om et fortsat EU-medlemskab, hvor det til sammenligning var omkring 70% i Italien. Det udførlige dataarbejde og stikprøvedesign bag ESS, har ved anvendelse af analysevejten sikret repræsentativitet, der giver

mulighed for at foretage generelle slutninger om landenes befolkning. For en effektiv udnyttelse af ESS' omfattende mængde spørgsmål, har en anvendelse af faktoreanalyse, ved kombination af flere variable, tilladt at afdække holdningsmønstre med relevans for holdningen til EU.

De endelige modeller udviste at immigrationsfaktorerne og tillid til henholdsvis nationale institutioner og myndigheder, var de mest signifikante variable og derved havde størst indflydelse på ønsket om at forblive i EU. Ligeledes fremgik det at tillid til Europa-Parlamentet egner sig som en indikation på, om respondenten bevæger sig i en retning, der er for eller imod EU-medlemskabet, da majoriteten af variablene i den logistiske og lineære model var tilfælles, og parametrene havde samme fortegn. Trods betydelige forskelle mellem landene, påvirkede variablene i den logistiske model holdningen til EU i samme retning, med undtagelse af tillid til nationale institutioner, der havde en modsatrettet korrelation for Polen i 2016, som blev sammenkædet med at Polen i 2016 var det eneste af de tre fokuslande, der havde en euroskeptisk regering.

Litteratur

- Balfour, R. and Robustelli, L. (2019). *Why Did Italy Fall Out of Love with Europe?* The international spectator or Instituto Affari Internazionali.
- Barigazzi, J. (2019). *What Europe is really fighting about.* Politico.
- Billiet, J., Meuleman, B., and Davidov, E. (2014). *The relationship between ethnic threat and economic insecurity in times of economic crisis: Analysis of European Social Survey data.* ESS. DOI: <https://doi.org/10.1093/migration/mnu023>.
- Bundeswahlleiter (2017). *Bundestag election 2017.* Bundeswahlleiter.
- Claridge, T. (2015). *Putnam on social capital – democratic or civic perspective.* Social capital research.
- Cohen, R. (2015). *Germany, Refugee Nation.* The New York Times.
- Davidov, E. and Meuleman, B. (2012). *Explaining Attitudes Towards Immigration Policies in European Countries: The Role of Human Values.* Journal of Ethnic and Migration Studies. DOI: <https://doi.org/10.1080/1369183X.2012.667985>.
- Davidov, E., Meulemann, B., Schwartz, S. H., and Schmidt, P. (2014). *Individual values, cultural embeddedness, and anti-immigration sentiments: Explaining differences in the effect of values on attitudes toward immigration across Europe.* KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie. DOI: <https://doi.org/10.1007/s11577-014-0274-5>.
- Draghi, M. (2019). *Speech by Mario Draghi, President of the ECB, on the award of Laurea honoris causa in law from Università degli Studi di Bologna, Bologna, 22 February 2019.* European Central Bank.
- EC (2021a). *Poland. Population: demographic situation, languages and religions.* European Commission.
- EC (2021b). *Reinforcing democracy and integrity of elections.* European Commission.
- ECJ (2020). *By refusing to comply with the temporary mechanism for the relocation of applicants for international protection, Poland, Hungary and the Czech Republic have failed to fulfil their obligations under European Union law.* Court of Justice of the European Union.
- EP (2018). *Brexit effect: Public opinion survey shows that EU is more appreciated than ever.* European Parliament.

- ESS (2009). *Exploring public attitudes, informing public policy. Selected findings from the first three rounds.* ESS.
- ESS (2021a). *ESS8 - 2016 Notes on data and fieldwork.* ESS.
- ESS (2021b). *ESS9 - 2018 Notes on data and fieldwork.* ESS.
- ESS (2021c). *Sampling.* ESS.
- Eurostat (2021a). *Asylum and first time asylum applicants - annual aggregated data (rounded).* TPS00191. EC.
- Eurostat (2021b). *Emigration by age and sex tabel MIGR_EMI2.*
- Fitzgerald, R. and Sibley, E. (2015). *The United Kingdom in the European context – top-line reflections from the European Social Survey.* ESS.
- Guillot, J. D. and Corlett, N. (2018). *Brexit effect: Public opinion survey shows that EU is more appreciated than ever.* European Parliament.
- Hagelüken, A. (2021). "Wir brauchen 400 000 Zuwanderer pro Jahr". SZ.
- Henley, J. (2020). *Support for Eurosceptic parties doubles in two decades across EU.* The Guardian.
- Keita, S. and Dempster, H. (2020). *Five Years Later, One Million Refugees Are Thriving in Germany.* CGDEV.
- Kinowska-Mazaraki, Z. (2021). *The Polish Paradox: From a Fight for Democracy to the Political Radicalization and Social Exclusion.* MDPI.
- Kvittingen, I. (2021). *Korruptionsskandale kan have mørket unge italienske vælgere for livet.* Videnskab.dk.
- Lynn, P. (2021). *Mail fra Peter Lynn.*
- Markowski, R. (2019). *Creating Authoritarian Clientelism: Poland After 2015.* Hague Journal on the Rule of Law. DOI: <https://doi.org/10.1007/s40803-018-0082-5>.
- McGinnis, D. P. (2021). *Anti-Immigrant Populism in Italy: An Analysis of Matteo Salvini's Strategy to Push Italy's Immigration Policy to the Far Right.* The Yale Review of International Studies.
- Narkowicz, K. (2018). "Refugees Not Welcome Here": State, Church and Civil Society Responses to the Refugee Crisis in Poland. International Journal of Politics, Culture, and Society. DOI: <https://doi.org/10.1007/s10767-018-9287-9>.
- Nerlich, C. and Schroth, J. (2018). *The economic impact of population ageing and pension reforms.* ECB.
- OECD (2013). *Government at a Glance 2013.* OECD.
- O'Rourke, N. and Hatcher, L. (2014). *A Step-by-Step Approach to Using SAS for Factor Analysis and Structural Equation Modeling.* SAS.
- Romei, V. (2019). *Record Italian emigration and low birth rate drive economic fears.* Financial Times.
- Sadurski, W. (2019). *Polish Constitutional Tribunal Under PiS: From an Activist Court, to a Paralysed Tribunal, to a Governmental Enabler.* Hague Journal on the Rule of Law. DOI: <https://doi.org/10.1007/s40803-018-0078-1>.

- Shotter, J. and Foy, H. (2020). *Ten years after the Smolensk air disaster, political scars remain*. Financial Times.
- Simionescu, M. (2020). *Italexit and the Impact of Immigrants from Italy on the Italian Labor Market*. MDPI. DOI: <https://doi.org/10.3390/jrfm14010014>.
- Sztompka, P. (2000). *Trauma wielkiej zmiany. Społeczne koszty transformacji*. ISP.
- Vasilopoulou, S. (2018). *The party politics of Euroscepticism in times of crisis: The case of Greece*. Sage Journals. DOI: <https://doi.org/10.1177/0263395718770599>.
- Villa, M. (2019). *Italy receives more asylum seekers from Germany than from Libya*. Politico.
- Wike, R., Poushter, J., Silver, L., Devlin, K., Fetterolf, J., Castillo, A., and Huang, C. (2019). *Views on the European Union across Europe*. Pew Research Center.
- Wike, R., Stokes, B., and Simmons, K. (2016). *Europeans Fear Wave of Refugees Will Mean More Terrorism, Fewer Jobs*. Pew Research Center.

En statistisk tilgang til Spritbådene mellem Danmark og Tyskland i et historisk perspektiv

Lars Holmgård Bak

Rigsarkivet, Aabenraa, Danmark

&

Nils Karl Sørensen¹

Økonomisk Institut, Syddansk Universitet, nks@sam.sdu.dk

Resumé

Flensborg fjord, der danner grænsen mellem Danmark og Tyskland var indtil ændringer i EU-lovgivningen klassificeret som internationalt farvand uden varebeskatning. Hård beskatning af alkohol, tobaksvarer og lignende gav grund for en rederibranche baseret på toldfrit salg med livlig trafik over fjorden.

Indledningsvis opstilles en mikroøkonomisk model for salget af toldfri produkter. Dernæst analyseres udviklingen baseret på arkivmateriale fra flere perioder. Det drejer sig om årvise og månedsvise observationer af passagerer og salgsstatistik fra perioden fra 1963 til 1969. Endvidere tegnes med udgangspunkt i et materiale fra begyndelsen af 1980erne en profil af den typiske spritbådspassager. Samlet opnås et tidsbillede af en forgangen epoke, hvor varebeskatningen i høj grad påvirkede forbrugernes adfærd.

Det findes, at elasticiteten på de produkter, der var afgiftsbelagte, var meget høj. Antallet af passagerer varierede især med dels afgifter og andre reguleringer dels med sæson og højtider. Endelig var den typiske passager en midaldrende person med moderat indkomst.

Modeller for social genetiske effekter hos grise og betydningen for dyrevelfærd

Bjarne Nielsen, Quantitative Genetics and Genomics, Aarhus University, Lizette V. Horndrup, Danish Agriculture & Food Council, and Birgitte Ask, SEGES, Breeding and Genetics

Grise er sociale dyr, som interagerer med de dyr, som de går sammen med. Der findes grise, som gør noget godt for de andre grise, og der findes grise, som har en negativ indflydelse på andre grise. I arbejdet med at øge dyrevelfærden i griseproduktionen og forbedre produktiviteten har forskere fra danske SEGES Svineproduktion, Aarhus Universitet og SRUC (Scotland's Rural College) undersøgt, om de sociale egenskaber hos grise kan forbedres gennemavl. Resultaterne tyder på, at det faktisk er muligt. Forskningen er gennemført med støtte fra GUDP (Grønt Udviklings- og Demonstrationsprogram) under Landbrugs- og Fiskeristyrelsen. Målet har været at udvikle en metode til at beregne en social avlsværdi til avlsgrise og dermed udvælge de grise med de bedste sociale gener som forældre til fremtidens grise for avlsselskabet DanBred.

Den generelle model for beskrivelse af sociale effekter i husdyrhold er som flg.

$$y_{ikm} = \mu_{ikm} + a_{ikm} + \sum_{j \neq f(i,k,m)}^{n_k} u_{jk} + \gamma_k + l_m + e_{ikm}$$

hvor y_{ikm} er fænotypen for et dyr i i gruppe $k = \{1, 2, \dots, K\}$, fra kuld $m = \{1, 2, \dots, M\}$; a_{ikm} er det genetiske potentielle af dyr $i = \{1, 2, \dots, n_{km}\}$; n_{km} er antallet af søskende fra kuld m i gruppe k ; u_{jk} er det socialgenetiske potentielle af gruppefælderne $j = \{1, 2, \dots, n_k\}$ i gruppe k med n_k individer; sammenhængen mellem i og j er givet af funktionen f hvor $j = f(i, k, m)$; γ_k er den fælles miljøeffekt for alle dyr i gruppe k ; l_m er den fælles effekt af søskende, som er født i kuld m og e_{ikm} er residualeffekten. Det antages at

$$\begin{pmatrix} a_{ikm} \\ u_{jk} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \sigma_{au} \\ \sigma_{au} & \sigma_u^2 \end{bmatrix} \right), \quad \gamma_k \sim N(0, \sigma_\gamma^2), \quad l_m \sim N(0, \sigma_l^2), \text{ og} \\ e_{ikm} \sim N(0, \sigma_e^2).$$

I omtalte projekt har vi anvendt ovenstående model på alle avlsrelevante dyr i avlspopulationen, hvilket er mellem 100.000 dyr og 1 mio. dyr for en given race. Som fænotype, y_{ikm} anvendte vi registerdata af tilvæksten, der måles i gram per dag og er målt på alle avlsrelevante grise i vækstperioden fra at grisene vejer 30kg til de vejer 100 kg. Som estimationsprocedure anvendte vi REML¹. Derefter kunne vi identificere to grupper af dyr. Dyr med henholdsvis højt og lav værdi af prædikteret \hat{u}_{jk} . For at undersøge effekten af den social genetiske komponent, \hat{u}_{jk} blev der etableret afkom af de to grupper. Samlet i de to grupper blev der produceret mere end 4728 afkom. En mere omfattende beskrivelse data og projekt findes i Ask et al (2021)²

Hypotesen var, at afkom efter grise med høj \hat{u}_{jk} havde en lavere risiko for negativ social adfærd end afkom efter grise med lav \hat{u}_{jk} . Med andre ord, at afkom af grise med en favorabel effekt på deres stiftellers tilvækst netop er grise, som viser en relativt lav grad af negativ social adfærd overfor deres stiftæller.

Afkom af de to grupper blev undersøgt for forskellig negativ social adfærd, som bl.a. antal rifter på ørene efter bid, og desuden grisenes reaktion på at komme ind i en vægt for at blive vejet.

Adfærdsdata blev bl.a. analyseret med en generaliseret lineær model af målingerne for de enkelte grise.

Lad Y være et kategorisk og ordnet adfærdsegenskab med k kategorier og lad $p(Y \leq k)$ være den kumulative sandsynlighed for at Y er mindre end eller lig med en specifik kategori $k = \{1, 2, \dots, k - 1\}$. Sandsynligheden for at være mindre end eller lig med en kategori k blev bestemt af

$$\log\left(\frac{p(Y \leq k)}{p(Y > k)}\right) = \text{logit}(p(Y \leq k)).$$

¹ Madsen P, Jensen J. A user's guide to DMU. Aarhus University, Center for Quantitative Genetics and Genomics. 2013. <https://dmu.ghpc.au.dk/DMU/Doc/Curre nt/>. Accessed 3 Jan 2021.

² Ask B, Pedersen LV, Christensen OF, Nielsen HM, Turner SP, Nielsen B. Selection for social genetic effects in purebreds increases growth in crossbreds. *Genetics Selection Evolution*. 2021;53(1)

Lad derefter y_i være den observerede kategori af Y for gris i . Derved blev den ordinære logistiske regressionsmodel

$$\text{logit}(p(y_i \leq k)) = \mathbf{x}_i' \mathbf{b} + \beta_a \hat{a}_i + \beta_u \hat{u}_i + \gamma_j$$

hvor $\mathbf{x}_i' \mathbf{b}$ korrigere for de systematiske miljøeffekter i vektoren \mathbf{x}_i målt på gris i . Miljøeffekterne var bl.a. er givet ved årstid, sektion i besætning, gruppestørrelse, gennemsnitsalder i gruppen og grisens egen vægt. Dernæst har vi regressionsfaktorerne β_a og β_u for de prædikterede genetiske effekter af avlsværdi for tilvækst, \hat{a}_i og social genetiske avlsværdi for tilvækst \hat{u}_i . De resterende tilfældige miljø-gruppeeffekt af gruppe j er givet ved $\gamma_j \sim N(0, \sigma_\gamma^2)$.

Modellerne blev estimeret med Maximum Likelihood ved at brug af R-pakken Ordinal³.

Resultater

Analyserne viste, at grupper med øget genetisk potentiale for social avlsværdi, dvs. øget værdi af \hat{u}_i havde lavere forekomst af ørerifter end øvrige grupper.

Regressionskvotienten blev bestemt til -0,094 ($p=0,021$). Ligeledes var der gunstig effekt af social avlsværdi på sværhedsgraden af ørerifter, hvor regressionskvotienten blev estimeret til -0,22 ($p=0,05$).

Der var også en positiv effekt af social avlsværdi, når grisene skulle vejes ved omkring 90 kg. Grise i grupper med høj social avlsværdi var mindre aktive under vejningen, hvilket af nogle adfærdsforskere tolkes som mindre angstfyldt adfærd.

Regressionskvotienten blev estimeret til -0,146 ($p=0,027$).

Tabel 1: Regressionseffekter af genetisk avlsværdi for tilvækst og social effekt relateret til tilvækst samt tilfældig gruppevarians for tre forskellige egenskaber.

Egenskab	Antal ørerifter		\hat{u}_i		\hat{a}_i		Gruppe varians
	Estimat	p-værdi	Estimat	p-værdi	Estimat	p-værdi	
Antal rifter på øre, ved ca.34 kg	-	-	-0.094	0.014	0.004	0.021	0
Sværhedsgraden af ørerifter betinget på antal ørerifter, ved ca.34 kg	0.518	<0.0001	-0.22	0.05	-0.0052	0.25	1.24
Aktivitet i vejekasse, ved ca.90 kg	-	-	-0.146	0.027	-0.007	0.008	0.34

³Christensen R. “Ordinal—Regression Models for Ordinal Data.” R package version 2018.8-25, <http://www.cran.r-project.org/package=ordinal/>. 2018

Deck height during transport of weaners – piglet height and microclimatic conditions inside the vehicle

Leslie Foldager^{1,2*}, Marianne Kaiser¹, Guoxing Chen³, Line D. Jensen¹, Jens Kr. Kristensen⁴, Cecilie Kobek-Kjeldager¹, Karen Thodberg¹, Li Rong³, Mette S. Herskin¹

¹ Dept. of Animal Science, Aarhus University, DK8830 Tjele

² Bioinformatics Research Centre, Aarhus University, DK8000 Aarhus C

³ Dept. of Civil and Architectural Engineering, Aarhus University, DK8000 Aarhus C

⁴ Dept. of Electrical and Computer Engineering, Aarhus University, DK8830 Tjele

* Email: leslie@anis.au.dk

Abstract

The present study was part of a policy support request commissioned by the Ministry of Environment and Food of Denmark, focusing on deck height when transporting weaner pigs on long and short journeys. Former legislation on deck heights have been based on an equation from Vorup and Barton-Gade (1991) relating live weight and average height based on 87 pigs (25-160 kg) and 21 sows (130-260 kg), including only 16 pigs in 26-40 kg. Part 1 of the study determined height of pigs weighing approximately 5-40 kg and aimed to establish new or confirm validity of old equations predicting height from live body weight in this weight interval. Prediction equations were built using 1435 pigs (47% castrates, 53% females) from 9 Danish herds and validated on 179 pigs (54% castrates, 46% females) from a 10th herd. The equation from Vorup and Barton-Gade (1991) systematically overfitted the height of these pigs, whereas an equation developed by Condotta et al. (2018) using 150 pigs (age 4-20 weeks) predicted equally well as models developed in the present study. Part 2 of the study examined microclimatic conditions during transport of pigs weighing 20-25 kg inside two decks on the lorry (deck heights 70 and 90 cm) and two on the trailer (deck heights 60 and 80 cm). Sixteen journeys for each of 8 and 23 h duration were done in Denmark from June 2021 to March 2022, ensuring transport under varying weather conditions. Data consisted of GPS tracking (position and speed), how much shutters were open (cm), use of mechanical ventilation (0/1), weather information, temperature from two sensors outside lorry and trailer, and temperature, relative humidity and CO₂ from sensors placed inside each of the 12 compartments, 3 on each deck with pigs. As per contract, results are not allowed to be revealed before handing in the advisory report to the ministry in July. In addition to results from part 1, results from analyses of relations between deck height, microclimate and external factors will be presented at the symposium and in the proceeding.

Introduction

The research presented in this paper was part of a policy support request commissioned by the Ministry of Environment and Food of Denmark, focusing on deck height when transporting weaner pigs. An advisory report answering this was handed in to the ministry July 5, 2022, and an extended summary (in Danish) was published 10 days later as an advisory memo from DCA (Herskin et al., 2022). The aim was to clarify: "For pigs weighing 20-25 kg, and transported according to the requirements of short journeys (up to 8 hours excluding loading and unloading time) as well as long journeys (up to 23

hours excluding loading and unloading time), at what deck height is the wording of the Council Regulation (Council Regulation (EC) No. 1/2005), (adequate ventilation above the animals when they are in a naturally standing position, without on any account hindering their natural movement), lived up to.”

The present paper only focus on: 1) the relation between pig weight and height of pigs in the weight interval 5-40 kg and 2) microclimatic conditions inside compartments containing pigs during experimental journeys of pigs with an average weight of 20-25 kg. The former of these two points was carried out during the first phase of the study during which e.g., the experimental design was decided, methodology and equipment for collection of data in multi-deck vehicles of relatively low deck height was developed and contracts were signed with external partners like the pig producer, the trading company and the transport company.

One task during the first phase of the study was to decide which deck heights to examine during the experiment. Former recommendations on deck heights had been based on an equation from Vorup and Barton-Gade (1991) relating height of pigs to live body weight on basis of 87 pigs (25-160 kg) and 21 sows (130-260 kg), including only 16 pigs in 26-40 kg and none in the range of interest, 20-25 kg. The scientific literature on the height-weight relation seemed scarce and we therefore decided to conduct an experimental trial to establish new equations for pigs weighing approximately 5-40 kg or confirm validity of old equations when predicting height from live body weight in this weight interval. The outcome of this study will soon be submitted to an international journal but preliminary results were presented in an advisory memorandum to the ministry (Foldager et al., 2021), handed in and published 17 December 2021.

Manuscripts presenting various parts of the second phase are now being prepared for submission to international journals so to safeguard this process, the present paper will mainly extract relevant parts of the memorandum (Foldager et al., 2021) and of the Danish advisory memo (Herskin et al., 2022).

Material and methods

First phase – measurement of weight and height

The experimental trial during the first phase of the study was carried out from October 2021 to February 2021 on 10 Danish pig herds with production of castrated and female weaners. The herds were balanced between the two major breeding systems in Denmark; DanBred (DB) and Danish Genetics (DG). Moreover, the herds were selected with the intent to take into account an anticipated distribution of management practice where two out of three Danish pig producers buy their gilts while the last third breed their own gilts ('homebred'). Nevertheless, the combination DG-homebred could not be found, possibly because DG is a relatively new constellation. Origin of semen was not taken into account during the selection of herds. The pigs were randomly selected following a

procedure that ensured good coverage across ages and thus weights and heights, and with five pigs from each of the selected pens. In each herd, the measurements were collected over two consecutive days and in addition, the sex of the pig was determined.

Herd 1 was used for setting up the experiments but once procedures had been settled, a final sampling was completed as for the other herds. This sample consisted of 179 pigs (46% females) weighing 4.7-46.0 kg (avg.=19.8; SD=10.34) with a height of 24.1-57.2 cm (avg.=38.7; SD=7.54). Since some fine-tuning hereafter was found relevant to e.g., avoid measuring pigs weighing more than 40 kg, this final sample was set aside for validation of prediction equations. During validation, six pigs from herd 1 that weighed 41.0 to 46.0 kg were used separately. The remaining 1435 pigs from herds 2 to 10 (53% females; range: 45-70%) weighing 2.8-40.6 kg (avg.=16.7, range: 13.4-23.1; SD=9.43, range: 6.11-11.64) with a height of 19.3-55.2 cm (avg.=36.4, range: 33.3-40.6; SD=8.25, range: 4.99-9.17) were used as training data for development of these equations.

Second phase – microclimatic conditions during transport

The second phase of the study was carried out from June 2021 to March 2022 using two parallel block-randomised semi-controlled field trials; one for 8 h (short) journeys and one for 23 h (long) journeys. Both trials involved pigs from the same herd and using the same commercially certified transport vehicle, a 4-deck lorry with a 5-deck trailer shown in Fig. 1. During experimental journeys, pigs were only carried on the first and third deck (counted from the bottom).

The vehicle was customised so that commercial deck heights of roughly 60 and 70 cm in the trailer and lorry could be compared with an increase of approximately 20 cm to around 80 and 90 cm, respectively. Adaptations ensured that each deck height on the trailer (60/80 cm) and on the lorry (70/90 cm) could be placed on both first and third deck; giving four combinations for the vehicle. The floor on the third deck was adapted so that placements of mechanical ventilator fans and shutters for natural ventilation were comparable between first and third deck within lorry and within trailer. The usual commercial separation of the deck into three 5.4-5.9 m² compartments holding pigs was applied. To follow regulations, the number of pigs per compartment was 30-40 for short journeys and 31-36 for long journeys; varying due to differences in compartment size, average body weight and different space requirements for short and long transport.

The journeys were randomised such that all four possible combinations of deck heights and vertical placement (first/third deck) were used for both journey lengths within each of four blocks; two in the warmer period of the year (block 1 and 2 in June-September) and two in the colder period of the year (block 3 and 4: October-March). All 32 journeys (16 short and 16 long) were carried out in Jutland, Denmark, using mainly highway and as far as possible the same routes for each journey of same length. The results presented in the current paper concerns only the period between departure from the herd and until

arrival at a commercial assembly barn. After some further examination and recording at the assembly barn, the pigs were no longer involved in the scientific study, and exported as part of normal commercial practice.

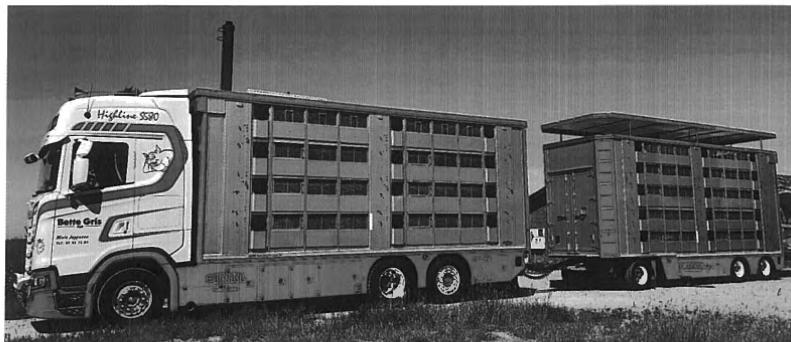


Figure 1 Photo of the vehicle (4-deck lorry with 5-deck trailer) used for experimental trials of the second phase of the study. The photo shows a situation during the process of loading pigs onto the trailer with roof lifted, shutters for natural ventilation fully opened and separate shutters in front of the mechanical ventilation fans also pulled down. (Photo: AU)

Measurement of microclimatic conditions

Complementary to the data usually collected during commercial pig transport (see below), data was logged for how much shutters were open (cm), outside temperature from two sensors on the side of the lorry and under the bottom of the trailer, and inside temperature, relative humidity and CO₂ from sensors placed in each of the 12 compartments, 3 on each deck with pigs. An extra temperature sensor (not considered in the present paper) was placed in each of the four rear-end compartments during the last four journeys. Fig. 2 shows a sketch of the experimental setup of the vehicle.

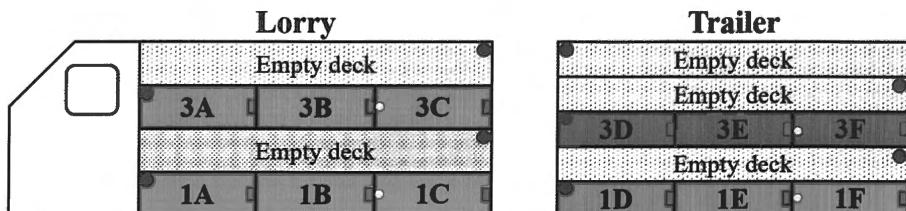


Figure 2 Sketch of the experimental setup of the vehicle (4-deck lorry with 5-deck trailer) and the placement of built-in temperature sensors (purple bullets), study loggers with temperature and CO₂ sensors (yellow rectangles) and extra temperature sensors (yellow bullets) used only for the last four journeys.

In addition to study sensors, built-in sensors indicating use/opening of doors and roofs, use of mechanical ventilation (0/1) and inside temperature (one sensor per deck; Fig. 2) were attached to a DataCold 600 logger (Carrier Transicold Europe S.C.S., France). Recordings of shutter opening were also captured by this logger by a customised solution provided by Locus A/S (www.locus.dk) and Carrier Transicold Scandinavia A/S (www.carrier.com/truck-trailer/). These companies also assisted with retrieving data from the global positioning system (GPS) equipment (LinkItAll) tracking position, distance travelled and average speed of the vehicle, which is mandatory along with the temperature logging at 15-min intervals. With exception of two journeys in October, however, this information was provided at 5-min intervals for this study. The driver of the vehicle followed industry practice, and based on data from the built-in sensors, as well as knowledge and experience, adjusted opening/closing of ventilation shutters, as well as use of mechanical ventilation, and, when stationary, could also lift the roof.

Weather data were retrieved from OpenWeather (<https://openweathermap.org>) via Python scripts on basis of the GPS position. Thus, weather data were available at the same timestamps as GPS positions with exception from the first two journeys where only data from the location at the beginning of a clock hour had been stored. Nevertheless, weather data were hourly averages at the specific location and thus on a cruder scale than the information from sensors. To enable merging on a finer grid, weather information was carried backwards from current timestamp to the minute after last timestamp (or 5 min), i.e. covering the same period as the average vehicle speed.

Statistical analyses

Statistical analyses were carried out using R version 4.1.2 (R Core Team, 2021) and a significance level of 0.05. Linear and nonlinear mixed effects models were estimated with the *lme* and *nlme* functions, respectively, from the *nlme* package v. 3.1-153.

Height-weight relation

The equation developed by Vorup and Barton-Gade (1991) for prediction of height from live body weight was the following second order polynomial

$$f_{V1991}(W) = 38.8639 + 0.4272 \cdot W - 0.000838 \cdot W^2 \quad (1)$$

In the advisory memorandum (Foldager et al., 2021), two model types were examined: 1) polynomial regressions and 2) nonlinear growth curve models. Later we became aware of a study by Condotta et al. (2018), evaluating and updating physical dimension data for grow-finish stages of modern pigs including standards published by ASABE (2020). Based on 150 growing-finishing pigs in the age range 4-20 weeks, Condotta et al. (2018) developed the following power function for predicting height from weight

$$f_{C2018}(W) = 14.4 \cdot W^{0.35} \quad (2)$$

Power function equations are obtained by exponentially back-transforming log-linear models and is a third type of models considered in the current paper. The height-weight data from the 1614 pigs are shown in Fig. 3 along with the predictions by Eq. 1 and 2.

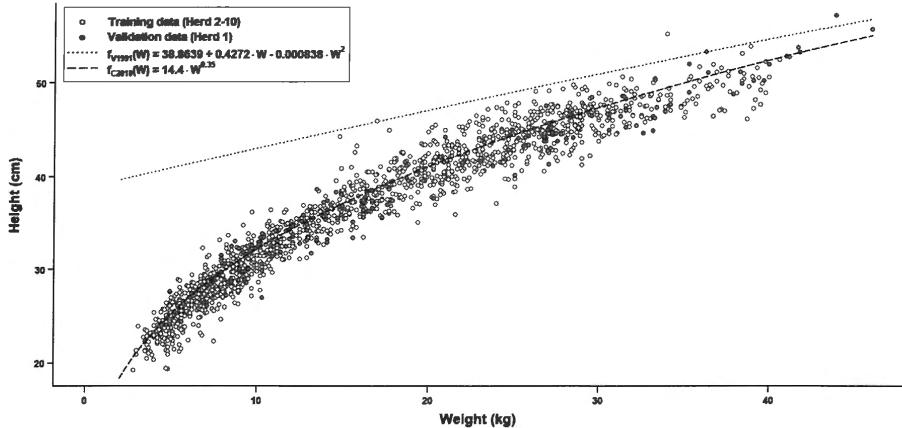


Figure 3 Scatter plot of height vs live body weight of pigs from training data ($N = 1435$) and validation data (grey colour; $N = 179$). The dotted curve is the prediction by Eq. 1, i.e. by the 2nd order polynomial in Vorup and Barton-Gade (1991). The dashed curve is the prediction by Eq. 2, i.e. the power function in Condotta et al. (2018).

The polynomial regression models used for modelling height (H) as a function of weight (W) were linear mixed effects models

$$H_{ijkl} = \beta_0 + \sum_{q=1}^Q \{\beta_q \cdot W_{ijkl}^q\} + \gamma X + \eta_{j(i)} + \varepsilon_{ijkl} \quad (3)$$

where i is an index for herd, j index barn section (in herd), k index pen (of section in herd) and $l = 1, \dots, n_{ijk}$ are the pigs measured in pen k of section j in herd i . Fixed effects parameters are the intercept β_0 and the coefficients β_q , $q = 1, \dots, Q$, where Q is the polynomial order. Other potential factors, covariates, or interactions are indicated by the row vector γ and design matrix X corresponding to these other variables, e.g. breed and sex. The random effect η of section within herd and the residual error ε are assumed normal distributed with zero mean and variances σ_η^2 and σ_ε^2 , respectively. In addition, we included a compound symmetry correlation within pen, i.e. $\text{cor}(\varepsilon_{iikl}, \varepsilon_{iikl'}) = \rho$ for two pigs l and l' from the same pen and $\rho = 0$ for pigs from different pens. The final predictions were made on population level, i.e. based only on the fixed effects estimates.

Three nonlinear mixed effects 3 parameters growth models were examined: asymptotic, Gompertz and logistic. For the sake of brevity and since these were not giving the best predictions, further details are left aside but can be found in Foldager et al. (2021).

The third type of model was based on a linear regression of $\log(\text{height})$ on $\log(\text{weight})$

$$\log(H_{ijkl}) = \beta_0 + \beta_1 \cdot \log(W_{ijkl}) \quad (4)$$

followed by exponential back-transformation giving the power function

$$H_{ijkl} = f_p(W_{ijkl}) = \exp(\beta_0) \cdot W_{ijkl}^{\beta_1} = \alpha \cdot W_{ijkl}^{\beta} \quad (5)$$

The derived parameter $\alpha > 0$ ensures a height > 0 (for weight > 0) even if the intercept from Eq. 4 is negative. The function in Eq. 5 will be strictly increasing if $\beta > 0$, strictly decreasing if $\beta < 0$ and constant ($= \alpha$) if $\beta = 0$. To estimate this model type, we use the model of Eq. 3 with log-transformed heights and weights, and $Q = 1$.

Prediction performance of the models were examined in the validation herd ($i = 1$) by the root mean squared error (RMSE), a coefficient of determination (R^2), and the mean absolute error (MAE) as defined in the equations below, where N_1 is the validation sample size, \hat{H} and H are predicted and observed values and \bar{H} the average H .

$$RMSE = \sqrt{\frac{1}{N_1} \sum_{j=1}^{J_1} \sum_{k=1}^{K_{1j}} \sum_{l=1}^{L_{1jk}} (H_{1jkl} - \hat{H}_{1jkl})^2} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{j=1}^{J_1} \sum_{k=1}^{K_{1j}} \sum_{l=1}^{L_{1jk}} (H_{1jkl} - \hat{H}_{1jkl})^2}{\sum_{j=1}^{J_1} \sum_{k=1}^{K_{1j}} \sum_{l=1}^{L_{1jk}} (H_{1jkl} - \bar{H}_{1jkl})^2} \quad (7)$$

$$MAE = \frac{1}{N_1} \sum_{j=1}^{J_1} \sum_{k=1}^{K_{1j}} \sum_{l=1}^{L_{1jk}} |H_{1jkl} - \hat{H}_{1jkl}| \quad (8)$$

Decisions on the final order of the polynomial (Q_{final}) and inclusion of other fixed effects were based on χ^2 likelihood ratio test. As an extra check, prediction performance was calculated also for polynomial orders $Q_{\text{final}} - 1$ and $Q_{\text{final}} + 1$.

Microclimatic conditions: temperature and CO₂

For each response variable, a separate model was created for each combination of journey length and vehicle part: short and long journeys in the lorry and trailer. The temperature (°C) from study sensors was recorded once per minute and the CO₂ concentration (ppm) was recorded every minute but stored as average at 10 min intervals.

Both of these were smoothed within journey (and sensor) with the R function *loess* to remove sudden unrealistically high peaks. For CO₂, the loess fits were also used to fill in observations every minute; enabling easy merging with other sources of information.

From the vehicle's average speed, an indicator function (Drive; 0/1 for average speed > 5 km/h) was created, interpreted as the states "stopped" and "driving" (State) and used include driving speed (Speed) only when the vehicle was driving. Mechanical ventilation were generally only used sporadically and not included in analyses. Due to moderately high and high correlations between outdoor temperature (WeaTemp) and respectively relative humidity and total shutter opening (left + right side), only WeaTemp and wind speed (WindSp) were also included in models in addition to the experimental factors: Compartment (see Fig. 2), deck height (Height) and period (Block).

The temperature inside each compartment and CO₂ concentration during transport was analysed on the level of the GPS data (approximately at 5 min-intervals or 15 min-intervals for two journeys) in a linear mixed effects model with random effect of journey (via departure date) and an AR(1) correlation structure to account for repeated measurements from the same compartment within journey. The initial fixed effects were main effects and all two-way interactions between: Compartment, Height, State and Block, all as categorical factors, and the continuous variables WeaTemp, WindSp and Drive:Speed (speed when driving). The models were reduced stepwise, starting with the least significant interaction effect. When no more interactions could be removed, non-significant main effects (not being part of an interaction) were removed in the same stepwise manner. The main effects of Compartment, Height and State stayed in the model even if non-significant to enable quantification of these effects. For CO₂, only data from block 3 and 4 were used due to many missing recordings in block 1 and 2. CO₂ data were also less complete in block 3 and 4 than were temperature.

The final models are shown as footnotes to the result tables. Mean estimates of categorical factors (i.e., intercepts) were calculated with continuous variables set to zero using *emmeans* from the *emmeans* package v. 1.7.1-1. Compartments were compared within deck and between decks within horizontal position using *glht* from the *multcomp* package v. 1.4-17 for these nine pre-selected contrasts with adjustment for multiple testing by the default one-step procedure of *glht*. Adjusted p-values are denoted p_{adj} .

Results

Height-weight relation

Inclusion of herd characteristics and sex of the pig was examined but only sex found to be of statistically significant importance. In the polynomial regressions, interaction between sex and (powers of) weight were also tested to investigate if sex specific curves would fit best. Nevertheless, the differences between castrates and female pigs were too small to be of any practical importance and only unisex equations will therefore be

presented here. The obvious confounder of time since weaning was only known for approximately half of the pens and thus could not be included; anyway, may also not be useful in practical applications and was highly correlated with both height (Pearson's correlation coefficient, $\rho = 0.84$) and weight ($\rho = 0.81$).

Overall, the two best prediction models, among those presented here, were the third order polynomial and the power function (Table 1). Thus, we recommend either the following third order polynomial

$$f_1(W) = 15.5 + 2.06 \cdot W - 0.0500 \cdot W^2 + 0.000514 \cdot W^3 \quad (9)$$

or the following power function

$$f_2(W) = 14.0 \cdot W^{0.353} \quad (10)$$

both of which are shown in Fig. 4 with 95% prediction curves. These functions are valid for weights between 5 and 40 kg and should not be used outside this range, though they will probably fit reasonably down to 3 kg.

Table 1 Performance of models for prediction of height from weight for the 179 pigs from the validation herd, divided into those weighing ≤ 40 kg and > 40 kg, respectively. The measures shown are the root mean squared error (RMSE, Eq. 6), a coefficient of determination (R^2 , Eq. 7) and the mean absolute error (MAE, Eq. 8)

Model	<i>Pig weight ≤ 40 kg</i>			<i>Pig weight > 40 kg</i>		
	(N=173)			(N=6)		
	RMSE	R^2	MAE	RMSE	R^2	MAE
Polynomial (Eq. 3)						
2 nd order	1.66	0.945	1.33	5.77	-11.2	5.49
3 rd order	1.59	0.949	1.26	2.13	-0.648	1.87
4 th order	1.59	0.949	1.26	2.72	-1.69	2.48
Nonlinear growth						
Asymptotic	1.63	0.947	1.29	4.42	-6.13	4.20
Gompertz	1.66	0.945	1.32	4.98	-8.03	4.76
Logistic	1.70	0.942	1.36	5.42	-9.71	5.20
Power function (Eq. 5)						
2 nd order polyn. (Eq. 1) ¹	9.10	-0.661	8.38	1.78	-0.158	1.71
Power function (Eq. 2) ²	1.81	0.934	1.46	1.31	0.377	0.770

¹Vorup and Barton-Gade (1991); ²Condotta et al. (2018) earlier

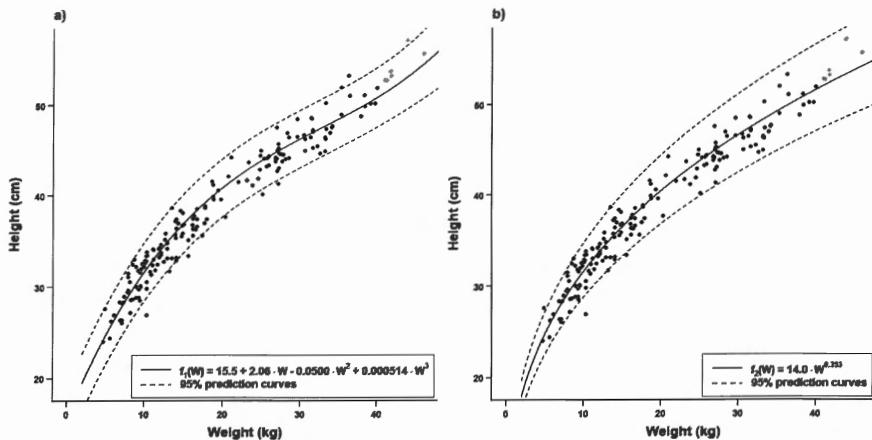


Figure 4 Scatter plot of height (cm) and live body weight (kg) for 179 pigs from the herd used for validation of the prediction equations; grey points: weight > 40 kg ($N = 6$). Predicted height from weight (solid curve) with 95% prediction bands (dashed curves) as obtained by: a) the 3rd order polynomial in Eq. 9 and b) the power function in Eq. 10.

Microclimatic conditions: temperature

Results from models of the mean temperature during transport are shown in Table 2. On short journeys, the temperature in the lorry was higher at 70 cm compared to 90 cm with a mean difference of 1 ± 0.5 °C. In the trailer for short journeys and in both lorry and trailer for long journeys, there were no differences between deck heights. There was an effect of State for short journeys in the lorry with a negligible difference: 0.03 ± 0.014 °C lower temperature when stopped, far less than the ± 0.5 °C precision of the measurement. In contrast there was an effect of compartment in the horizontal direction of the vehicle for all four combinations of journey length (short/long) and vehicle part (lorry/trailer) and with the same general direction; warmest in the front-end compartment, coldest in the rear-end and with the middle in between but closest to the temperature of the front-end. None of the vertical comparisons showed significant differences.

For short journeys in the lorry, compartment 1A and 1B differed from 1C by 6 ± 0.8 °C and 4 ± 0.8 °C, respectively, while 1A did not differ significantly from 1B. Similarly on deck 3, compartment 3A and 3B differed from 3C by 6 ± 0.8 °C and 4 ± 0.8 °C, respectively. Here 3A also tended to differ from 3B (2 ± 0.8 °C; $p_{adj} = 0.064$). On short journeys in the trailer, the temperatures in 1D and 1E were higher than in 1F by 5 ± 0.8 °C and 3 ± 0.8 °C, whereas 1D and 1E did not differ significantly. Similarly, compartment 3D and 3E were 7 ± 0.8 °C and 5 ± 0.8 °C higher than in 3F and 3D tended to be higher than the middle 3E (2 ± 0.8 °C; $p_{adj} = 0.056$).

Table 2 Summary of model output on the mean temperature (°C) during transport

Variable	Levels	Short journeys ¹		Long journeys ²	
		Mean ± SE	p-value	Mean ± SE	p-value
Lorry ^a	Height (deck)	70 cm	22 ± 0.5	0.009	21 ± 0.4
		90 cm	21 ± 0.5		20 ± 0.4
	Compartment	1A	24 ± 0.7	< 0.001	22 ± 0.6
		1B	22 ± 0.7		21 ± 0.6
		1C	18 ± 0.7		17 ± 0.6
		3A	24 ± 0.7		23 ± 0.6
		3B	22 ± 0.7		21 ± 0.6
		3C	18 ± 0.7		16 ± 0.6
	State	Driving	21 ± 0.4	0.033	20 ± 0.3
		Stopped	21 ± 0.4		20 ± 0.3
Trailer ^b	Height (deck)	60 cm	21 ± 0.4	0.251	20 ± 0.3
		80 cm	21 ± 0.4		20 ± 0.3
	Compartment	1D	23 ± 0.6	< 0.001	23 ± 0.6
		1E	21 ± 0.6		19 ± 0.6
		1F	19 ± 0.6		17 ± 0.6
		3D	24 ± 0.6		24 ± 0.6
		3E	22 ± 0.6		20 ± 0.6
		3F	17 ± 0.6		17 ± 0.6
	State	Driving	21 ± 0.4	0.159	20 ± 0.2
		Stopped	21 ± 0.4		20 ± 0.2

^{1a} Temperature ~ Compartment + Height + Block + State + WeaTemp + WindSp + Block:WeaTemp + Block:WindSp + WeaTemp:WindSp + Compartment:Drive:Speed

^{1b} Temperature ~ Compartment + Height + Block + State + WeaTemp + WindSp + Drive:Speed + Block:WindSp + Compartment:Drive:Speed + WeaTemp:Drive:Speed + WindSp:Drive:Speed

^{2a} Temperature ~ Compartment + Height + Block + State + WeaTemp + WindSp + Drive:Speed + Compartment:WeaTemp + Block:WeaTemp + Compartment:Drive:Speed + Block:Drive:Speed + WeaTemp:Drive:Speed + WindSp:Drive:Speed

^{2b} Temperature ~ Compartment + Height + Block + State + WeaTemp + WindSp + Drive:Speed + Compartment:WeaTemp + Block:WeaTemp + Compartment:Drive:Speed + Block:Drive:Speed + WeaTemp:Drive:Speed + WindSp:Drive:Speed

On long journeys in the lorry, 1A and 1B were 6 ± 0.8 °C and 5 ± 0.8 °C higher than 1C and 3A and 3B were 7 ± 0.8 °C and 4 ± 0.84 °C higher than 3C. Compartments 1A and 1B did not differ whereas 3A was 3 ± 0.8 °C higher than 3B. Finally for long journeys in the trailer, 1D and 1E were 6 ± 0.8 °C and 2 ± 0.8 °C higher than 1F, respectively. Compartments 1D and 3D had mean temperatures, which were 4 ± 0.8 °C higher than 1E and 3E, respectively. Both 3D and 3E were higher than 3F with mean differences of 7 ± 0.8 °C and 3 ± 0.8 °C, respectively.

Microclimatic conditions: CO₂

Summary of the output from models on the mean temperature during transport are shown in Table 3. For all journey lengths and decks, there was an effect of compartment in the horizontal direction of the vehicle with a higher level in the front compartments. Vertically, there were no significant compartment differences for 1A vs 3A, 1B vs 3B and 1C vs 3C in the lorry, and for 1D vs 3D, 1E vs 3E and 1F vs 3F in the trailer.

Table 3 Summary of model output on the mean CO₂ (ppm) level during transport

Variable	Levels	Short journeys ¹		Long journeys ²		
		Mean ± SE	p-value	Mean ± SE	p-value	
Lorry ^a	Height (deck)	70 cm	3230 ± 281	0.129	2660 ± 238	0.084
		90 cm	2710 ± 281		2210 ± 238	
	Compartment	1A	3910 ± 456	< 0.001	3010 ± 355	< 0.001
		1B	3030 ± 456		3030 ± 355	
		1C	1910 ± 529		1740 ± 408	
		3A	3770 ± 456		3590 ± 355	
		3B	3760 ± 456		2120 ± 355	
		3C	1440 ± 456		1120 ± 355	
	State	Driving	2960 ± 210	0.027	2433 ± 195	0.337
		Stopped	2980 ± 210		2433 ± 195	
Trailer ^b	Height (deck)	60 cm	2870 ± 287	0.533	2580 ± 184	0.176
		80 cm	3110 ± 286		2250 ± 183	
	Compartment	1D	4440 ± 498	0.001	4110 ± 317	< 0.001
		1E	3020 ± 496		1840 ± 319	
		1F	2400 ± 492		1520 ± 318	
		3D	4050 ± 495		3920 ± 317	
		3E	2080 ± 495		1720 ± 318	
		3F	1930 ± 492		1390 ± 317	
	State	Driving	2980 ± 203	0.115	2420 ± 130	0.433
		Stopped	3000 ± 203		2420 ± 130	

^{1a} CO₂ ~ Compartment + Height + Block + State + WeaTemp + Drive:Speed + Block:WeaTemp + Block:Drive:Speed + WeaTemp:Drive:Speed

^{1b} CO₂ ~ Compartment + Height + Block + State + WeaTemp + Drive:Speed + Compartment:WeaTemp + Height:WeaTemp + Compartment:Drive:Speed + WeaTemp:Drive:Speed

^{2a} Temperature ~ Compartment + Height + Block + State + WeaTemp + WindSp + Drive:Speed + Compartment:WeaTemp + Block:WeaTemp + Compartment:Drive:Speed + Block:Drive:Speed + WeaTemp:Drive:Speed + WindSp:Drive:Speed

^{2b} Temperature ~ Compartment + Height + Block + State + WeaTemp + WindSp + Drive:Speed + Compartment:WeaTemp + Block:WeaTemp + Compartment:Drive:Speed + Block:Drive:Speed + WeaTemp:Drive:Speed + WindSp:Drive:Speed

For the lorry on short journeys this was manifesting in the mean CO₂ concentration being 2000 ± 685 ppm higher in compartment 1A compared to 1C. On the third deck, 3A and 3B were 2330 ± 631 ppm and 2320 ± 631 ppm higher than 3C. There were no differences between 1A and 1B, 1B and 1C or between 3A and 3B. There was an effect of State but with a difference of only 14 ± 6.3 ppm this has no practical implication. On short journeys in the trailer the picture was about the same with 1D being 2040 ± 699 ppm higher than 1F but not significantly higher than 1E and 1E not higher than 1F. On the third deck, 3D was 1970 ± 699 ppm higher than 3E and 2120 ± 697 ppm higher than 3F, while the mean CO₂ level in 3E was not different from that in 3F.

On long journeys in the lorry, the CO₂ concentration of compartments 1A and 1B tended to be 1280 ± 504 ppm and 1290 ± 504 ppm higher than 1C ($p_{adj} = 0.081$ and $p_{adj} = 0.077$). The concentration in 3A was 1470 ± 463 ppm higher than 3B and 2470 ± 463 ppm higher than 3C. Finally for long journeys in the trailer, the CO₂ concentration in 1D was 2280 ± 450 ppm higher than 1E and 2590 ± 449 ppm higher than 1F. Likewise, the concentration was 2200 ± 449 and 2530 ± 448 higher in 3D than in 3E and 3F.

Discussion and conclusions

Weight-Height relation

Disregarding marginal sex differences (Foldager et al., 2021), the best predictions were obtained by a third order polynomial (Eq. 9) and by a power function (Eq. 10). The second order polynomial in Eq. 1 developed by Vorup and Barton-Gade (1991) systematically overestimates the height of all but four of the 1614 pigs, whereas the power function in Eq. 2 developed by Condotta et al. (2018) predicted equally well as the best models from the present study. The first figure (“Figur 1”) in Vorup and Barton-Gade (1991) actually also reveals that the height is overestimated by their prediction curve for nine out of the ten pigs in their weight group of 25-34 kg.

The table concerning deck height in vehicles transporting pigs weighing > 40 kg used until now (BEK nr. 26 of 13/01/2020, 2020) relies on the Vorup and Barton-Gade (1991) equation and continued use of this for pigs weighing ≤ 40 kg can therefore not be recommended. Instead using Eq. 9 or Eq. 10 can give proper relations from which a new table for this weight group can be determined and used also for pigs weighing 20-25 kg.

Microclimatic conditions

For the conditions given and within the ranges compared in the second phase of the present study, deck height only had few and minor effects on the chosen indicators of good ventilation during transport; temperature and CO₂ in compartments holding pigs. Irrespective of the deck height, the position of compartments on the vehicle (in the horizontal plane, i.e. from rear-end to front-end) influenced the microclimatic conditions significantly. This was the case for the lorry as well as the trailer, where temperature

and CO₂ concentration inside front-end compartments both were higher than in the rear-end compartments. Typically, the conditions of the middle compartments were most comparable to the front-end though in some cases significantly lower.

From built-in temperature sensors positioned in the outer wall of the lorry and trailer, the driver had access to average temperatures at 15-min intervals. Relatively large temperature differences were observed between the built-in sensors in the outer wall and study sensors near the animals (Herskin et al., 2022) and have attracted some attention from stakeholders and journalists. This observation calls for further investigation as it indicates that drivers may not know the actual temperature near the pigs, and thus have poor chances of knowing when adjustment of shutter openings or use of mechanical ventilation may be relevant to reduce the temperature or CO₂ level.

Acknowledgement

We thank the drivers of the vehicle including Niels Jeppesen, the owner of the transport company Bette Gris, who is furthermore acknowledged for letting us drill extra holes and attaching equipment in his vehicle. Khem Raj Gautam was part of the project for a period as post doc and is thanked for his contribution to retrieving weather data.

References

- ASABE (2020). Dimensions of Livestock and Poultry. ASABE Standard ASAE D321.2 MAR1985 (R2020). American Society of Agricultural and Biological Engineers, St. Joseph, MI. <https://elibrary.asabe.org/standards.asp>
- Bek nr. 26 of 13/01/2020 (2020). Law No. 26, 13/01/2020 on the protection of animals during transport. [Danish] <https://www.retsinformation.dk/eli/lt/2020/26>
- Condotta ICFS, Brown-Brandl TM, Stinn JP, Rohrer GA, Davis JD, Silva-Miranda KO (2018). Dimensions of the Modern Pig. Transactions of the ASABE 61(5):1729-1739. <https://doi.org/10.13031/trans.12826>
- Council Regulation (EC) No 1/2005 of 22 December 2004 on the protection of animals during transport and related operations and amending Directives 64/432/EEC and 93/119/EC and Regulation (EC) No 1255/97. <https://www.eumonitor.eu>
- Foldager L, Kaiser M, Herskin MS (2021). Note on prediction of height from weight for pigs weighing 5 to 40 kg. 10 pages. Advisory memorandum from DCA – Danish Centre for Food and Agriculture, Aarhus University. <https://dca.au.dk/publikationer>
- Herskin MS, Kobek-Kjeldager C, Kaiser MØ, Thodberg K, Jensen LD, Chen G, Zhang G, Rong L, Foldager L (2022). Krav til indvendig højde ved transport af smågrise. Rådgivningsnotat fra DCA – Nationalt Center for Fødevarer og Jordbrug, Aarhus Universitet. [Danish advisory report] <https://dca.au.dk/publikationer>
- R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>
- Vorup P, Barton-Gade P (1991). Dimensioner på danske svin anno 1991 og dets betydning for trans-portforhold. Slagteriernes Forskningsinstitut. [Danish report]

Er det ok at stoppe et forsøg før tid?

Helle M. Sommer, SEGES Innovation

Det er ofte set, at man følger akkumulerede data under udførelsen af et forsøg, herunder også at man også foretage opgørelser af den primære responsvariable. Det er dog også almindeligt kendt, at gentagende signifikantest øger den overordnet type 1 fejl, og derfor bør signifikansniveauet justeres for derved at overholde den samlede type 1 fejl på det ønsket niveau (typisk 5%). De gentagende inspektioner og signifikantest øger sandsynligheden for, at mindst én af analyserne undervejs ('interim' analyser) viser en signifikant forskel under 0-hypotesen. Med andre ord, så øges risikoen for, at en af analyserne viser en signifikant forskel, selv om der reelt ikke er det. Det er derfor vigtigt, at stramme kravet til hvornår, der erklæres en signifikant forskel ved interim analyserne således, at den overordnede type 1 fejl holdes under kontrol. Sekventiel hypotesetestning påvirker ikke kun signifikansstyrken (type 1 fejl) men bevirket også, at den maksimale test størrelse for forsøget skal være større, end hvis der ikke blev foretaget analyser undervejs. Det er derfor vigtigt, at sekventielt hypotesetestningsforsøg planlægges, før forsøget udføres. Den maksimale test størrelse opnås, hvis ingen af interim analyserne giver anledning til at stoppe eller justere forsøget. Der kan være flere årsager til at analysere eller stoppe et forsøg undervejs. Årsagen kan både være af tidsmæssig, økonomisk eller informationsmæssig art. Der kan være interesserantere, som skal motiveres med information om forsøgets 'tilstand' for at fortsætte i forsøget, navnlig hvis forsøget strækker sig over en længere periode. Der kan være usikkerhed om de indgående antagelser ved dimensioneringen (f.eks. varianseestimatet), som man gerne vil verificere og evt. justere for. Eller der kan være tidsmæssige og/eller økonominiske interesser i at reducere omfanget af afprøvningen, hvis der åbenlyst ikke kan opnås den ønskede effekt ('stop for futility') eller hvis effekten er overbevisende stærk, og man således allerede på et tidligt tidspunkt har vist signifikant effekt ('stop for efficacy').

Forsøg, hvor man foretager interim analyser, benævnes ofte 'Sequential trials/designs' eller 'Sequential group trails/designs', hvor sidstnævnte henviser til at analyserne foretages efter en gruppe af data og ikke som førstnævnte, hvor analyserne kan foretages efter hver prøvetagning. Ønsker man at kunne foretage ændringer i antallet af prøver der skal anvendes i næste gruppe på baggrund af resultaterne fra den forrige interim analyse, da skal man anvende 'Adaptive designs'. 'Adaptive designs' dækker også over forholdene ved 'Sequential designs'.

Generelt gælder der, at adgangen til inspektion og interim analyser før forsøget er afsluttet, kun bør tilgås af personer med relevant ekspertise og som er uafhængig af personer som er involveret i at udføre forsøget (FDA 2019).

Forsøgstype der undersøges i nærværende indlæg

For at holde det simpelt er der kun betragtet forsøg med 2 forsøgsgrupper (kontrol og intervention) og hvor data i de 2 forsøgsgrupper (K og I) er normalfordelte og har samme varians. Mange af de nedenfor nævnte metoder kan også anvendes for andre fordelinger eller omskrives til at gælde for disse. Der opstilles en 0-hypotese for om forskellen mellem de 2 forsøgsgrupper er signifikant mod alternativet, at de ikke er signifikant forskellige fra hinanden. Det relevante 2-sided test problem er giver ved

$$H_0: \mu_I - \mu_K = 0 \quad \text{mod} \quad H_A: \mu_I - \mu_K \neq 0$$

og den forskel man ønsker at identificere, er givet ved $\delta = \mu_I - \mu_K$ med variansen $\sigma^2 / (\sum n_{I,i}) + \sigma^2 / (\sum n_{K,i})$, hvor og $\sum n_{I,i}$ og $\sum n_{K,i}$ er antallet af prøver i hver af forsøgsgrupperne til og med den i 'te interim analyse. Det totale antal prøver er delt op i grupper med hhv. $n_{I,i}$ og $n_{K,i}$ prøver i hver forsøgsgruppe. I alt er der N grupper og dermed N interim analyser.

Stop for signifikant effekt ('efficacy')

De 3 mest anvendte metoder til at fastlægge stop kriterierne (α') for signifikant forskel er Pocock, Haybittle & Peto og O'Brien & Fleming (Chin & Lee, 2008). Pocock's stop kriterie er kontant over alle interim analyserne, Haybittle & Peto's er konstant over de første $N-1$ interim analyser og har et højere stop kriterie ved den sidste analyse, og O'Brien & Fleming's er konstant stigende over de forskellige interim analyser. Ved disse 3 ovennævnte metoder anvendes der et fast antal prøver i hver datagruppe over de i interim analyser og som er ens for de 2 forsøgsgrupper, dvs. $n_{I,i} = n_{K,i} = n$. Dermed er variansen for δ givet ved $2\sigma^2/(in)$. Lan & DeMets opererer med et fleksibelt antal prøver i hver gruppe og et ikke-på-forhånd fastlagt antal interim analyser ved bestemmelsen af stop kriterierne for signifikans.

Pococks metode er i det følgende gennemgået mere i detaljer og med eksempler.

Mange af forholdene omkring adaptiv design gør sig også gældende for de andre 2 metoder. Vha. numerisk integration fandt Pocock frem til nogle passende stop-krav ved interim analyserne samt til antallet af prøver ($2n$) der skal tages ved hver gruppe (trin). Grænseværdier for signifikans (α') er fundet ved at anvende sammenhængen mellem test-styrken ($1 - \beta$) og teststørrelsen, $Z = \frac{\delta}{\sigma\sqrt{2/(in)}}$ samt $z_{1-\alpha'/2}$.

$$(1 - \beta) = 1 - \Phi(z_{1-\alpha/2} - Z)$$

Fastsættelsen af grænseværdien for en given kombination af α, β, N , blev fundet ved at vælge den værdi, der bevirkede, at den gennemsnitlige overordnede α var lig 5% under den alternative hypotese. Til dette formål blev der ved simulering bestemt det gennemsnitlige antal trin (eller grupper), der skulle til for at opnå signifikans. Ved på forhånd at bestemme sig for α, β, N samt angive δ som funktion af σ kan man ved Pocock metode bestemme antallet af $2n$ prøver, der indgår i hvert trin (gruppe) og anvende Poccoks grænseværdi for signifikans (benævnt nominal significance level).

Group sequential designs for a normal response with known variance σ^2 , α (overall significance) = .05, $1 - \beta$ (power under $H_A: \mu_A - \mu_B = \delta$) = .95 and several choices of N (maximum number of groups)

Maximum number of groups, N	Nominal significance level, α'	Required number* of patients per group, $2n$	Maximum number* of patients, $2nN$	Average number* of patients until trial stopped under H_A
1	.05	51.98	52.0	52.0
2	.0294	28.39	56.8	37.2
3	.0221	19.73	59.2	33.7
4	.0182	15.19	60.8	32.2
5	.0158	12.38	61.9	31.3
10	.0106	6.50	65.0	29.8
20	.0075	3.38	67.6	29.5

* Multiply each value by σ^2/δ^2 .

Kilde: Pocock, 1977

Tabellen ovenfor gælder for mange typer responsvariable (normal, binomial, Wilcoxon, model med justering for kovariater og varierende antal i grupperne (n)).

Der er stor fordel ved at gå fra ingen interim analyse til blot én interim analyse under den alternative hypotese. Man får dog ikke meget mere ud af at foretage mere en 5 interim analyser (tabel ovenfor). Hvis man regner med, at der er en reel forskel mellem de 2 behandlinger (at den alternative hypotese er sand), så kan man forvente at spare en del prøver. F.eks. hvis man har valgt at have 5 interim analyser, så falder det forventede antal prøver med 40% (fra 52* til 31,3*; hvor tallene ganges med σ^2/δ^2) og det maksimale antal stige med blot 19% (fra 52* til 61,9*; hvor tallene ganges med σ^2/δ^2). I nedenstående tabel er der givet et eksempel med $\delta = 0,3\sqrt{2}\sigma$ for bedre at kunne forholde sig til hvad ovenstående tabel viser.

Comparison of fixed, sequential and group sequential designs for $\alpha = .05$ and $1 - \beta = .95$ under $H_A : \delta = 0.3 \sqrt{2} \sigma$. (Normal response with known variance)

	Maximum number of patients	Average number of patients under H_A
Fixed design	289	289
Sequential designs		
SPRT	∞	148
Restricted plan	396	162
RST plan	400	170
RST group sequential designs		
2 groups	316	207
3 groups	329	187
5 groups	344	174
10 groups	361	166
20 groups	376	164

O'Brien & Flemings kriterie er bestemt ved at anvende Z-værdi der afhænger af i $Z_c(i) = Z_{OBF} \sqrt{(N/i)}$, hvor Z_{OBF} er fundet ved simulering således, at den overordnet type 1 fejl under 0-hypotesen er overholdt. Haybittle & Petos stop kriterie er mere eller mindre arbitrært valgt. Sammenligning af stop kriterierne for de 3 metoder, Pocock, Haybittle & Peto og O'Brien & Fleming ses af nedenstående tabel for forsøg med hhv. $N=1, 2, 3, 4$ og 5 interim analyser samt et α niveau på 5%.

Stop kriterie α'				
N	Interim analysenr, i	Pocock	Haybittle & Peto	O'Brien & Fleming
2	1	0,0294	0,002	0,0054
	2	0,0294	0,05	0,00492
3	1	0,0221	0,001	0,0006
	2	0,0221	0,001	0,0151
	3	0,0221	0,05	0,0471
4	1	0,0182	0,001	0,00005
	2	0,0182	0,001	0,0039
	3	0,0182	0,001	0,0184
	4	0,0182	0,05	0,0412
5	1	0,0158	0,001	0,000005
	2	0,0158	0,001	0,0013
	3	0,0158	0,001	0,0085
	4	0,0158	0,001	0,0228
	5	0,0158	0,05	0,0417

Lan & DeMets (1983) foreslår en alfa-forbrugsfunktion som gør, at man ikke behøver at bestemme antallet af interim analyser på forhånd. Ved hjælp af denne funktion

bestemmes signifikansgrænserne for de forskellige trin af interim analyserne. Første signifikansgrænse α_1' kan beregnes, men de efterfølgende skal findes ved numerisk integration.

$$\alpha_1' = \sqrt{t_1} \Phi^{-1} \left\{ 2\Phi(z_{\gamma/2\alpha}/\sqrt{t_1}) - 1 \right\}$$

hvor t_1 er andelen af prøver ud fra det maksimale planlagt antal, som indgår i trin 1.

Sidenhen er der foreslået adskillige andre planlagte procedure for sekventielle hypotesetestninger med interim analyser, som tillader modifikation af designet for den resterende del af forsøget uden at gå på kompromis med det overordnet signifikansniveau på f.eks. 5%. Med udvidelsen af disse features, som også inkluderer stop af forsøg pga. umuligheden i at komme i mål, samt justering af prøvestørrelse, da benævnes designet ofte 'adaptiv design'.

Stop for uduelighed af forsøg (futility)

Stop-kriterierne i sekventielle gruppe designs er ikke lige så undersøgt som stop-kriterier for signifikant effekt. Der er dog i litteraturen beskrevet et par metoder hvorpå man kan stoppe et forsøg pga. manglende effekt.

- 1) På baggrund af de indsamlede data fra interim analysen beregnes sandsynligheden for at opnå en signifikant effekt ud fra den effekt-forskel man ønskede at kunne identificere. Er denne sandsynlighed meget lille f.eks. <10% så standses forsøget
- 2) Disse forfattere anvender en beta-forbrugsfunktion
DeMets & Ware og Li, Herrmann & Rauch
- 3) En konsensus er at anvende en p-værdigrænse på 50%

Ønsker man at stoppe flest mulige forsøg pga. 'futility', men stadig bevare en ønsket teststyrke, da kan man anvende en beta-forbrugsfunktion til at opstille stop-kriterierne for 'futility'.

Justering af prøvestørrelsen

Hvis et forsøg ved en interim analyse hverken stoppes pga. af signifikans eller mangel på samme, så kan man på baggrund af de akkumulerede data justere antallet af prøver for at optimere prøvestørrelsen, der skal anvendes i det næste trin (gruppe). Er p-værdien fra forrige trin f.eks. tæt på at være under den valgte signifikansgrænse, så kan man i det næste trin nøjes ned en mindre prøvestørrelse end ellers planlagt fra start, og omvendt kan prøve antallet justeres op, dog højest til maks-prøvestørrelsen, som er

bestemt fra start. Over de seneste årtier har der været udviklet flere metoder for 'resampling' eller 'recalculation' som basere sig på betinget power beregning.

Fordele og ulemper ved adaptiv design

Fordele

- I gennemsnit færre prøver, reducerer udgifterne
- Reducere tidsforbruget
- Øger sandsynligheden for succes – for at finde en bestemt forskel f.eks. ved at øge prøveantallet, hvis dimensioneringsforudsætningerne ikke holdt.
- Lettere at overbevist interessertere om at deltage/medvirke, når der ligger en plan for at afslutte forsøget, hvis det ikke duer eller hvis man har opnået signifikant effekt før tid.

Ulemper

- Færre data og dermed dårligere bestemt parameterestimater
- Mere tidskrævende at planlægge og udføre. Herunder at nå at foretage interim analyserne inden næste gruppe af data i forsøget potentielt skal sættes i gang.
- Svære at tolke p-værdier og skal derfor tolkes med forsigtighed – fokus bør i højere grad være på, at der er signifikant forskel mellem de 2 behandlinger på et 5% niveau frem for den enkelte p-værdi
- Kompliceret at korrigere for biased estimerater (Jennison & Turnbull, 1999)
- Stoppes eller nedjusteres forsøget ikke undervejs, da anvendes 'maks prøve antallet' som er lidt større end ved et klassisk ikke-adaptivt design.

Referencer

- Chin & Lee**, 2008. Principles and Practice of Clinical Trial Medicine
- O'Brien & Fleming**. *A Multiple Testing Procedure for Clinical Trials*. Biometrics, Vol. 35, No. 3, pp. 549-556
- FDA**, 2019. Adaptive Designs for Clinical Trials of Drugs and Biologics – Guidance for Industry.
- Jennison & Turnbull**, 1999. *Group Sequential Methods with Applications to Clinical Trials*. CRCPress
- Lan & DeMets**, 1983. *Discrete Sequential Methods with Applications to Clinical Trials*. Biometrika, Vol 70, No. 3, pp.659-663
- Pocock**, 1982. *Interim Analyses for Randomized Clinical Trials: The Group Sequential Approach*. Biometrics, Vol. 38, No. 1, pp. 153-162
- Pocock**, 1977. *Group Sequential Methods in the Design and Analysis of Clinical Trials*. Biometrika, Vol. 64, No. 2, pp. 191-199
- DeMets & Ware**, 1980. *Group sequential methods for clinical trials with a one-sided hypothesis*. Biometrika, Vol. 67, No. 3, pp. 651-660
- Li, Herrmann & Rauch**, 2020. *Optimality criteria for futility stopping boundaries for group sequential designs with a continuous endpoint*. BMC MedResMethodol. Vol 20, No. 1, pp. 280

En statistisk sammenligning af danske skoleelevers læsefærdigheder før og efter Folkeskolereformen 2014

Josefine Palmø og Sophie A. Steiness
Økonomisk Institut, Københavns Universitet

Abstract

This paper aims to clarify whether Danish students were affected by the education reform of 2014 "Folkeskolereformen". One of the objectives of the reform was to reduce social background's significance on the students' reading skills.

The analysis is based on the data from OECD's PISA program from 2009 and 2018 as both years had reading as the main area of focus. The final model is estimated using ordinary least square estimation. This analysis finds that the following explanatory variables *socioeconomic background, socioeconomic composition of the student's school, sex, immigration status, metacognition strategies in reading when summarising, and metacognition strategies in reading when understanding and remembering* to be of significance when estimating students' reading skills.

In conclusion the analysis shows no significant development in the average student's reading skills since the reform. The same applies to the significance of *social background, socioeconomic composition of the student's school, sex, immigration status and metacognition strategies in reading when understanding and remembering on students' reading proficiency*. However, this paper does find a significant increase in the importance of *metacognition strategies in reading when summarising between the two PISA studies*. The results regarding no significant development show similarities to the results of the 5-year follow-up research program which until 2018 mapped how well the reform was implemented.

Parts of the work described in this paper have been submitted as our Bachelor thesis, fall 2021 (Steiness, 2021).

1 Indledning

"Reformen bør have danske Pisa-resultater" (Folkeskolen.dk, 2013) og "Vi skal gøre skolen endnu bedre" (Berlingske, 2013) lød overskrifterne i danske medier december 2013. Få uger efter var aftalen om Folkeskolereformen 2014 på plads. En reform, som senere er udskældt for ikke at lykkedes med sine overordnede mål. Denne undersøgelse stiller skarpt på danske skoleelevers læsefærdigheder før og efter Folkeskolereformen 2014. I denne sammenhæng vil vi også se på betydningen af social baggrund, som reformen havde som delmål. I tråd med lignende studier finder vi, at niveauet for læsefærdighederne målt ved PISAs læsescore er uændret af reformen.

Vi benytter PISA-undersøgelserne for 2009 og 2018. Vores endelige model for begge år er en multipel lineær model udledt vha. OLS estimation, der består af *socioøkonomisk baggrund, klassekammerateffekten, køn, immigrationsstatus, metakognitive læsestrategier ved resume af tekst samt metakognitive læsestrategier ved erindring og forståelse af tekst*. Med undtagelse af de metakognitive læsestrategier er ovenstående variable typisk med i PISA-analyser. Ved brug af hypotesetest laver vi en statistisk sammenligning på tværs af de to år, der undersøger om der er sket en ændring i elevernes faglige niveau og i betydning af variablene. Ud fra vores tests finder vi ingen statistisk signifikant ændring i læsescoren samt i betydningen af de forklarende variable med undtagelse af metakognitive strategier ved resume af tekst.

Problemstillingen er økonomisk relevant, da det er vigtigt for Danmark at forblive konkurrencedygtig på det internationale marked, hvor arbejdsstyrken af veluddannede også stiger. En måde at opnå dette er ved at investere i folkeskolen, da det er der fundamentet for uddannelse dannes. Denne tankegang var drivkraften bag Folkeskolereformen 2014.

Folkeskolereformen 2014 blev vedtaget d. 20. december 2013. Reformen udsprang fra et politisk ønske om at have danske skoleelevers læse-, matematik- og naturfagsfærdigheder op på niveau med OECD-gennemsnittet og udvikle alle elever fagligt, så de får chancen for at udfolde deres potentielle samtidig med at Danmark forbliver internationalt konkurrencedygtige (Regeringen (A, B, F) samt V og O, 2013, pp. 1).

Reformen består af tre målsætninger samt fire resultatmål, se tekstdokumentet nedenfor.

- 1) Folkeskolen skal udfordre alle elever, så de bliver så dygtige, de kan.
 - Mindst 80 pct. af eleverne skal være gode til at læse og regne i de nationale test
 - Andelen af de allerdygtigste elever i dansk og matematik skal stige år for år
- 2) Folkeskolen skal mindske betydningen af social baggrund i forhold til faglige resultater.
 - Andelen af elever med dårlige resultater i de nationale test for læsning og matematik uanset social baggrund skal reduceres år for år
- 3) Tilliden til og trivslen i folkeskolen skal styrkes blandt andet gennem respekt for professionel viden og praksis.
 - Elevernes trivsel skal øges

(Rosenkrantz-Theil, 2021, pp. 5).

De tre overordnede målsætninger indfries igennem tre indsatsområder:

1) "En længere og mere varieret skoledag med mere og bedre undervisning og læring". Implementeringen består af flere fagtimer i dansk, matematik og engelsk og dertil mere motion, lektiehjælp og flere udflugter. 2) "Et kompetencecelført af lærere, pædagoger og skoleledere". Kompetencecelføret indebærer, at 95 pct. af undervisningen skal udføres af kompetence undervisere i 2020. Målet er senere blevet udskudt til 2025. 3) "Få klare mål og regelforenklinger", der imødekommer udfordringerne omkring styringen af folkeskolen ved at lægge ansvaret over på kommunerne og indføre klare faglige mål (Regeringen (A, B, F) samt V og O, 2013).

2 Data

PISA er et OECD-program, der tester 15-årige i læsning, matematik og naturfag. Programmet begyndte i 2000 og foretages hvert tredje år på skoler i OECD-lande. Danmarks fjerde og syvende PISA-undersøgelse blev gennemført i 2009 og 2018, hvor hoveddømmet var læsning. De to år er derfor brugbare til at belyse effekten af Folkeskolereformen i forhold til elevers læsefærdigheder. I begge stikprøver er der en oversampling af elever med en anden etnisk herkomst end dansk (Christensen, 2019, pp. 8).

I PISA-undersøgelsen er der, ligesom mange andre dataser, manglende værdier. Vi benytter derfor multipel imputering. De manglende værdier kan give et indtryk af, at påvirkningen fra en variabel er anderledes end den faktisk er. Når vi ser på socioøkonomisk baggrund (ESCS), i tabel 3, understøttes denne pointe: Det er i højere grad elever med en læsescore under det danske gennemsnit for det pågældende år, som mangler at besvare ESCS. I det vi vælger at imputere data bliver hældningen på regressionslinjen for ESCS støjtere end i figur 1. Dette billede er gældende for alle de forklarende variable. Vi opnår derfor mere information om variablenes påvirkning af læsescoren ved imputering. Den eneste variabel, hvor det ikke er mulig at imputere, er immigrationsstatus, da metoden ikke understøtter en dummyvariabel med kun tre mulige kategorier.

År	Manglende værdier i alt	Andel under gennemsnit	Andel over gennemsnit
2009	127	93,7 pct.	6,3 pct.
2018	226	80,1 pct.	19,9 pct.

Tabel 1: Andelen af manglende værdier for ESCS

Kilde: PISA 2009 og 2018, egne beregninger i SAS

I det følgende vil vi uddybe de benyttede variable i analysen:

Score – læsescoren betegner den gennemsnitlige PISA læsescore for den i ’te elev på skole j , hvor p er antallet af læsetestscorer det pågældende undersøgelsesår:

$$Score_{ij} = \frac{1}{p} \sum_{i=1}^p læsetestscore_{ij}.$$

Af tabel 2 ses at den gennemsnitlige læsescore er steget fra 2009 til 2018 med 6.2137 point. Stigningen betyder, at den danske læsescore gik fra at ligger under til over OECD-gennemsnittet i perioden.

År	Variabel	N	Gennemsnit	Standard afv.	95% konfidensinterval af gns.
2009	Score	5924	494.9162	8.9542	477.0967 512.7357
	ESCS	5924	0.2899	0.1082	0.0745 0.5953
	S_ESCS	5924	0.2642	0.1068	0.0517 0.4767
	METASUM	5924	0.1632	0.0914	-0.0187 0.3451
	UNDREM	5924	0.1496	0.0924	-0.0342 0.3334
2018	Score	7657	501.1299	7.5701	486.0649 516.1950
	ESCS	7657	0.5085	0.0644	0.3803 0.6367
	S_ESCS	7657	0.5125	0.0667	0.3797 0.6453
	METASUM	7657	0.1716	0.0851	0.0023 0.3410
	UNDREM	7657	0.0186	0.0822	-0.1450 0.1822

Tabel 2: Gennemsnit og spredning af forklarende variable – imputeret data, vægtet

Kilde: PISA 2009 og 2018, egne beregninger i SAS

ESCS – socioøkonomisk baggrund står for economic, social and cultural status og er et indeks for den i'et elev på skole nr. j's socioøkonomiske baggrund. Den består af elevernes besvarelser om deres forældres uddannelsesniveau (PARED), erhvervsmæssige status (HISEI) og information om specifikke ejendele i husholdningen (HOMEPOS). Figur 1 viser en positiv sammenhæng mellem elevers læsescore og deres socioøkonomisk baggrund uafhængigt af køn¹. Ud fra dette forventer vi, at elevers faglige præstationer er påvirket af deres socioøkonomiske baggrund. En betydningsfuld teoretiker, der er kommet med et bud på hvorfor og hvordan den sociale baggrund spiller ind på elevernes faglige resultater, er den franske sociolog *Pierre Bourdieu* (1930-2002).

En af hans hovedpointer er, at uddannelsessystemet har en iboende sorteringsfunktion, som reproducerer den dominerende klassenes status i samfundet. Dette sker bl.a. ved overførsel af kulturel kapital² på tværs af generationer (Andersen, 2013, pp. 369-376). Hvis vi følger den tankegang, vil børn af f.eks. akademikerforældre, alt andet lige, have en højere kulturel kapital. De får lettere ved at afkode den dominerende kultur i uddannelsessystemet og klarer sig bedre end børn med lavere kulturel kapital.

Når aftalepartnerne bag reformen indførte en længere og mere varieret skoledag, med bl.a. øget lektiehjælp og flere udflygter, kan det tolkes som et ønske om at kompensere for manglende kulturel kapital hos nogle elever og herved højne deres faglige niveau.

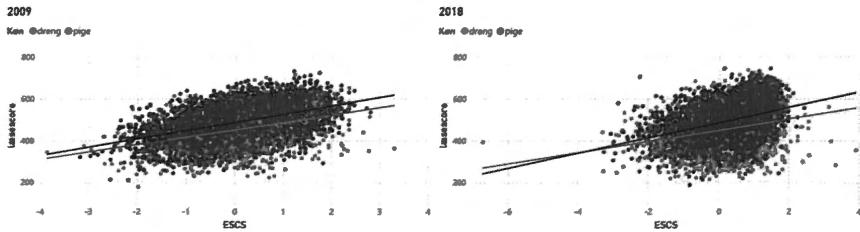
S_ESCS – klassekammerateffekten er et udtryk for skole j's gennemsnitlige socioøkonomiske sammensætning af elever og beregnes på baggrund af ESCS: (Christensen (2019, pp. 151); Bay (2018, pp. 51, 54-55)):

$$S_ESCS_j = \frac{1}{n_j} \sum_{i=1}^{n_j} ESCS_{ij},$$

hvor n_j er antal elever på skole j. Beregningen af variablen gør, at vi forventer, at klassekammerateffekten vil følge samme tendens som ESCS.

¹bemærk: figuren bygger på rådata for at vise den oprindelig tendens

²Bourdies tre kapitalformer: *Økonomisk kapital* omhandler penge og materielle ressourcer. *Kulturel kapital* omhandler uddannelsesniveau og finkulturelle ferdigheder såsom at beherske koderne i finkulturen. En elev har kulturel kapital hvis de koder, som de lærer i hjemmet og i skolen hjælper dem til at tolke kulturens referencesystem. *Social kapital* omhandler sociale relationer og de ressourcer, som en elev får herfra.



Figur 1: Sammenhæng mellem socioøkonomisk baggrund og læsescore fordelt på køn
Kilde: PISA 2009 og 2018, egne beregninger i Microsoft Power BI

Metakognitive læsestrategier

METASUM og UNDREM omhandler metakognitive læsestrategier. Vi omtaler dem samlet som "de metakognitive læsestrategier" i resten af undersøgelsen. Formålet med læsestrategierne er, at eleven bliver en aktiv del af læringsprocessen. Det betyder, at eleverne er bevidste om hvorfor en strategi er god, hvordan forskellige opgaver skal gribes an og hvordan vanskeligheder håndteres i problem løsningen (Geisnæs (2017, pp. 20); Dansk Clearinghouse (2014, fodnote pp.13)). Elevernes svar på nytthigheden af de forskellige strategier sammenlignes med en ekspertgruppens rangering. Høje indeksværdier angiver, at elevens rangering ligner eksperternes. For begge variable skalerer PISA indeksværdierne i to trin: først tildeles hver elev en score mellem 0 og 1 på baggrund af deres forståelse af læsestrategiernes nytthighed. Dernæst standardiseres scorene, så indekset får en middelværdi på 0 og standard afvigelse på 1 på tværs af alle OECD-lande (OECD, 2010, pp. 113). Vi forventer, at en god forståelse for nytteige læsestrategier giver en bedre læsescore: en elev, som kender strategierne, vil benytte dem. Derfor forventer vi et positivt fortegn for de metakognitivt læsestrategier.

METASUM – metakognitive læsestrategier ved resume af tekst laves på baggrund af følgende:

- 1) Elevens egen rangering af nytthigheden af forskellige strategier ved resume-skrivning. Eleven kan vælge mellem strategierne:
 - (A) Jeg skriver et resume. Efterfølgende tjekker jeg, at hvert afsnit i teksten er med i resumeet, da hvert afsnit skal være med.
 - (B) Jeg prøver at kopiere så mange sætninger fra teksten som muligt
 - (C) Før jeg skriver resumeet, læser jeg teksten igennem så mange gange som muligt
 - (D) Jeg gennemtjekker omhyggeligt, at de vigtigste fakta fra teksten er med i resumeet
 - (E) Jeg læser teksten, mens jeg understreger de vigtigste sætninger. Derefter skriver jeg med mine egne ord resumeet.
- 2) En ekspertgruppens rangering af nytthigheden af strategierne: $DE > AC > B$

Frit oversat fra engelsk (OECD, 2010, pp. 113).

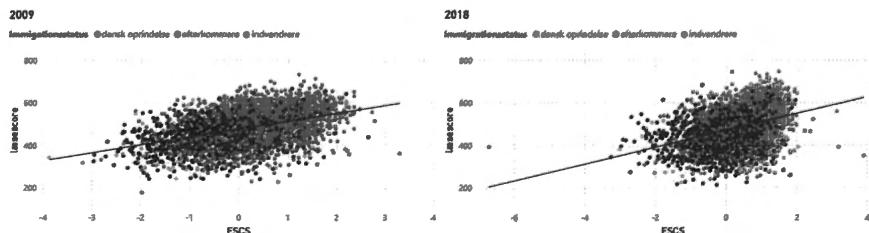
UNDREM – metakognitive læsestrategier ved forståelse og erindring af tekst laves ud fra følgende:

- 1) Elevens egen rangering af hvor nyttige forskellige strategier er. De kan vælge mellem:
 - (A) Jeg koncentrerer mig om de dele af teksten, som er lette at forstå
 - (B) Jeg læser hurtigt teksten to gange
 - (C) Efter jeg har læst teksten, diskuterer jeg indholdet med andre
 - (D) Jeg understreger vigtige dele af teksten
 - (E) Jeg opsummerer teksten i mine egne ord
 - (F) Jeg læser teksten højt for en anden person
- 2) En ekspertgruppens rangering af nyttigheden af strategierne: *CDE > ABF*

Frit oversat fra engelsk (OECD, 2010, pp. 113).

SEX – køn PISA laver en binær dummyvariabel for køn med inddelingen: "pige"(tildelt kvinde ved fødslen) eller "dreng"(tildelt mand ved fødslen). I figur 1 ser vi, at den lineære tendenslinje for piger i 2009 ligger på et højere niveau med en svagt stejleste holdning end tendenslinjen for drenge. Det betyder, at pigers gennemsnitlige læsescore i 2009 er højere end drenge. I 2018 er tendenslinjen for piger ligeledes stejlere, men niveaumæssigt veksler den mellem at være over og under drenge tendenslinje. Dette kan skyldes outliers, som ligger sydvest i grafen og gør drenge tendenslinje fladere. Vi vælger ikke at gøre mere ved denne samme andre outliers (Steiness, 2021, pp. 14), da den udgør en ubetydelig andel af det samlede data.

IMMIG – immigrationsstatus er en dummyvariabel med inddelingen: Dansk oprindelse, indvandrere eller efterkommere³. Figur 2 illustrerer sammenhængen mellem socioøkonomisk baggrund og læsescore for begge år fordelt på immigrationsstatus. Uanset immigrationsstatus, så følger observationerne en samlet positiv, lineær tendens. Niveaumæssigt ligger elever med "dansk oprindelse" højest, dernæst elever, som er "efterkommere", og lavest ligger elever, der er "indvandrer". Andelen af elever med immigrationsstatus "indvandrere" er stort set uændret fra 2009 til 2018. Andelen af elever med immigrationsstatus "efterkommere" stiger med 2,54 pct. point fra 2009 til 2018. Andelen af elever med immigrationsstatus "dansk oprindelse" falder cirka tilsvarende, se tabel 3.



Figur 2: Sammenhæng mellem socioøkonomisk baggrund og læsescore fordelt på immigrationsstatus
Kilde: PISA 2009 og 2018, egne beregninger i Microsoft Power BI

³Ifølge PISA er et barn med to forældre af dansk oprindelse, hvis mindst én forælder er født i Danmark. Barnet er efterkommer, hvis begge forældre er født i udlandet, men barnet selv er født i Danmark. Endelig vil barnet betegnes som indvandrer, hvis barnet selv samt forældrene er født i udlandet.

	Køn "pige"	Køn "dreng"	IMMIG Dansk oprindelse	IMMIG Efterkommere	IMMIG Indvandrere
2009	50.5049	49.4951	91.3783	5.8625	2.7593
2018	49.9548	50.0452	89.3467	8.4058	2.2475

Tabel 3: Andel (pct.) af køn og immigrationsstatus

Kilde: PISA 2009 og 2018, egne beregninger i SAS

3 Analyse og resultater

3.1 Modelselektion

PISAundersøgelsen er baseret på et totbins stikprøvedesign: først udtrækkes skolerne og så eleverne. Stikprøven er disproportionalt allokeret, da udtræksandsynligheden og deltagelsesprocenten varierer på tværs af skoler samt elever, og oversampling af bestemte skoletyper og landsdele forekommer (Egelund, 2010, pp. 160). Vi benytter derfor PISAs stikprøvevægte i vores model for at opnå middelrette parameterestimater.

Der findes to typer af vægt i PISAs datasæt: en samlet vægt (PISAs Final Student Weight) og replikationsvægte. Den samlede vægt sikrer, at niveauet for parameterestimaterne ligner det sande niveau. Replikationsvægtene sikrer, at estimatorens varians er den sande varians. Vi medtager altid begge vægttyper.

De tidligere beskrevne forklarende variable er udvalgt ved baglænselimination. Baglænselimination giver mulighed for simpelt at betragte effekter fra mulige forklarende variable og alene beholde de variable, som i sig selv eller betinget på andre forklarende variable, er signifikante for læsescoren. Vores endelige model for begge år, er:

$$\begin{aligned} score_i = & \beta_0 + \beta_1 ESCS_i + \beta_2 S_ESCS_i + \beta_3 METASUM_i + \beta_4 UNDREM_i \\ & + \delta_1 SEX_i + \delta_2 IMMIG_{i,\text{dansk oprindelse}} + \delta_3 IMMIG_{i,\text{efterkommere}} + \epsilon_i, \quad (1) \end{aligned}$$

hvor følgende dummies angiver: SEX_i er 1 hvis eleven er pige og ellers 0. $IMMIG_{i,\text{dansk oprindelse}}$ er 1 hvis eleven er af dansk oprindelse og ellers 0, samt $IMMIG_{i,\text{efterkommere}}$ er 1 hvis eleven er efterkommer og ellers 0.

Vi har i vores bachelorprojekt (Steiness, 2021, pp. 16-18) udført en dybdegående modelkontrol, der viser at den endelige model opfylder MLR. 1-6. OLS estimatoren er konsistent og middelret med normalfordelte fejled.

3.2 Modelresultater

I det følgende fortolker vi den endelige model. Fælles for parameterestimaterne er (tabel 4), at de er positive og signifikante på 1 pct. signifikansniveau – med undtagelse af immigrationsstatus for efterkommere i 2018. Der er altså ingen signifikant forskel mellem at være indvandrer eller efterkommer i 2018.

Vi ser, at socioøkonomisk baggrund har signifikant betydning i begge undersøgelsesår for elevers læsescore. Elever med høj social, kulturel og økonomisk kapital opnår en højere gennemsnitlig læsescore. Estimatet for ESCS er, med en total ændring på 0,47 point, næsten uændret i perioden.

Parameterestimatet for METASUM stiger fra 2009 til 2018. Det tegner et billede af, at det bliver vigtigere for elevernes læsescore, at de er mere bevidste om læsestrategierne ved resume af tekst gennem perioden. For UNDREM tegner der sig det modsatte billede fra 2009 til 2018.

For 2009 og 2018 ser vi, at piger i gennemsnit får en score på hhv. 16,15 og 12,15 point højere end drenge. Samme tendens gælder for elever med dansk oprindelse og efterkommere i forhold til elever, som er indvandrer. Ifølge den estimerede model vil en pige med dansk oprindelse i gennemsnit opnå 63,08 flere point i læsning end en dreng, som er indvandrer i 2009. I 2018 er tallet 49,23 point højere.

Forklaringsgraden for den endelige model er 0,3987 i 2009 og 0,3481 i 2018, jf. tabel 5. Vi kan altså forklare hhv. 39,87 pct. og 34,81 pct. af variationen i læsescoren for 2009 og 2018. Under forudsætning af opretholdelse af MLR.1-6, kan

Variabel	Estimat	Standardfejl	t-værdi	Pr> t
2009	Intercept	424.7184	3.7727	112.58 <.0001
	ESCS	19.9196	1.3041	15.28 <.0001
	S_ESCS	24.0238	3.6834	6.52 <.0001
	sex Pige	16.1522	1.8466	8.75 <.0001
	sex Dreng	0.0000	0.0000	.
	IMMIG Indvandrere	0.0000	0.0000	.
	IMMIG Dansk oprindelse	46.9223	3.9631	11.84 <.0001
	IMMIG Efterkommere	19.5052	4.8196	4.05 0.0001
2018	METASUM	22.2897	1.2922	17.25 <.0001
	UNDREM	19.8764	1.299	15.30 <.0001
	Intercept	432.1302	6.0713	71.18 <.0001
	ESCS	20.3858	1.5072	13.53 <.0001
	S_ESCS	28.3827	3.7844	7.50 <.0001
	sex Pige	12.1453	2.3905	5.08 <.0001
	sex Dreng	0.0000	0.0000	.
	IMMIG Dansk oprindelse	37.0842	6.1970	5.98 <.0001
	IMMIG Efterkommere	7.7216	6.3976	1.21 0.2310
	IMMIG Indvandrere	0.0000	0.0000	.
	METASUM	26.5594	1.2960	20.49 <.0001
	UNDREM	18.6276	1.2017	15.50 <.0001

Tabel 4: Parameterestimater for den endelige model, 2009 og 2018

Kilde: PISA 2009 og 2018 og egne beregninger i SAS

vi konkludere at socioøkonomisk baggrund, klassekammerateffekten, køn, immigrationsstatus og de metakognitive strategier påvirker elevers læsescore positivt. Det er lig vores forventninger for modellen, som beskrevet i afsnit 2. Om ændringerne er signifikante, tester vi i det efterfølgende afsnit.

k	År	Analyse	Diskussion	
		Model	A1	A2
R ²	2009	0.3987	0.3255	0.2346
	2018	0.3481	0.2982	0.1686

Tabel 5: Forklaringsgrad for den endelige model og alternativ model 1 & 2

Kilde: PISA 2009 og 2018, egne beregninger i SAS

3.3 Statistisk sammenligning

For at se om der er sket en signifikant ændring i modelvariablene fra 2009 til 2018 benytter vi en dobbeltsidet t-test på et 5 pct. signifikansniveau. Den dobbeltsidet t-test er approksimativt normalfordelt og har en kritisk værdi på 1.96, (Wooldridge, 2015, pp. 745).

Ændring i læsescoren Der er en absolut stigning på 6,203 point i den gennemsnitlige læsescore fra 2009 til 2018. For at teste om stigningen er signifikant opstiller vi hypoteserne:

$$\mathcal{H}_0 : E(score_{2009}) = E(score_{2018})$$

$$\mathcal{H}_A : \mathcal{H}_0 \text{ ikke opfyldt.}$$

Vi beregner t-værdien vha. oplysninger fra tabel 2:

$$|t| = \left| \frac{E(score_{2009}) - E(score_{2018})}{\sqrt{\sigma_{score_{2009}}^2 + \sigma_{score_{2018}}^2}} \right| = \left| \frac{494,9262 - 501,1299}{\sqrt{8,95424^2 + 7,570149^2}} \right| = 0,5299.$$

T-værdien ligger udenfor det kritiske område og nulhypotesen opretholdes. Vi afviser, at der er en signifikant stigning i den gennemsnitlige læsescore fra før Folkeskolereformen til efter – eleverne er ikke blevet dygtigere i læsetesten.

Ændring i ESCS Forskellen i parameterestimaterne for ESCS på 0,4662 enheder fra 2009 til 2018 testes med samme metode under følgende hypoteser:

$$\mathcal{H}_0 : \beta_{1,2009} = \beta_{1,2018}$$

$$\mathcal{H}_A : \mathcal{H}_0 \text{ ikke opfyldt.}$$

Vi beregner t-værdien vha. oplysninger fra tabel 4:

$$|t| = \left| \frac{\beta_{1,2009} - \beta_{1,2018}}{\sqrt{\sigma_{\beta_{1,2009}}^2 + \sigma_{\beta_{1,2018}}^2}} \right| = \left| \frac{19,9196 - 20,3858}{\sqrt{1,3041^2 + 1,5072^2}} \right| = 0,2339.$$

T-værdien er udenfor det kritiske område, så nulhypotesen opretholdes. Der er ingen signifikant ændring i betydningen af elevers socioøkonomisk baggrund på læsescoren fra før til efter vedtagelsen af Folkeskolereformen.

Ændring i de resterende variable Vi gentager metoden for alle forklarende variable og får resultaterne i tabel 6.

Variabel	Nulhypoteze	T-værdi	Kritiske værdi	Konklusion
Læsescore	$E(score_{2009}) = E(score_{2018})$	-0,5299	1,96	\mathcal{H}_0 opretholdes
ESCS	$\beta_{1,2009} = \beta_{1,2018}$	0,2339	1,96	\mathcal{H}_0 opretholdes
S.ESCS	$\beta_{2,2009} = \beta_{2,2018}$	0,8254	1,96	\mathcal{H}_0 opretholdes
METASUM	$\beta_{3,2009} = \beta_{3,2018}$	2,333	1,96	\mathcal{H}_0 afvises
UNDREM	$\beta_{4,2009} = \beta_{4,2018}$	0,7058	1,96	\mathcal{H}_0 opretholdes
sex Pige	$\delta_{1,2009} = \delta_{1,2018}$	1,3265	1,96	\mathcal{H}_0 opretholdes
IMMIG Dansk oprindelse	$\delta_{2,2009} = \delta_{2,2018}$	1,3374	1,96	\mathcal{H}_0 opretholdes
IMMIG Efterkommere	$\delta_{3,2009} = \delta_{3,2018}$	1,4711	1,96	\mathcal{H}_0 opretholdes

Tabel 6: Overblik af hypotésetest
Kilde: PISA 2009 og 2018, egne beregninger

Alt i alt opretholdes hypoteserne om ingen ændring i den gennemsnitlige læsescore samt i betydningen af socioøkonomisk baggrund, klassekammerateffekten, UNDREM, køn og immigrationsstatus fra før til efter Folkeskolereformen. Som den eneste afviser vi nulhypotesen om ingen ændring i betydningen af METASUM fra 2009 og 2018. Vi kan altså ikke afvise, at der er sket en ændring i betydningen af metakognitive læsestrategier ved resume af tekst fra før til efter vedtagelsen af Folkeskolereformen.

4 Diskussion

I det følgende vil vi diskutere vores metode samt resultater i relation til følgforskningsprogrammet foretaget i forbindelse med Folkeskolereformen 2014.

4.1 Data- og metodevalg

Datavalg Vi har valgt at bruge PISA-undersøgelserne i vores analyse, da aftaleparterne bag Folkeskolereformen lagde vægt på, at danske skoleelever i 2013 var ”omkring gennemsnittet i OECD i dansk, matematik og naturfag, når de forlader folkeskolen” og at der samtidig var debat om de seneste PISA-resultater fra 2012 samt folkeskolens mulighed til at påvirke resultaterne. Christine Antorini udtalte i den forbindelse ”Vi har allerede gjort en del, og med reformen løfter vi det faglige niveau endnu mere. (...) Men PISA-resultaterne peger på, at vi står med en udfordring, når det handler om at udvikle både de fagligt svage og de fagligt stærke elevers potentiale.”. Her ses en klar politisk forventning fra den daværende regering om, at Folkeskolereformen ville påvirke PISA-resultaterne positivt, hvorfør det er hensigtsmæssigt, at benytte PISA som datakilde.

Alternative modeller Generelt er det svært at undgå udeladt variabel bias (MLR.4) i en model. Det kan antages, at børns trivsel har en effekt på deres faglighed i skolen og derfor kan der være udeladt variabel bias i modellen. I de følgende alternativ modeller tager vi forholdsbehold for dette og undersøger om der er medtaget for mange variable i den endelige model. Regressionsoutput for alternativ modellerne kan findes i (Steiness, 2021, bilag A.2):

Alternativ 1 (A1) inkluderer køn, IMMIG, METASUM og UNDREM, og forskellen ligger i, at vi undlader ESCS samt S_ESCS. Det samme billede mht. signifikans tegner sig i A1 som i den endelige model. Forklарingsgraden falder med 7,32 pct. point for 2009 og 4,99 pct. point for 2018, sammenlignet med den endelige, se 5. Heraf kan vi altså se, at socioøkonomisk baggrund og klassekammerateffekten er betydningsfulde i forklaringen af læsescoren, hvilket bekræfter politikerne bag Folkeskolereformen og Bourdias antagelser om kulturel kapitals betydning ift. faglige resultater. A1 fremstår derfor ikke som et bedre modelvalg.

Alternativ 2 (A2) inkluderer ESCS, S_ESCS, køn og IMMIG, og undlader METASUM og UNDREM. Det samme billede mht. signifikans tegner sig i A2 som i den endelige model. Forklарingsgraden falder med 16,41 pct. point i 2009 og 17,95 pct. i 2018 sammenlignet med den endelige model, hvilket er det største fald mellem den endelige model og en alternativ model. Så de to mål for metakognitive strategier er betydningsfulde, og A2 er ikke en bedre model.

4.2 Diskussion af modelresultater

Vi finder, at socioøkonomisk baggrund, klassekammerateffekten, de metakognitive læsestrategier, køn og immigrationsstatus er signifikante variable til at forklare læsescoren for både før og efter Folkeskolereformen. Men er det egentlig troværdigt? Vi dykker ned i kausaliteten i modellen og genovervejer udeladte variable.

ESCS og klassekammerateffekten – kausalitet Vi finder, at en høj klassekammerateffekt alt andet lige medfører en høj læsescore. Men kausaliteten kan vendes: Dygtige elever (høj læsescore) klynger sig sammen, hvilket skaber en høj klassekammerateffekt. Simplificeret virker det plausibelt jf. Bourdieus teori om uddannelsessystemets iboende sorteringsfunktion. En elev har høj kulturel kapital i kraft af sine forældre f.eks. hvis de er akademikere. Akademiker-forældre kan have mere kapital til at benytte det frie skolevalg og sende deres barn til en skole, hvor den kulturelle kapital er højere. Dette skaber en illusion om at der eksisterer en – eller bidrager til en forsterket – klassekammerateffekt.

METASUM – kausalitet og sammenligning med relevant forskning Vi antager, at de metakognitive læsestrategier bevirkir en høj læsescore, men kausaliteten kan også vendes. Elever med en høj læsescore kan være mere opmærksomme på effektive læsestrategier og scorere højere i de metakognitive læsestrategier. Samtidig kan elever med lav læsescore have problemer med at læse og forstå opgavespørørgsmålene og derfor score lavt i strategierne.

Der er ikke meget dansk forskning vedr. metakognitive strategier. Rambøll kortlægger i ”Forskningskortlægning matematik”(Dansk Clearinghouse, 2014) 19 udenlandske studier omhandlende metakognitive matematikstrategier, (red. systematisk problemløsning). Alt i alt tegner kortlægningen et billede af, at ”systematisk problemløsning har en positiv effekt på elevernes faglige præstationer”(Dansk Clearinghouse, 2014, pp. 27). Selvom vores undersøgelse omhandler læsefærdigheder, kan der drages ligheder med kortlægningen og vi finder samme resultat i forklaringen af læsescoren.

4.3 Diskussion af sammenligning

I det følgende sættes vores hypotesetests resultater i relation til VIVEs rapport "Folkeskolens udvikling efter reformen" (Børne- og Undervisningsministeriet, 2020), som opsamler resultaterne af følgeforskningsprogrammet.

Ingen signifikant ændring i læsescoren og implementering af reformen Vi finder ingen signifikant ændring i elevernes PISA læsescore fra før til efter Folkeskolereformen. Vi kan ikke se en effekt af Folkeskolereformens formål om et fagligt løft af eleverne i perioden. I rapporten "Folkeskolens udvikling efter reformen" er den overordnede konklusion, at der ingen effekt er i de tre overordnede mål i perioden om elevernes faglighed (Børne- og Undervisningsministeriet, 2020), hvilket vores resultat stemmer overens med.

Derudover finder rapporten ingen forskel, uanset hvor langt skolerne er nået med implementeringen (Børne- og Undervisningsministeriet, 2020, pp. 15-16). Skolernes oplevelse af implementeringen er præget af manglende ressourcer og struktur for hvordan indsatsområderne opnås. Der efterlyses retningslinjer for implementering af tiltag i undervisningen herunder bevægelse i faglige aktiviteter. Dette bidrager også til den manglende implementering, da folkeskolerne i høj grad selv skulle strukturere hvordan og hvornår, det skulle foregå (Børne- og Undervisningsministeriet, 2020, pp. 44). Den manglende målbare effekt begrundes Børne- og Undervisningsministeriet (2020) med, at implementeringen endnu ikke er fuldt gennemført. Rapporten skriver at det tager mellem 5-15 år at implementere større skolepolitiske reformer. Samtidigt forventer de først at se et fald før et egentligt fagligt løft (Børne- og Undervisningsministeriet, 2020, pp. 15). Vores undersøgelse kan være påvirket af samme problematik, da PISA 2018 også kun er 4 år efter reformen blev vedtaget.

En anden forklaring, som rapporten peger på, ift. den manglende effekt er den sideløbende implementering af en række ekstra, nye skolepolitiske tiltag – bl.a. øget inklusion, nedlukninger og sammenlægninger af skoler. Rapporten mener altså ikke, at nedgangen skyldes reformen, men udefrakommende faktorer.

En tredje mulig forklaring, som følgeforskningen ikke kommer ind på, er stigningen i lærermangel de seneste år (Danmarks Lærerforening, 2021, pp.1). Manglende fagligt uddannet personale kan have betydning for om skolerne prioritérer implementeringen. Desuden kan manglen på fagligt uddannet personale gå ud over elevernes udbytte og folkeskolerne kan have svært ved at opfylde de Fælles Mål⁴.

Afslutningsvist bemærkes det, at reformen kom i krydset på lærerlockouten i 2013, hvorfor det kan tænkes, at lærerne ikke var motiveret for at gennemføre reformen og i højere grad søgte job væk fra folkeskolen.

Ingen signifikant ændring i ESCS, S_ESCS, køn og IMMIG Hypoteserne om ingen ændring i betydningen af klassekammerateffekten, socioøkonomisk baggrund, køn og immigrationsstatus fra før til efter Folkeskolereformen oprettholdes. Følgeforskningen har samme overordnede konklusion for socioøkonomisk baggrund, køn og immigrationsstatus. VIVE fastslår desuden, at de tre sidstnævnte variable forklarer 17-20 pct. af karakterforskellen i 9.klasse. (Børne- og Undervisningsministeriet, 2020, pp.18-20)

Rapporten nævner ikke klassekammerateffekten, men da vores konklusioner er ens ift. betydningen af social baggrund – klassekammerateffekten er baseret på ESCS – forventer vi ingen ændring i betydningen af den som følge af reformen. En politik der kan have påvirket variablen, er de kommunalpolitiske tiltag om sammenlægninger af skoler, men det er ikke noget vi undersøger i forbindelse med dette projekt.

Signifikant ændring i METASUM Vi afviser nulhypotesen om ingen ændring i betydningen af metakognitive læsestrategier ved resume af tekst ift. elevers læsescore. Så det er for den gennemsnitlige elev i 2018 vigtigere at være bevidst om effektive strategier ved resume-skrivning, end det er i 2009, for at opnå en høj læsescore.

Totusindeogni var det første PISA-år med variablene for de metakognitive læsestrategier (OECD, 1 sd, pp. 23). Det formodes at eleverne ikke havde kendskab til strategierne på forhånd og svært ved at forstå opgaven. Efterfølgende kan lærerne have lagt vægt på strategierne op til PISA-testene, så eleverne er forberedt på spørgsmålene om nytigheden af strategierne, hvilket i METASUMs tilfælde kan have gjort en forskel. Tiltaget om at kommunikere de fælles

⁴"Fælles Mål er nationale mål, der beskriver, hvad eleverne skal lære i skolens fag og emner samt børnehaveklassen" Børne- og Undervisningsministeriet (2021)

læringsmål ud til eleverne ifm. reformen kan have medvirket til en øget bevidsthed om egen læring og dermed en større bevidsthed om nyttige strategier. (Danmarks Statistik, 1 sd). Vi finder det interessant, at nulhypotesen for den anden variabel vedrørende metakognitive læsestrategier, UNDREM, bliver oprettholdt. Der er altså en forskel indenfor strategityperne.

5 Konklusion

Vores undersøgelse foretager en statistisk sammenligning vha. hypotesetest af danske skoleelevers læsefærdigheder før og efter Folkeskolereformen 2014. Vi baserer sammenligningen på PISA-undersøgelsen fra 2009 samt 2018 og finder, at variablene for socioøkonomisk baggrund, klassekammerateffekt, køn, immigrationsstatus og metakognitive læsestrategier bidrager til forklaringen af elevers læsescore i PISA for 2009 og 2018.

Vores analyser viser ingen signifikant ændring i læsescoren over tid, se afsnit 3.3. Vi finder ingen ændring i betydningen af socioøkonomisk baggrund på læsescoren og vi kan ikke påvise en ændring i betydningen af social baggrund ift. faglige resultater. Denne konklusion er lig følgeforskningen vedr. Folkeskolereformen 2014's implementering.

For de resterende variable – klassekammerateffekten, immigrationsstatus, køn og UNDREM – finder vi ligeledes ingen signifikant ændring i betydningen på læsescoren fra 2009 til 2018.

Vi mäter en signifikant ændring i betydningen af metakognitive læsestrategier for resume af tekst (METASUM) fra før til efter Folkeskolereformen. Vi ser, at den gennemsnitlige elev vil opleve en stigning på 4,27 point i deres læsescore fra 2009 til 2018, som følge af metakognitive læsestrategier. Om denne signifikante ændring skyldes Folkeskolereformen, kan vi ikke konkludere.

6 Litteratur

- Andersen, H. (2013). *Klassisk og moderne samfundsteori*. Hans Reitzels Forlag.
- Bay, H. (2018). *Klassekammerateffekten i PISA*. Danmark: Symposium i anvendt statistik.
- Berlingske (Tilgået d. 21. oktober 2021, skrevet 3. december 2013). Artikel: "Antorini: Vi skal gøre skolen endnu bedre" af Cathrine Bloch. berlingske.dk/politik/antorini-vi-skal-goere-skolen-endnu-bedre.
- Børne- og Undervisningsministeriet (2020). Folkeskolens udvikling efter reformen – En vidensopsamling om folkeskolereformens følgeforskningsprogram 2014-2018. uvm.dk/-/media/filer/uvm/aktuelt/pdf21/maj/210528-folkeskolens-udvikling-efter-reformen-en-vidensopsamling-om-folkeskolereformen.pdf.
- Børne- og Undervisningsministeriet (Tilgået 3. november 2021, sidst opdateret 25. juni 2021). Hjemmeside "om fælles mål". uvm.dk/folkeskolen/fag-timetal-og-overgange/faelles-maal/om-faelles-maal.
- Christensen, V. T. (december 2019). Pisa 2018. Danske unge en international sammenligning. Bind 1 - Resultatrapport. uvm.dk/-/media/filer/uvm/int/191203-pisa-2018-danske-unge-i-en-international-sammenligning.pdf.
- Danmarks Lærerforening (1. november 2021). Analyse: "4 ud af 10 læreruddannede arbejder ikke i folkeskolen" af Mie Dalskov Pil og Anton Lykkehoff. dlf.org/pressemeddelelser/2021/november/antallet-af-lærere-uden-for-folkeskolen-stiger.
- Danmarks Statistik (Tilgået d. 31. oktober 2021, s.d.). Hjemmeside om "PISA 2022: Sådan foregår testen". dst.dk/da/Sites/pisa-2022/testen.
- Dansk Clearinghouse, R. (maj 2014). Forskningskortlægning matematik, (mathematical literacy). dpu.au.dk/fileadmin/edu/Udgivelser/Clearinghouse/Ramboell/FORSKNINGSKORTLAEGNING_MATEMATIK_MATHEMATICAL_LITERACY.
- Egelund, N. (2010). Pisa 2009. Danske unge – en international sammenligning, Bind 2 - Teknisk rapport. uvm.dk/-/media/filer/uvm/udd/folke/pdf11/110920-pisa-teknisk-rapport-2009-ashx.pdf.
- Folkeskolen.dk (3. december 2013). Artikel: "Antorini: Reformen bør have danske Pisa-resultater" af Maria Becher Trier. folkeskolen.dk/537423/antorini-reformen-boer-haeve-danske-pisa-resultater.
- Geisnes, D. og Olund Kirkegaard, P. (2017). Metakognition, selvregulering og konsolidering – fra teori til praksis. ucn.dk/Files/Billeder/ucn/Kurser-og-videreuddannelser/Metakognition-selvregulering-konsolidering-BUTM.pdf.
- OECD (2010). Pisa 2009 results: Learning to learn. <https://www.oecd.org/pisa/pisaproducts/48852630.pdf> oecd.org/pisa/pisaproducts/48852630.pdf
- OECD (Tilgået d. 26. oktober 2021, s.d.). Instrument development and summary of results from the PISA-based test for schools pilot in 2012. oecd.org/pisa/aboutpisa/PISA-Based%20test%20for%20Schools%20Instrument%20Design%20and%20Development.pdf.
- Regeringen (A, B, F) samt V og O (7. juni 2013). Fagligt løft af folkeskolen. uvm.dk/-/media/filer/uvm/udd/folke/pdf14/okt/141010-endelig-aftaletekst-7-6-2013.pdf.
- Rosenkrantz-Theil, P. (27. maj 2021). Redegørelse til folketingen om folkeskolereformen. altinget.dk/misc/210528-Redegeorelse-til-Folketingen-om-folkeskolereformen.pdf.
- Steiness, S.A. og Palmø, J. (november 2021). En statistisk sammenligning af danske skoleelevers læsefærdigheder før og efter folkeskolereformen 2014. Ikke publiceret, adgang ved kontakt til forfattere.
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

The Eigenvector-Eigenvalue Identity Applied to Fast Calculation of polSAR Scattering Characterization

Allan Aasbjerg Nielsen[✉]

Abstract—Unlike the original Cloude-van Zyl decomposition of reflection symmetric polarimetric synthetic aperture radar (polSAR) data, a recently suggested version of the decomposition for full/quad pol data relies on the Cloude-Pottier mean alpha angle ($\bar{\alpha}$) to characterize the scattering mechanism. $\bar{\alpha}$ can be calculated from the eigenvectors of the coherency matrix. By means of the eigenvector-eigenvalue identity (EEI), we can avoid the calculation of the eigenvectors. The EEI finds $\bar{\alpha}$ by means of eigenvalues of the 3×3 coherency matrix and its 2×2 minor(s) only and is well suited for fast array-based computer implementation. In this letter with focus on computational aspects, we demonstrate fast EEI-based determination of $\bar{\alpha}$ on X-band Flugzeug synthetic aperture radar (F-SAR) image data over Vejers, Denmark, including a detailed example of calculations and computer code.

Index Terms—Anisotropy, coherency matrix, complex covariance matrix, entropy, Flugzeug synthetic aperture radar (F-SAR), Hermitian matrix, mean alpha angle ($\bar{\alpha}$), polarimetric synthetic aperture radar (polSAR), X-band.

I. INTRODUCTION

IN THE covariance matrix formulation of multilook polarimetric synthetic aperture radar (polSAR) image data each pixel is described by a complex 3×3 matrix

$$\mathbf{C} = \begin{bmatrix} \langle S_{hh} S_{hh}^* \rangle & \sqrt{2} \langle S_{hh} S_{hv}^* \rangle & \langle S_{hh} S_{vv}^* \rangle \\ \sqrt{2} \langle S_{hv} S_{hh}^* \rangle & 2 \langle S_{hv} S_{hv}^* \rangle & \sqrt{2} \langle S_{hv} S_{vv}^* \rangle \\ \langle S_{vv} S_{hh}^* \rangle & \sqrt{2} \langle S_{vv} S_{hv}^* \rangle & \langle S_{vv} S_{vv}^* \rangle \end{bmatrix}.$$

This matrix is Hermitian also known as self-adjoint, i.e., the matrix is equal to its own conjugate transpose, $\mathbf{C} = \mathbf{C}^H$, the superscript H denotes transpose and complex conjugate (which is denoted by *), $\langle \cdot \rangle$ denotes ensemble averaging. The trace, the determinant, and all eigenvalues are real and nonnegative.

The Hermitian coherency matrix is $\mathbf{T} = \mathbf{NC}^T$ [1] where

$$\mathbf{N} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & \sqrt{2} & 0 \end{bmatrix}.$$

The full/quad pol covariance matrix \mathbf{C} and the coherency matrix \mathbf{T} have the same eigenvalues.

The Cloude-Pottier decomposition [1], [2] of the coherency matrix \mathbf{T} gives us entropy, H , anisotropy, A , and the mean alpha angle, $\bar{\alpha}$. The sorted eigenvalues of \mathbf{T} are $\lambda_1 \geq \lambda_2 \geq \lambda_3$. As a measure of target disorder, entropy H is defined as

$$H = - \sum_{i=1}^3 p_i \log_3 p_i \quad \text{with} \quad p_i = \frac{\lambda_i}{\sum_{i=1}^3 \lambda_i}$$

$0 \leq H \leq 1$. Low entropy is associated with a single dominant scattering mechanism, high entropy with random scattering.

Manuscript received March 1, 2022; revised April 7, 2022; accepted April 20, 2022. Date of publication April 22, 2022; date of current version May 12, 2022.

The author is with the DTU Compute, Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Lyngby, Denmark (e-mail: alan@dtu.dk).

Digital Object Identifier 10.1109/LGRS.2022.3169994

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see <https://creativecommons.org/licenses/by/4.0/>

Anisotropy A can be considered a measure of departure from azimuthal symmetry and is defined as

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3}$$

$0 \leq A \leq 1$. Low anisotropy is associated with azimuthal symmetry.

Fast array-based calculation of eigenvalues for 3×3 (and 2×2) Hermitian matrices is described in [3].

The eigenvectors e_i , $i = 1, 2, 3$, of \mathbf{T} are of the form

$$e_i = \exp(j\phi_i) \begin{bmatrix} \cos \alpha_i \\ \sin \alpha_i \cos \beta_i \exp(j\delta_i) \\ \sin \alpha_i \sin \beta_i \exp(j\gamma_i) \end{bmatrix}$$

where α_i , β_i , δ_i , and γ_i are angles which characterize the scattering. Note that the angle ϕ_i is arbitrary: e_i may be rotated by any angle ϕ_i , i.e., e_i may be replaced by $e_i \exp(j\phi_i)$ and still be a solution to the eigenproblem.

Here, we are interested in the α_i only. For the squared norm $|e_{i1}|^2$ of the first component e_{i1} of e_i we have

$$\begin{aligned} |e_{i1}|^2 &= \exp(j\phi_i) \cos \alpha_i \exp(-j\phi_i) \cos \alpha_i \\ &= \cos^2 \alpha_i \end{aligned}$$

(the squared norm is independent of the arbitrary rotation angle ϕ_i). For $\bar{\alpha}$, we have

$$\bar{\alpha} = \sum_{i=1}^3 p_i \alpha_i \quad \text{with} \quad \alpha_i = \arccos |e_{i1}|.$$

$\bar{\alpha} = 0^\circ$ is associated with single bounce (surface) scattering, $\bar{\alpha} = 45^\circ$ with scattering from differently oriented dipoles (typically vegetation), and $\bar{\alpha} = 90^\circ$ with (conductive) double bounce scattering.

Calculating eigenvectors is more cumbersome than calculating eigenvalues, and it seems we need the first component of the eigenvectors to calculate the α_i and $\bar{\alpha}$. Sections II–V show that we can obtain these angles without the eigenvectors by means of the so-called eigenvector-eigenvalue identity (EEI) and give computationally oriented examples.

II. EIGENVECTOR-EIGENVALUE IDENTITY

Let \mathbf{H} be an $n \times n$ Hermitian matrix with eigenvalues $\lambda_1(\mathbf{H}) \geq \dots \geq \lambda_n(\mathbf{H})$ and eigenvectors e_i (we reuse the symbols e_i). We write e_{ij} for the j th component of e_i . \mathbf{H} has $(n-1) \times (n-1)$ minors \mathbf{M}_j , $j = 1, \dots, n$, made from \mathbf{H} by deleting the j th row and the j th column from \mathbf{H} . These minors are also Hermitian with eigenvalues $\lambda_1(\mathbf{M}_j) \geq \dots \geq \lambda_{n-1}(\mathbf{M}_j)$. For these eigenvalues we have the so-called Cauchy interlacing inequalities

$$\lambda_i(\mathbf{H}) \geq \lambda_i(\mathbf{M}_j) \geq \lambda_{i+1}(\mathbf{H})$$

for $i = 1, \dots, n-1$.



Fig. 1. Part of Vejers, Denmark, approximately the area covered by the F-SAR data, from Google Earth. We see (summer) houses, caravan sites with a swimming pool and service buildings (a grocery store and a petrol station), the main road called Vejers Havvej (where the text "Vejers" appears) going mostly east-west, dirt roads, vegetation, etc.

According to [4]–[6] we have the following so-called eigenvector-eigenvalue identity (EEI) for the squared norm $|e_{ij}|^2$ of the components of the eigenvectors e_i :

$$|e_{ij}|^2 \prod_{k=1, k \neq i}^n (\lambda_i(\mathbf{H}) - \lambda_k(\mathbf{H})) = \prod_{k=1}^{n-1} (\lambda_i(\mathbf{H}) - \lambda_k(\mathbf{M}_j)).$$

A related expression due to Jacobi appeared already in 1834.

As stated above the phase of any individual e_i is arbitrary, therefore the relative phase between e_{ik} and e_{jk} , $i \neq j$, is arbitrary. However, the relative phases between the components of any e_i , say between e_{ij} and e_{ik} for $j \neq k$, is not arbitrary.

Despite the simple nature of the EEI and the mature state of development of linear algebra, this identity was not widely known until recently [6].

For the Cloude-Pottier decomposition of full/quad polSAR data, $\mathbf{H} = \mathbf{T}$, the 3×3 coherency matrix and the minor \mathbf{M}_1 comes from \mathbf{T} by removing the first row and the first column, so \mathbf{M}_1 is 2×2 . In this case, we get for the squared norm of the first component of each of the three eigenvectors

$$\begin{aligned} |e_{11}|^2 &= \frac{(\lambda_1(\mathbf{T}) - \lambda_1(\mathbf{M}_1))(\lambda_1(\mathbf{T}) - \lambda_2(\mathbf{M}_1))}{(\lambda_1(\mathbf{T}) - \lambda_2(\mathbf{T}))(\lambda_1(\mathbf{T}) - \lambda_3(\mathbf{T}))} \\ |e_{21}|^2 &= \frac{(\lambda_2(\mathbf{T}) - \lambda_1(\mathbf{M}_1))(\lambda_2(\mathbf{T}) - \lambda_2(\mathbf{M}_1))}{(\lambda_2(\mathbf{T}) - \lambda_1(\mathbf{T}))(\lambda_2(\mathbf{T}) - \lambda_3(\mathbf{T}))} \end{aligned}$$

$$|e_{31}|^2 = \frac{(\lambda_3(\mathbf{T}) - \lambda_1(\mathbf{M}_1))(\lambda_3(\mathbf{T}) - \lambda_2(\mathbf{M}_1))}{(\lambda_3(\mathbf{T}) - \lambda_1(\mathbf{T}))(\lambda_3(\mathbf{T}) - \lambda_2(\mathbf{T}))}.$$

These expressions are well suited for fast array-based computer implementation. We obtain the a_i and \bar{a} from $|e_{11}|$ as above in Section I, here without calculating the eigenvectors.

III. IMAGE DATA EXAMPLE

In a project headed by the Danish Ministry of Defense Acquisition and Logistics Organization (DALO) the Flugzeug synthetic aperture radar (F-SAR) system of the German Aerospace Center (DLR) has been used to acquire data at several locations in Denmark and Greenland. The F-SAR offers the possibility of exploring the performance of a high resolution, fully polarimetric SAR system with five frequency bands in the range from ~ 300 to ~ 10 GHz [7].

The Cloude-Pottier decomposition is illustrated with airborne F-SAR X-band data (frequency 9.60 GHz corresponding to a wavelength of ~ 3 cm). The data were acquired on 20 June 2017 covering a central part of the village Vejers on the Danish west coast, Fig. 1. The scene covers the east-west main road (called Vejers Havvej where the text "Vejers" appears in Fig. 1) going to and from the sea, campsites with caravans to the south of the road, a swimming pool, parking lots, and several service buildings (grocery store, petrol station, etc.) as well as many small and big houses north and south of the road. These

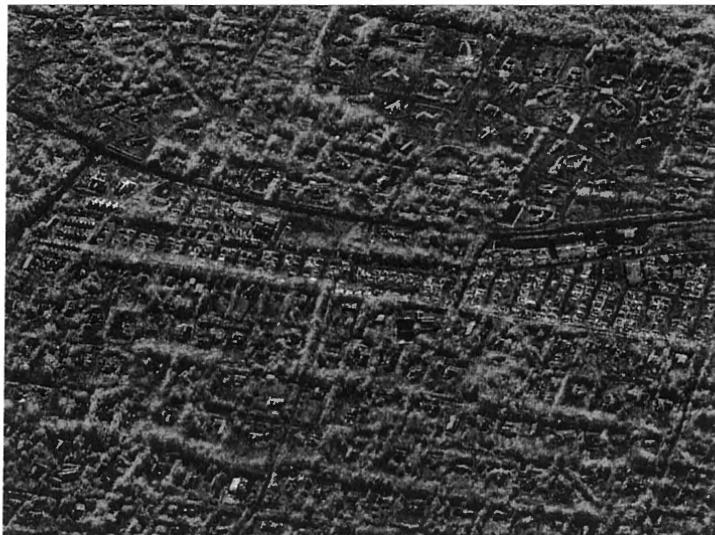


Fig. 2. Full/quad pol version of Cloude-van Zyl decomposition. Red objects are mainly buildings which are characterized by double bounce reflection. Green objects are mainly characterized by volume scattering (typically vegetation), or dihedral targets which are rotated with respect to the flight line which is east-west, here right-left, looking from north. Blue areas are mainly objects which are characterized by single bounce reflection. The yellow caravans, for example, are characterized both by double bounce and rotated dihedrals. Range increases from top to bottom of the image.



Fig. 3. Cloude-Pottier mean alpha angle, $\bar{\alpha}$ in degrees. In accordance with remarks in Section I and with Fig. 2, roads have low values for $\bar{\alpha}$, vegetation has $\bar{\alpha}$ around 45° , and parts of most of the buildings and caravans have very low or very high values (single or double bounce).

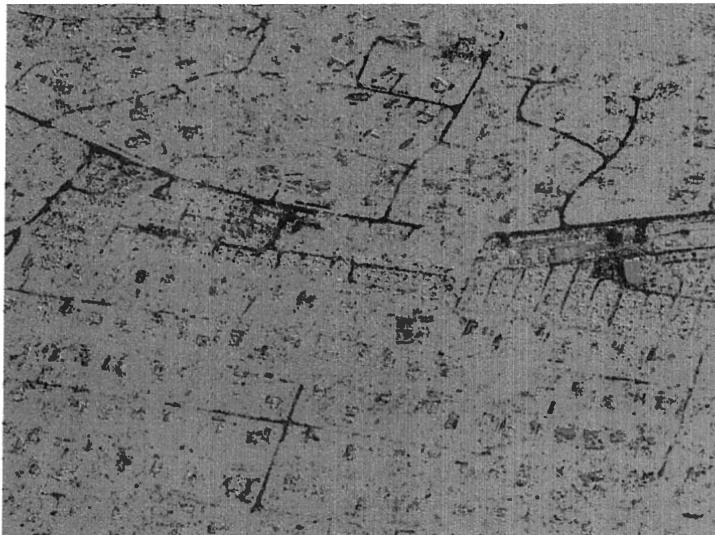


Fig. 4. RGB representation of the Cloude–Pottier decomposition of the coherency matrix. Red is anisotropy (A), green is entropy (H), blue is mean alpha angle ($\bar{\alpha}$). Man-made structures, mostly houses and caravans, appear red-magenta-purple-blue (i.e., high or intermediate values of A and/or $\bar{\alpha}$ with low values of H), vegetation green (i.e., high values of H with lower values of A and $\bar{\alpha}$), and roads black-dark gray (i.e., low values of all three parameters).

data are dealt with also in [8] which introduces a full/quad pol version of the (original, reflection symmetry-based) Cloude–van Zyl decomposition [9], [10].

Fig. 2 shows this full/quad pol version of the Cloude–van Zyl decomposition for a 3000 rows by 4800 columns scene. Red objects are mainly buildings which are characterized by double bounce reflection. Green objects are mainly characterized by volume scattering (typically vegetation), or dihedral targets which are rotated with respect to the flight line. Blue areas are mainly objects that are characterized by single bounce reflection. The yellow caravans, for example, are characterized both by double bounce and rotated dihedrals.

Fig. 3 shows the Cloude–Pottier mean alpha angle ($\bar{\alpha}$). In accordance with remarks in Section I and with Fig. 2, roads have low values for $\bar{\alpha}$, vegetation has $\bar{\alpha}$ around 45°, and parts of most of the buildings and caravans have very low or very high values (single or double bounce).

Fig. 4 shows anisotropy (A), entropy (H), and mean alpha angle ($\bar{\alpha}$) as red-green-blue (RGB). Man-made structures, mostly houses and caravans, appear red-magenta-purple-blue (i.e., high or intermediate values of A and/or $\bar{\alpha}$ with low values of H), vegetation green (i.e., high values of H with lower values of A and $\bar{\alpha}$), and roads black-dark gray (i.e., low values of all three parameters).

Array-based calculations take around 0.37 μ s per pixel for determination of H , $\bar{\alpha}$ and A carried out in eight-byte precision with MATLAB R2022a on a MacBook Pro from 2019, 2.3 GHz 8-Core Intel i9 processor, 64 GB 2667 MHz DDR4 memory. This is a speed-up of a factor of ~55 over an implementation with calls to MATLAB function svd in for-loops over all pixels.

IV. WORKED EEI EXAMPLE

For a detailed illustration of the EEI calculations, we select a pixel on a roof with double bounce ($H = 0.0573$, $\bar{\alpha} = 87.2^\circ$ and $A = 0.6946$). For this pixel, we have covariance and coherency matrices as shown at the top of the next page. From MATLAB function eig, we get eigenvalues $\lambda_i(T)$ for T (first row above the horizontal line) and corresponding (column) eigenvectors e_i

25.7837	0.2325	0.0419
0.0206 – $j0.0306$	-0.9477 – $j0.2958$	-0.1084 – $j0.0347$
0.4709 – $j0.8811$	0.0347 + $j0.0097$	0.0151 – $j0.0197$
-0.0234	-0.1141	0.9932

The eigenvectors from eig are rotated such that the third components are real. Based on these eigenvectors we get

$$\begin{aligned} |e_{11}|^2 &= (+0.0206 - j0.0306)(+0.0206 + j0.0306) \\ &= 0.0014 \\ |e_{21}|^2 &= (-0.9477 - j0.2958)(-0.9477 + j0.2958) \\ &= 0.9857 \\ |e_{31}|^2 &= (-0.1084 - j0.0347)(-0.1084 + j0.0347) \\ &= 0.0130. \end{aligned}$$

If we use MATLAB function svd instead, we get the same eigenvalues $\lambda_i(T)$ (first row above the horizontal line) and differently rotated corresponding (column) eigenvectors e_i

25.7837	0.2325	0.0419
-0.0369	0.9928	0.1138
-0.9938 + $j0.1020$	-0.0360 + $j0.0011$	-0.0084 + $j0.0233$
0.0131 + $j0.0194$	0.1089 – $j0.0340$	-0.9459 + $j0.3029$

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} 13.937 & -0.196 + j0.393 & -12.735 - j0.097 \\ -0.196 - j0.393 & 0.059 & 0.207 + j0.358 \\ -12.735 + j0.097 & 0.207 - j0.358 & 12.062 \end{bmatrix} \\ &= \begin{bmatrix} 13.9 & 0.4 \exp(j116.4^\circ) & 12.7 \exp(-j179.6^\circ) \\ 0.4 \exp(-j116.4^\circ) & 0.1 & 0.1 \exp(j60.0^\circ) \\ 12.7 \exp(j179.6^\circ) & 0.4 \exp(-j60.0^\circ) & 12.1 \end{bmatrix} \\ \mathbf{T} &= \begin{bmatrix} 0.2648 & 0.9373 + j0.0967 & 0.0082 + j0.0249 \\ 0.9373 - j0.0967 & 25.7347 & -0.2847 + j0.5311 \\ 0.0082 - j0.0249 & -0.2847 - j0.5311 & 0.0585 \end{bmatrix} \\ &= \begin{bmatrix} 0.26 & 0.94 \exp(j5.9^\circ) & 0.03 \exp(j71.8^\circ) \\ 0.94 \exp(-j5.9^\circ) & 25.74 & 0.60 \exp(j118.2^\circ) \\ 0.03 \exp(-j71.8^\circ) & 0.60 \exp(-j118.2^\circ) & 0.06 \end{bmatrix}. \end{aligned}$$

The eigenvectors from svd are rotated such that the first components are real. Based on these eigenvectors, we get

$$\begin{aligned} |e_{11}|^2 &= (-0.0369)^2 = 0.0014 \\ |e_{21}|^2 &= (+0.9928)^2 = 0.9857 \\ |e_{31}|^2 &= (+0.1138)^2 = 0.0130 \end{aligned}$$

i.e., the same solution as the one based on eig. Rotation angles to obtain eigenvectors based on svd from eigenvectors based on eig are

$$\begin{bmatrix} e_1 & e_2 & e_3 \\ -0.5589 - j0.8292 & -0.9546 + j0.2980 & -0.9524 + j0.3050 \\ \exp(-j124.0^\circ) & \exp(j162.7^\circ) & \exp(j162.2^\circ) \end{bmatrix}.$$

For the $\alpha_i = \arccos |e_{ii}|$ (or $\alpha_i = \cos^{-1} |e_{ii}|$), we get

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 87.8850^\circ \\ 6.8722^\circ \\ 83.4644^\circ \end{bmatrix} \quad \text{with} \quad \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} 0.9895 \\ 0.0089 \\ 0.0016 \end{bmatrix}$$

resulting in $\bar{\alpha} = 87.2^\circ$.

\mathbf{T} has first minor (delete first row and column from \mathbf{T})

$$\mathbf{M}_1 = \begin{bmatrix} 25.7347 & -0.2847 + j0.5311 \\ -0.2847 - j0.5311 & 0.0585 \end{bmatrix}$$

with eigenvalues $\lambda_i(\mathbf{M}_1) = [25.7489 \ 0.04438]$. Note that the Cauchy interlacing inequalities hold.

For the squared norm $|e_{11}|^2$ of the first components e_{ii} of \mathbf{e}_i the EEI gives

$$\begin{aligned} |e_{11}|^2 &= \frac{(25.7837 - 25.7489)(25.7837 - 0.04438)}{(25.7837 - 0.2325)(25.7837 - 0.0419)} \\ &= 0.0014 \\ |e_{21}|^2 &= \frac{(0.2325 - 25.7489)(0.2325 - 0.04438)}{(0.2325 - 25.7837)(0.2325 - 0.0419)} \\ &= 0.9857 \\ |e_{31}|^2 &= \frac{(0.0419 - 25.7489)(0.0419 - 0.04438)}{(0.0419 - 25.7837)(0.0419 - 0.2325)} \\ &= 0.0130 \end{aligned}$$

i.e., the same solution as obtained by means of the (differently rotated) eigenvectors of the coherency matrix.

V. CONCLUSION

The Cloude-Pottier mean alpha angle, $\bar{\alpha}$, can be determined by means of the first components of the eigenvectors \mathbf{e}_i of the 3×3 Hermitian coherency matrix. These eigenvectors may be rotated by any angle ϕ_i , i.e., \mathbf{e}_i may be replaced by $\mathbf{e}_i \exp(j\phi_i)$ and still be a solution to the eigenproblem.

Alternatively, $\bar{\alpha}$ may be calculated based on the eigenvalues of the coherency matrix and its first 2×2 minor by means of the so far little known EEI without calculating eigenvectors. This facilitates fast array-based computer implementation of the $\bar{\alpha}$ angle calculation.

MATLAB code with this array-based implementation will be made available from the author's homepage <https://people.compute.dtu.dk/alan> under "Publications".

REFERENCES

- [1] J.-S. Lee, M. R. Grunes, T. L. Ainsworth, L.-J. Du, D. L. Schuler, and S. R. Cloude, "Unsupervised classification using polarimetric decomposition and the complex Wishart classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 5, pp. 2249–2258, Sep. 1999, doi: 10.1109/36.789621.
- [2] R. S. Cloude and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 68–78, Jan. 1997, doi: 10.1109/36.551935.
- [3] A. A. Nielsen, "Fast matrix based computation of eigenvalues and the Loewner order in PolSAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 242–246, Feb. 2020, doi: 10.1109/LGRS.2019.2952202.
- [4] K. T. Löwner, "Über monotone matrixfunktionen," *Mathematische Zeitschrift*, vol. 38, pp. 177–216, Dec. 1934. [Online]. Available: <https://eudml.org/doc/168495>
- [5] S. Hensley, "The eigenvector-eigenvalue identity and radar polarimetry," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2021, pp. 2–5, doi: 10.1109/IGARSS47720.2021.9554916.
- [6] P. B. Denton, S. J. Parke, T. Tao, and X. Zhang, "Eigenvectors from eigenvalues: A survey of a basic identity in linear algebra," *Bull. Amer. Math. Soc.*, vol. 59, no. 1, pp. 31–58, Jan. 2022, doi: 10.1090/bull/1722.
- [7] A. Reigber et al., "Very-high-resolution airborne synthetic aperture radar imaging: Signal processing and applications," *Proc. IEEE*, vol. 101, no. 3, pp. 759–783, Mar. 2013, doi: 10.1109/JPROC.2012.2220511.
- [8] A. A. Nielsen, P. J. Connetable, H. Skriver, and K. Conradsen, "A version of the Cloude/van Zyl decomposition for full/quasi pol synthetic aperture radar data," Tech. Univ. Denmark, Tech. Rep., Lyngby, Denmark, Apr. 2022, doi: 10.11581/DTU.00000237.
- [9] J. van Zyl, "Application of Cloude's target decomposition theorem to polarimetric imaging radar data," *Proc. SPIE*, vol. 1748, pp. 184–191, Feb. 1993, doi: 10.1117/12.140615.
- [10] J. J. van Zyl, M. Ariji, and Y. Kim, "Model-based decomposition of polarimetric SAR covariance matrices constrained for nonnegative eigenvalues," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3452–3459, Sep. 2011, doi: 10.1109/TGRS.2011.2128325.

Anders Milhøj

anders.milhøj@econ.ku.dk

Nyheder i SAS

Hvornår kommer nyhederne

For nogle år siden kom nyhederne til de analytiske procedurer i de nye opdateringer af hele SAS-systemet til fx en Windows-PC. I de senere år kom nyhederne i separate opdateringer af de analytiske pakker hyppigere typisk med højest to års mellemrum. Disse opdateringer indeholdt på en gang forbedringer af alle pakkerne STAT; ETS, OR etc, så der kom så mange forbedringer på en gang på brugerens PC, at reinstallationen var ulejligheden værd. Disse opdateringer er omtalt i mange tidligere SAS-indlæg ved Symposierne i Anvendt Statistik. Den seneste kom med opdateringen til version 15.1 af de analytiske procedurer i november 2018, som tidligere omtalt i symposieindlæg.

I august 2020 kom endnu en opdatering til den nuværende version 15.2 af de analytiske procedurer, men den var blot en teknisk opdatering, som rettede diverse fejl og ikke indeholdt decidederede nyheder. Opdateringen var ellers væsentlig nok, fordi den også indeholdt et endelig farvel til Windows-SAS afhængighed af Java. Denne afhængighed har gennem årene lejlighedsvis givet fejl i behandlingen af output fra SAS, når Java versionen på brugerens PC kom ud af takt med SAS-versionen.

I de senere år har SAS satset voldsomt på den nye cloud-baserede platform SAS Viya. På denne platform findes Visual Analytics, som er brugbar til at danne grafik intuitivt og hurtigt. SAS Viya indeholder også mange muligheder inden for Machine Learning med muligheder for fx neurale netværk, der tidligere kun kunne tilgås via Enterprise Miner. Samlet kaldes den del af SAS Viya for 'SAS® Visual Data Mining and Machine Learning' forkortet VDMML.

SAS Viya indeholder også en adgang til editoren SAS STUDIO og dermed en adgang til en kodebaseret SAS svarende til en installation på Windows-PC, så det er muligt at kombinere allerede udviklet SAS-kode med de nyere faciliteter. SAS Viya indeholder også muligheder for vekselvirking med 'open source' miljøet, idet der fx findes en Proc Python, så SAS-applikationer kan vekselvirke med Python-kode. Man kan også kombinere med Jupyter notebooks.

SAS-Viya kører i 'skyen' dvs på SAS-servere, som de fleste brugere ikke behøver interesser sig for. Disse servere kan køre parallelt, så det samlede system af store brugere kan anvendes som en stor 'talknuser'. I systemets terminologi kaldes det samlede system *SAS Cloud Analytics Service*, forkortet CAS, som alt køres på. Data omlægges til en særlig filtype 'sashdat', når fx SAS-datasæt med filtypen 'sas7bdat' lægges op.

Bogstavet H i filtypen sashdat står for Hadoop, som er Apaches cloudbaserede databasesystem symboliseret ved en gul elefant.

Kommercielle SAS-brugere kan naturligvis købe sig ind i de dele af SAS Viya, de har brug for. Min adgang til Viya går gennem portalen 'SAS Viya for Learners', der er gratis til rådighed for lærere og studerende. Denne vej giver adgang til det meste, men ikke alt, og den er naturligvis begrænset, så den ikke giver adgang til store talknuser-opgaver, så den ikke kan misbruges til kommersielt brug.

På denne platform kommer drypvist mindre nyheder i form af nye procedurer og nye optioner til eksisterende procedurer. De nyeste er fra sommeren 2022, se Figur 1, men de er dog endnu ikke alle til rådighed for 'gratister' som mig. Et 'Action SET', som der ifølge Figur 1 er mange nye af, er en stump kode, der kalder SAS-procedurer; fx for Proc HMM, som vist i Figur 1. Ideen er, at disse action sets kan kaldes fra en session uden at burgeren nødvendigvis kender meget til syntaksen. En session kan i denne forbindelse også foregå i R eller i Python.

Figur 1 Udpluk af nyheder i SAS Viya sommeren 2022

The screenshot shows a web browser window with the URL https://documentation.sas.com/doc/en/pgmsascdc/v_029/pgmssw/n06mb8hpokgy6n0zls9o6obgv4v.htm. The page title is "What's New in SAS Viya Programming". On the left, there is a sidebar with navigation links for "Welcome to SAS Programming Documentation", "What's New", "Long-Term Support Releases" (with "2022.1 (May 2022)" highlighted), and "Stable Releases". The main content area displays the "2022.1 (May 2022)" release notes, which include sections for "Changes to Application Themes", "SAS Compute Server: Parquet LIBNAME Engine", "SAS Conversation Designer", "Cross-Environment Data Access", "SAS D2 Language", "DS2 and FEDSQL Procedures", and "SAS Econometrics". Under "SAS Econometrics", there are links to "Deep Learning for Econometrics (deepEcon) Action Set", "Hidden Markov Model (hmm) Action Set", "Market Attribution Action (marketattribution) Set", "Qualitative and Limited Dependent Variable Modeling (qjlm) Action Set", and "Sequential Monte Carlo (smc) Action Set". At the bottom of the page, there are links for "Privacy Statement", "Terms of Use", and "Copyright © SAS Institute Inc. All Rights Reserved". The status bar at the bottom right shows the date "02-06-2022".

Et Machine Learning eksempel

I dette eksempel anvendes et datasæt, der bruges til at analysere konsekvenserne for den nye afgørelse fra 24. juni 2022 af USA's Højesteret om abort. Dommen afgjorde, at de amerikanske delstater havde lov til hver især at forbyde/tillade abort, hvilket altså fjør-

nede den føderale ret til fri abort. Før dommen blev afsagt, havde nogle delstater allerede en lovgivning på plads, der straks (eller senest efter 30 dage) ville forbyde abort i delstaten. Andre delstater havde allerede en lovgivning, der fortsat ville tillade abort i delstaten. Men der var mange delstater, der ikke før dommen havde en reaktion parat. Det giver tre kategorier af en responsvariabel.

Variablen grad er en ordnet, kategoriseret variabel, der reflekterer delstaternes holdning til abort efter højesteretsdommen 24/6 - 2022, der ophævede den føderale rettighed til fri abort. Den har værdierne

3 betyder, at delstaten i forvejen har en såkaldt 'trigger' lov, der automatisk forbyder abort (enten straks eller indenfor 30 dage) efter højesteretsdommen.

2 betyder, at delstaten ingen 'beredskabslov' har.

1 betyder, at delstaten har en lov, der på forhånd garanterer fri abort, trods den forventede højesteretsdom.

I eksemplet forsøges disse reaktioner forklaret/forudsagt ved hjælp af en række forklarende variable, heriblandt delstatens vælgerandel til Trump ved præsidentvalget i 2016 og til Biden i 2020. Dette bliver gjort ved brug af en standardmetode, der stilles til rådighed af Viya. Ideen er, at brugeren finder en skabelon, en pipeline, der så skal udfyldes med de data, man ønsker at analysere.

Figur 2 Begyndelsen af oversigten over færdiglavede pipelines

The screenshot shows a web browser window with the URL <https://vle032.vfe.sas.com/ModelStudio/>. The page displays a list of pre-built pipelines under the heading 'The Exchange'. The left sidebar has sections for 'Templates' (Nodes and Pipelines) and 'Data Sources'. The main area lists pipelines categorized by name:

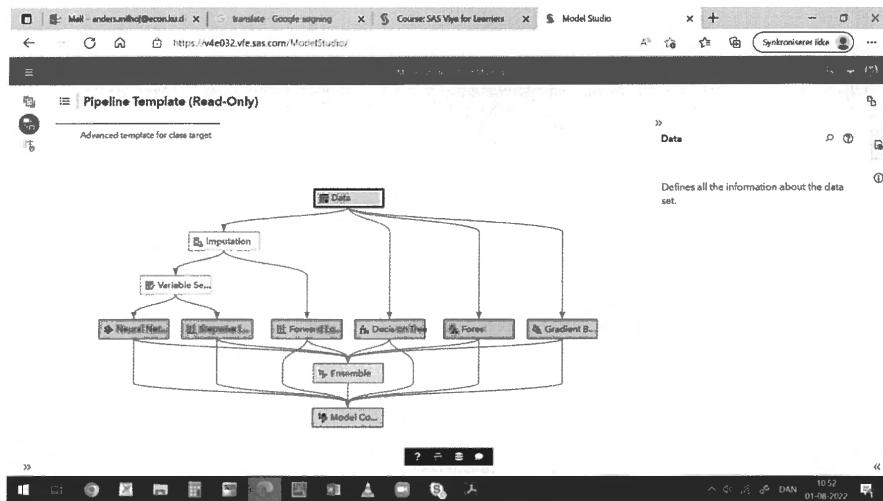
Name	Description
<u>Advanced demo</u>	Data mining pipeline that extends the intermediate template for ...
<u>Advanced template for class target</u>	Data mining pipeline for a class target that contains autotuned tr...
<u>Advanced template for interval target</u>	Data mining pipeline that extends the intermediate template for ...
<u>Advanced template for interval target with autotuning</u>	Data mining pipeline for an interval target that contains autotune...
<u>Advanced template for risk modeling</u>	Risk modeling pipeline that extends the basic risk modeling tem. ..
<u>Basic template for class target</u>	Data mining pipe ine that contains a Data, Imputation, Logistic R..
<u>Basic template for interval target</u>	Data mining pipeline that contains a Data, Imputation, Linear Reg. ..
<u>Basic template for risk modeling</u>	Risk modeling pipeline that contains an Interactive Grouping no...
<u>Blank template</u>	Data mining pipeline that contains only a data node.
<u>boom</u>	
<u>CPML demo</u>	
<u>CPML Demo Pipeline</u>	Logistic regression Pipeline

At the bottom right of the interface, there is a note: "This pipeline was created in the CPML store. It includes a forecastin..."

Figur 2 viser oversigten over hvilke templates, der er til rådighed. Når man ønsker en samlet template og ikke kun en begyndelse på et mere specifikt problem, skal man i menuen vælge 'Pipelines' og til denne opgave vælges underpunktet 'Data Mining and Machine Learning'. I dette tilfælde er ønsket at klassificere de amerikanske delstater alt efter hvilket af de tre niveauer af holdningen til fri abort, de har. Derfor vælges den næstøverste 'Advanced template for class target'.

På den måde opnår man et billede af en analytisk pipeline, se Figur 3. Den øverste kasse, 'Node' mærket 'Data', bruges til at specificere hvilket datasæt og hvilke variable, der skal bruges i analysen. Den lange linie med seks kasser, nodes, angiver seks forskellige algoritmer/metoder til at udføre klassifikationen. Længst til højre er 'Gradient Boosting', 'Forest' og 'Decision Tree', som alle er rene algoritmer. De tre kasser til venstre er 'Neural network' og to udgaver af logistisk regression, og disse tre tilbyder forbehandling af data i form af imputation af manglende værdier. To af dem har også muligheder for at reducere antallet af potentielle input variable i de to kasser mellem datakassen og den lange linie med metodemuligheder

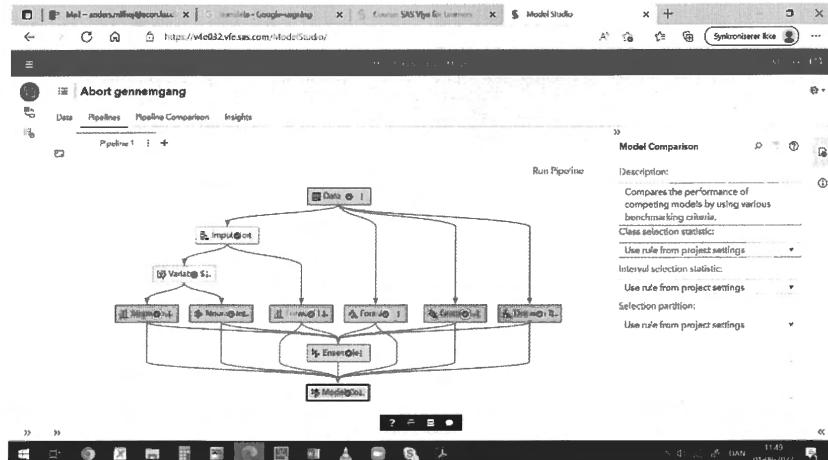
Figur 3 Den analytiske pipeline



Analysen går så ud på, at data indlæses i datakassen øverst og dernæst flyder videre ned til alle seks metodekasser; eventuelt gennem de to kasser med forbehandling af data. Når alle seks analyser er gennemført, kører resultaterne ned til 'Ensemble' kassen, hvori metoderne eventuelt kan kombineres for at forbedre det samlede resultat, der så glider ned til 'Model Comparisons' kassen. Bemærk at den sidste forbedring af resultaterne i Ensemble-kassen ikke er til rådighed for gratister som undertegnede; men resultaterne får heldigvis!

Når pipelinen er gennemløbet ved et klik på 'Run Pipeline', bliver kasserne markeret med et grønt tjek-mærke, dog undtagen Ensemble-kassen, der som nævnt ikke udføres. Den nederste kasse er 'Model Comparisons', som der er klikket på i Figur 4.

Figur 4 Resultatet af model sammenligningerne



Ved at klikke lidt rundt kommer et faneblad frem med navnet 'Insights'. Dette faneblad indeholder et par grafer og resultatet opsummeres ved sætningen *Project Summary: The champion model for this project is Neural Network. The model was chosen based on the KS (Youden) for the Test partition (0.5). The four most important factors are trump_2016, income_2018, violent_cat, and share_White_victim.*

Det er muligt at udvide den viste pipeline i Figur 4 med flere kasser. I dette eksempel vil det være relevant at studere den valgte metode til prædiktion nærmere. Neurale netværk giver som mange andre Machine Learning algoritmer ikke megen information om, hvorledes prædiktionen udføres, da det hele er gemt bort i diverse binære neuroner i flere lag.

For denne type analyser er formålet ofte at undersøge den valgte metode ved at score et nyt datasæt, hvor andre værdier af de forklarende variable bruges til at forudsige målvariablen. Det kan gøres ved at tilføje en kasse med 'Score Data'. I dette konkrete eksempel er det dog komplet irrelevant. I dette eksempel scores blot det brugte datasæt med alle delstater i USA.

Det kan gøres ved at se på et output datasæt, der dannes af proceduren ved at tilføje en ny kasse 'Data Exploration' under kassen med 'Neural Network'. Denne kasse danner middelværdier, varianser etc for alle variable i outputdatasættet, men selve datasættet

vises også ved et klik på fanen 'Output Data'. Med et klik mere ser man de første observationer og variable i outputdatasættet, se Figur 5.

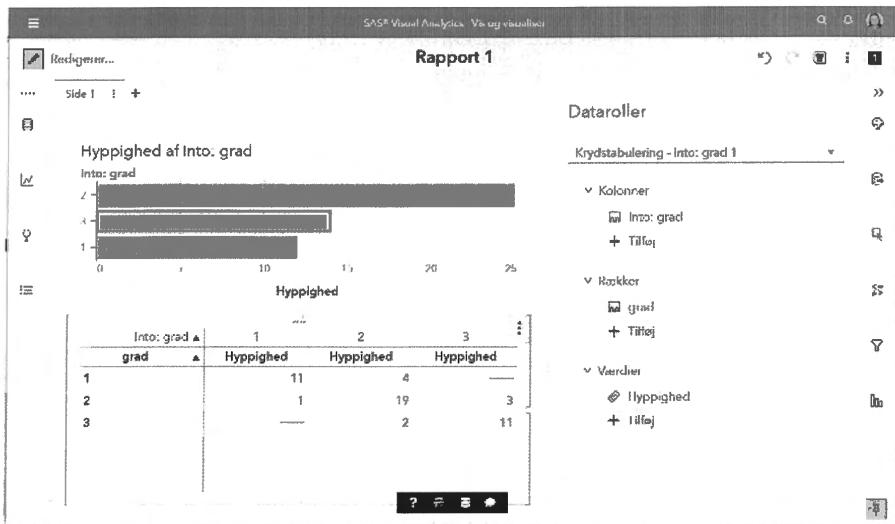
Figur 5 Output datasættet

State	gr...	Into: grad	Predicted: grad=1	Predicted: grad=2	Predicted: grad=3	Pr.
Alabama	2	3	0.0002570472	0.1634050507	0.8343379021	
Alaska	2	2	0.1098253489	0.6359371915	0.2542374596	
Arizona	2	2	0.0278512992	0.642056994	0.3300917068	
Arkansas	3	3	0.0002180547	0.1751486893	0.824633256	
California	1	1	0.895136191	0.1030375313	0.0018262777	
Colorado	1	2	0.1199223331	0.8321659902	0.0479117668	
Connecticut	1	1	0.829548365	0.1599148236	0.0102303399	
Delaware	1	2	0.0742407239	0.8723255365	0.0534337396	

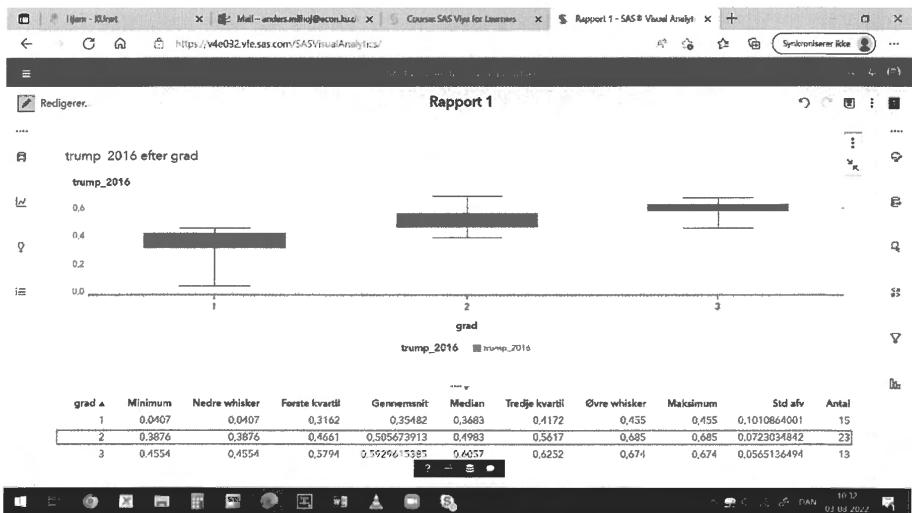
I menuen til venstre er en oversigt over de dannede datasæt, hvoraf der dog kun er et enkelt. Det kan gemmes ved klik på disketten, hvorefter man i en menu skal navigere hen til det areal, man som bruger har adgang til. Her lægges datasættet med filtypen sashdat jf. ovenfor. Dette datasæt kan så studeres enten ved at downloade datasættet, hvilket kan lade sig gøre som fx en csv fil, eller ved at analysere det med Visual Analytics for at blive inden for rammerne af Viya eller med SAS kode i SAS Studio i Viya. I det følgende anvendes Visual Analytics i SAS Viya.

Figur 6 viser en krydstabel med den korrekte grad som rækker og den forudsagte grad som søjler. Det ses, at der ikke er stater, der fejlklassificeres fra det ene yderpunkt til det andet yderpunkt; men der er stater, der misklassificeres mellem midtergruppen og ydergrupperne. For at se hvordan metoden rent faktisk virker, angiver Figur 7 og Figur 8 gennemsnittene af Trumps vælgerandel i 2016 for hver af kategorierne i den originale variable grad og for den forudsagte grad og også boxplots for vælgerandelen. Tabbellerne viser, at Trumps vælgerandel i 2016, som ifølge analyserne var den mest benyttede i kategoriseringerne, viser det forventede, nemlig at stater med høj tilslutning til Trump er mest markante abortmodstandere, mens stater med meget lav tilslutning til Trump er mest markante aborttilhængere.

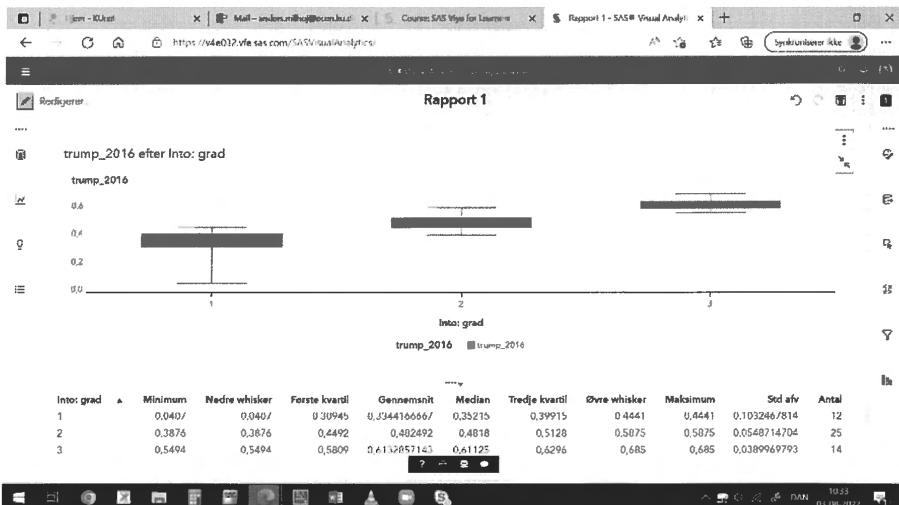
Figur 6 Misklassifikationstabell



Figur 7 Trumps vælgerandel i 2016 for hver kategori af grad.



Figur 8 Trumps vælgerandel i 2016 for hver kategori af den forudsagte grad.



Som et kuriosum betragtes i Tabel 1 data for staten Kansas, der havde en vejledende folkeafstemning d. 2. august 2022, dvs efter Højesterets dom, om abortspørgsmålet i forbindelse med primærvælgene før midtvejsvalget til november, hvor der bl.a. skal vælges ny guvernør.

Tabel 1 Resultater af analysen for staten Kansas

Grad	Into Grad	Ssh for grad 1	Ssh for grad 2	Ssh for grad 3	Trump 2016
2	2	0,001	0,542	0,457	0,567

Resultater af klassifikationen er korrekt, da den korrekte grad 2 blev forudsagt. Det neurale netværk tildeler en umådelig lav sandsynlighed for grad 1, hvilket nok først og fremmest skyldes, at staten valgte Trump i 2016, og dermed at Kansas' parlament nok ikke er så liberalt, at det rent faktisk på forhånd ville tilkendegive, at abort skulle være en rettighed.

Sandsynligheden for at Kansas skulle have haft en trigger lov, der på forhånd sagde, at abort skulle forbydes straks efter højesteretsdommen er stor, 45.7%. Det harmonerer ikke med resultatet af afstemningen 2. august, hvor op imod 60% af stemmerne var for retten til fri abort.

Danske økonometriske makromodeller

i medgang og modgang¹

Professor(em), dr. scient. adm. Jesper Jespersen, Roskilde Universitet

En sen start i slutningen af 1960erne

Økonometriske eksperter og samfundsøkonometriske prognosemodeller har i stigende grad vundet indpas ved tilrettelæggelsen af og begrundelse for den økonometriske politik. Denne artikel tager sit afsæt i disse modellers første spæde indtog i det danske økonommiljø i slutningen af 1960erne. Set i et internationalt, men også skandinavisk perspektiv overraskende sent tidspunkt. Blandt de danske professorer i hhv. København og Aarhus var der ingen tradition for at kombinere makroøkonometrisk teori og data ved brug af statistiske metoder. Naturligvis blev der undervist i 'anvendt statistik' på økonomistudiet; men endnu ikke i makro-økonometri'. I det hele taget var der på økonomistudiet i Danmark fortsat en noget tøvende holdning til brug af matematik (og statistik). Det var f.eks. først med studiereformen i 1970, at matematik blev et obligatorisk fag på polit-studiet.

Der var således adskillige årsager til, at det danske modelarbejde kom så relativt sent i gang. Her er det næppe nogen overdrivelse at pege på, at 'professorvældet', der dominerede frem til slutningen af 1960erne, havde et vist medansvar. Men nok så vigtigt er det at pege på, at udarbejdelsen af årlige nationalregnskabstal først var kommet i gang i løbet af 1950erne. Det betød, at konsistente tidsserier for de væsentligste makroøkonometriske variable ikke gik længere tilbage end 1953. Da makroøkonometri helt overvejende er baseret på tidsrækkeanalyse, så var det først fra midten af 1960erne, at der ville være et tilstrækkeligt antal år (dvs. observationer) og dermed frihedsgrader ved regressionsanalysen til, at det gav mening at benytte statistiske regnemetoder og de dertil knyttede tests for signifikans.

Disse forhold ændrede sig dog afgørende i slutningen af 1960erne. Det skete naturligvis med inspiration fra internationale modelmiljøer, der dengang var centreret omkring Jan Tinbergen i Holland, Lawrence Klein i USA og Ragnar Frisch og Leif Johansen i Oslo, hvor flere yngre forskere havde været på studieophold. Her havde arbejdet med makroøkonometriske modeller stået på i mere end 20 år. For ikke at nævne, at det internationale *Econometric Society* var blevet stiftet helt tilbage i midten af 1930erne.

På de økonometriske institutter i København og Århus, i Det økonometriske Råds sekretariat og sågar i Danmarks Nationalbank blev der i de følgende år ansat yngre medarbejdere med modelerfaring fra udlandet. Modelarbejdet tog yderligere fart med ansættelsen af Erling Jørgensen som kontorchef i det nyoprettede 'analyse og prognosekontor' (10. kontor) i Danmarks Statistik i 1969. Han ønskede at opbygge et dansk (uafhængigt) modelmiljø i samarbejde med lektor Ellen Andersen, hvor nationalregnskabsstatistikken helt naturligt skulle spille den centrale rolle.

Ikke mindst blev kontakten til Lawrence Klein af stor betydning. Det blev organiseret i begyndelsen af 1970erne, at Klein fik et længere ophold ved Økonomisk Institut, KU, hvilket

¹ Denne artikel har i en tidligere og delvis redigeret form været offentliggjort i e-tidsskriftet, *Kritisk Debatt*, april 2022

bidrog til at løfte modelarbejdet. Ikke mindst da det blev sikret, at den danske model (ADAM) i Danmarks Statistik kunne blive en del af Kleins internationale modelprojekt LINK, der sigtede mod at opbygge en verdensmodel, se Klein (1980).

Det danske modelarbejde blev indledt omkring 1970 i både Danmarks Statistik og Det økonomiske Råds sekretariat med opbygning af de keynesianske modeller med betegnelsen hhv. ADAM og SMEC. Begge modeller skulle primært anvendes ved udarbejdelsen af betingede prognoser for den makroøkonomiske udvikling 1-3 år ud i fremtiden og til kvantificering af kortsigtede effekter af den økonomiske politik. Modellernes udformning var i høj grad baseret på keynesianske analyser af den historiske udvikling. Netop fordi sigtet med disse modeller var korte fremskrivninger primært til brug ved de årlige finanslovsforhandlinger blev det efterspørgselskomponenterne, som begge modeller fokuserede på. Udbudssiden var mere rudimentært beskrevet. Dette tidlige modelarbejde er bl.a. dokumenteret i Andersen, 1975a, 1975b og DØR, 1974

Modellerne 'politiseres'

Disse to modeller, der havde mange fælles træk, var dominerende frem til slutningen af 1980erne; men de var begge som nævnt kortsigts-modeller, der ikke i deres konstruktion kunne belyse mere grundlæggende strukturændringer i dansk økonomi. Blev modellerne undtagelsesvist benyttet til beregninger, der rakte 5-10 år ud i fremtiden, var der en åbenbar tendens til, at 'modellen kørte af sporet' og gav nogle resultater, der mildt sagt var svære at fortolke. Det blev i stigende grad åbenbart, at økonomiens udbudsside måtte inddarbejdes i modellerne, for at de mere meningsfuldt kunne belyse mere langsigtede udviklingstendenser i dansk økonomi.

Med regeringsskiftet i 1982 og med betydelig inspiration fra bl.a. Storbritannien og USA begyndte Finansministeriet at efterspørge en model, der var designet til at gennemføre mere langfristede udbudsanalyser. En sådan strukturmodel skulle bruges til at belyse effekten af bl.a. reformer på arbejdsmarkedet og af en grundlæggende ændret pensions- og skattpolitik, der blev diskuteret livligt op gennem 1980erne.

Men statistiske analyser af strukturændringer kræver lange dataserier. Dertil kommer, at samfundsøkonomiske strukturer ofte ændres gennem direkte politiske indgreb, som f.eks. opbygning af velfærdsstaten, internationalt økonomisk samarbejde, som f.eks. medlemskab af EU og lignende. De mere langsigtede effekter heraf kan derfor være svære at fastlægge empirisk ved brug af traditionelle statistiske metoder, hvilket er indeholdt i den såkaldte Lucas-kritik af disse traditionelle makroøkonometriske modeller, se Lucas, 1976. Ønsket om at kunne benytte regnemodellen til sådanne mellemfristede, for ikke at sige langfristede fremskrivninger, satte de akademiske økonomer i et teoretisk dilemma: for hvad ved 'vi' egentlig om det (mellem)lange sigt? Her delte vandede sig mellem neoklassiske og (post)keynesianske økonomer. De ny-neoklassiske økonomer, bl.a. repræsenteret ved Robert Lucas, antager at en velorganiseret markedsøkonomi er selvregulerende og skaber fuld beskæftigelses-ligevægt af sig selv udelukkende bestemt af de underliggende markedsmæssige strukturer. Denne form for strukturmodel indebærer, at politikerne konjunkturpolitisk bør 'forholde sig i ro' og koncentrere sig om at afhjælpe markedsfejl gennem strukturreformer. De (post)keynesianske økonomer (f.eks. Joseph Stiglitz) er derimod mindre skråsikre med hensyn til at benytte økonomiske makromodeller til lange fremskrivninger, idet den underliggende usikkerhed både med hensyn til den fremtidige økonomiske adfærd og struktur er betydelig. Denne fundamentale usikkerhed indebærer, at jo længere der ses ud i fremtiden desto sværere bliver det på et kvalificeret

grundlag at beskrive den samfundsøkonomiske udvikling, se bl.a. Jespersen, 2007 og Jespersen (2022).

Da de ny-neoklassiske økonomer i stigende grad kom til at dominere den makroøkonomiske scene internationalt og ydermere var i samklang med de neoliberalre regeringer, der satte dagsordenen i de toneangivende vestlige lande, så blev det den (generelle) ligevægtsmodel (Lucas-modellen), der vandt indpas i modelmiljøerne. Som en konsekvens heraf blev ADAM i tæt samarbejde med Finansministeriets økonomer ligeledes omformet, så den ved længere fremskrivninger (prognoser?) fulgte et på forhånd fastlagt udviklingsforløb, der resulterede i bl.a. fuld (strukturel) beskæftigelsesligevægt, se Rosted, 2021, kap. 8 'Finansministeriets modelanwendung'. Denne á priori-antagelse om langsigtet markedsrigevægt har stor betydning især for resultatet af lange fremskrivninger. Hvis modellen(!) bliver slæbt ud af kurs af en eksogen begivenhed som f.eks. finanskrisen, så vil denne ændrede markedsstruktur sikre, at modellen(!) af sig selv vender tilbage til det af økonomerne på forhånd fastlagte hovedspor. (1) Det betød, at effekten af stød til efterspørgslen f.eks. *demand management politik*, i form af en ekspansiv (eller kontraktiv) finanspolitik, kun vil have en midlertidig effekt på beskæftigelsen – men derimod en varig forøgelse/forbedring af statsgælden.

Som en konsekvens af den generelle ligevægtsteoris stigende dominans på universiteterne og i de internationale institutioner bl.a. OECD, IMF og EU-kommisionen, blev der i 1990erne påbegyndt et arbejde med at opbygge en egentlig generel ligevægtsmodel fra grunden. Denne model fik navnet DREAM. Et akronym for Danish Rational Economic Agent Model, der netop angiver det teoretiske fundament baseret på mikroøkonomiske antagelser (bemærk 'antagelser'), at 'agenter handler rationelt'. Hvor 'rationalitet' er en helt speciel (og i forhold til almen psykologi afvigende) adfærd: individuel nytte- og profitmaksimering opgjort i kroner og ører, og i sin ekstreme (men ofte anvendte) form under antagelse af, at agenterne har sikker viden om fremtiden. Tilsvarende tages forudsætningen om 'generel ligevægt' meget bogstaveligt, idet der ved lange fremskrivninger antages permanent fuld strukturel beskæftigelse sikret gennem fuld pris- og lønfleksibilitet. DREAM-modellen benyttes til langsigtede fremskrivninger af dansk økonomi, hvor 2050 og 2060 er ofte benyttede terminalår. (2) Formålet er ofte at beregne konsekvenserne på den offentlige sektors balance af strukturaændringer for herigennem at belyse den såkaldte finanspolitiske holdbarhed. På mange måder en forudsigelig beregning, idet afsættet er den eksisterende samfundsøkonomiske struktur, der kendtegner dansk økonomi anno 2020. Når holdbarhedsberegningen forberedes, er der på forhånd lagt et spor for alle de eksogene variable bl.a. i form af vændret skatte- og udgiftspolitik frem til 2060. Beregningen gennemføres ved at der ændres på en af de centrale variable f.eks. arbejdsudbudet. Da det er en generel ligevægtsmodel, så vil beskæftigelsen ændres med et tilsvarende antal personer. Ikke overraskende betyder et øget arbejdsudbud en markant forbedring af de offentlige budgetter pga. af øgede skatteindtægter og mindskede sociale udgifter. Og 'vupti' det øger det 'økonomiske råderum', der er bestemt af saldoen på de offentlige budgetter.

Det var den af VK-regeringen nedsatte Velfærdskommission i 2003, der med en betydelig bevilling i ryggen videreførte DREAM-modellen, da den netop havde de langsigtede egenskaber, som regeringen efterspurgte. Det var især aflastningen af de offentlige budgetter, der var i fokus. Regeringen havde i kommissoriet peget på en forøgelse af pensionsalderen og en reduktion af efterlønnen, som mulige strukturaændringer. Da DREAM-modellen netop bygger på antagelsen om, at et øget udbud af arbejdskraft skaber sin egen efterspørgsel (efter arbejdskraft), var der ingen fare for, at disse reformer (i modellen) ville risikere at øge arbejdsløsheden, se f.eks. DREAM (2019). De forventede modelresultater udeblev ikke.

DREAM udskrev tabeller, der viste, hvorledes både beskæftigelsen og de offentlige budgetter ville forbedres i takt at disse såkaldte velfærdsreformer blev gennemregnet. (3)

Kan kortsigts- og langsigtsmodeller forenes?

Finansministeriets ønske om både at rådgive om den kortsigtede finanspolitik og den langsigtede strukturpolitik på et konsistent grundlag skabte et modelmæssigt dilemma, idet ADAMs styrke helt åbenbart var de mere kortsigtede konjunkturvurderinger til brug for de årligt tilbagevendende, finanslovsforhandlinger. Ved brug af den version af ADAM som Danmarks Statistik udarbejdede til brug ved mellemfristede beregninger af dermed af det 'finanspolitiske råderum', var der en stigende utilfreds i finansministeriet med resultaterne: tilpasningen til ligevegt gik for langsomt. Ligesom det var svært at sikre konsistens mellem resultaterne fra ADAM ved de meget lange fremskrivninger og de resultater som DREAM ville generere. Disse beregninger skulle helst ikke være i modstrid med hinanden og samtidig gerne have en forankring i den faktiske danske samfundsøkonomi.

Hidtil havde modeludviklingen som nævnt foregået i et samarbejde mellem Danmarks Statistik og Finansministeriet. Danmarks Statistik stod her som den uafhængige institution, der sikrede kvaliteten af ikke mindst det statistiske grundlag for de estimerede adfærdsrelationer, der udgjorde rygraden i ADAM. Det havde tidligere ført til så forskellige vurderinger mellem de to institutioner, så Finansministeriet i perioder benyttede sin egen tilpassede version af ADAM. Et forhold der for udenforstående iagttagere kunne forekomme som en politisering af modellen, se f.eks. Enhedslistens kritik (2018). Dette modsætningsforhold resulterede i, at der i 2017 blev sat et helt nyt modelarbejde i gang på initiativ af Finansministeriet. Konsulentfirmaet DREAM-gruppen (der reelt har ansvaret for DREAM-modellen) blev bedt om at opbygge en ny model, der har fået navnet MAKRO. Den skal inden for rammen af en og samme model kunne besvare spørgsmål, der relaterer sig til alle tre tidshorisonter og samtidig 'være baseret på en bred empirisk tilgang' og 'modelleringen af adfærd som udgangspunkt være teoretisk velfunderet' (DREAM-gruppen, 2020, s.4). Hvad der i praksis gemmer sig bag disse formuleringer, vil tiden vise, da modellen endnu ikke er færdigudviklet. Men den hidtidige grundantagelse, at modellen altid vender tilbage til hovedsporet i form af fuld realøkonomisk tilpasning i form af en generel (arbejds)markedsrigtigevægt, er der ikke ændret på. Ligesom adfærdsligningernes form og parameterverdier i højere grad vil være teoretisk bestemt (kalibrerede) end den ADAM-version som Danmarks Statistik havde ansvaret for. Forhold, der navnlig har betydning for tilpasningen på arbejdsmarkedet, forbrugssfæren og i udenrigshandlen. Der kan naturligvis også forekomme en vis træghed i tilpasningen til denne generelle markedsrigtigevægt; men MAKRO vil også have den egenskab, at i en lavkonjunktur 'kommer opsvinget til næste år'. Tilsvarende vil en 'ophedning' efterfølges af én 'blød landing', hvis blot politikerne sikrer, at budgetbalancen respekteres.

Kærgård (2020) udtrykker bl.a. i lyset af ovenstående uklare arbejdsdeling mellem Danmarks Statistik og finansministeriet en ikke ubetydelig kritik af det danske modeløkonomiske miljø, hvor han også savner faglige bidrag fra de universitetsansatte økonomers. Han efterlyser en saglig og seriøs diskussion gerne i fagtidsskrifter, idet han er betænkelig ved den dominerende rolle som Finansministeriets og dets model efterhånden har fået/tiltaget sig. Når MAKRO er færdigudviklet, vil der reelt kun være én model i Danmark – en nærmest monopolagtig tilstand. Tidligere havde der dog været et konkurrencelignende forhold mellem ADAM og SMEC, der havde givet anledning til livlige debatter bl.a. i Nationaløkonomisk tidsskrift med efterfølgende

bogudgivelser, bl.a. Jespersen, 1991. Hertil kommer, at 'fyringen' af Danmarks Statistik som den uafhængige garant for kvaliteten af det statistiske arbejde må mane til betænkelighed.

Modellernes fremtid?

Karakteristisk for de hidtil benyttede danske modeller har dog været, at kriser altid er kommet bag på modelbrugerne. Jeg erindrer kun en i dansk sammenhæng markant undtagelse, der ligger helt tilbage til 1973/74, hvor olieprisstigningen ramte dansk økonomi ret kontant. Her forudsagde Det økonomiske Råds formand ved brug af SMEC-modellen, som den absolut eneste, at BNP ville falde i 1974 – for første gang i efterkrigstiden! Det gav SMEC en forrygende start og anerkendelse, som det dog sidenhen har været vanskeligt at fastholde.

Op til krisen i 2008/09, hvor dansk økonomi helt åbenbart kørte i et højt gear, lod rådet fra Finansministeriet (og Det økonomiske Råd) ikke overraskende: 'ro på (i den økonomiske politik) – økonomien vil foretage en blød landing' for sådan var ligevægtsmodellen konstrueret. Men sådan gik det som bekendt ikke i 2008/09. Alligevel, da de værste bankkriser var overvundet, blev Finansministeriets modelversion af ADAM efter bragt i anvendelse, og så lod meldingen fra Finansministeriet atter: 'ro på – opsvinget kommer til næste år'. Det gjorde det mildt sagt heller ikke. Det danske BNP var først fem år senere tilbage på niveauet fra 2007. Professor Jesper Rangvid (2018) (4) har beregnet, at BNP i hvert af årene siden 2008 har ligget ca. 10 pct. under det i 2007 af modellen prognosticerede (dvs. mulige) BNP. Det betyder, at finanskrisen og den efterfølgende økonomiske politik indtil nu samlet set har kostet den danske samfundsøkonomi mere end 2.000 mia. kr. – og uden udsigt til at tabet på noget tidspunkt genvindes, snarere tværtimod.

Så blev dansk økonomi ramt af Corona'en. Atter var gode modelbaserede råd fraværende. En stribé hjælpepakker blev på ren ad-hoc basis skruet sammen kun understøttet af beregninger, der billedligt talt blev gennemført på bagsiden af en A4-kuvert (ligesom i de meget 'gamle' dage). Men nu da det ser ud til, at krisen ikke længere forværres, er Finansministeriets model atter taget i brug. Dansk økonomi vil – ifølge modellen – være tilbage på hovedsporet i 2022, hvorfor der ligefrem regnes med en stramning af finanspolitikken (af hensyn til budgetbalancen), hvorved den midlertidige 'suspension' af Budgetloven kan bringes til ophør.

Om brugen af de hidtidige prognosemodeller i lyset af erfaringerne fra hhv. finans-, statsgælds- og senest Covid19-krisen vil blive ændret i fremtiden, altså at MAKRO vil være out-dated før den overhovedet er taget i brug, kan det være svært at spå om, se dog Mette Frederiksens udtalelse. (5) De politiske vinde er ved at vende sig væk fra den neoliberalere markedsidéologi, og dermed også generel ligevægtsteori, ikke kun i Danmark, men også i USA og EU, ligesom de internationale institutioner allerede er i gang med at revurdere deres anbefalede økonomiske politik efter finans- og navnlig statsgældskrisen, se f.eks. IMF, 2016 (6).

Finanskriserne har i kombination med Covid19-krisen tilsammen slået et vindue op for en revurdering af makroøkonomisk teori, der minder om hhv. 1933 og 1979. Den i 2019 nylitrædte danske regering blev valgt på et program, der sigter mod at ændre den fordelings- og klimapolitiske dagsorden. Her har de hidtil benyttede modeller fejlet: væksten været svag, uligheden steget og miljø- & klimatilstanden forværet både nationalt (og internationalt). Det har slidt på de hidtidige modellers troværdighed. Det ligger derfor i kortene, at der, som en konsekvens af de eksisterende modellers åbenbare mangler og af de ændrede politiske vinde, vil komme et opbrud. Men hvornår og hvordan, tor jeg ikke spå, da det politiske establishment er splittet mellem en national velfærds- (rød) og international (grøn) klimadagsorden. Men ikke uinteressant er det, at det socialdemokratiske medlem af EU-parlamentet Niels Fuglsang (2021) netop har forsvarer en Ph.d.-afhandling, der klart viser, hvor dominerende en indflydelse

Finansministeriets modelbrug har haft på udformningen af den økonomiske politik. Et forhold som han ikke overraskende ser som et demokratisk problem, da samfundsøkonomi mildt sagt ikke er en eksakt videnskab.

Set i lyset af denne mere end 50 års lange erfaring med modelarbejde i Danmark først i Danmarks Statistik, i Det økonomiske Råds sekretariat og efterfølgende brugen af modellerne ikke mindst i finansministeriet igennem de seneste 20 år gjort, sildt på modellernes eller måske snarere modelbrugernes troværdighed. Det skyldes ikke mindst den i mine øjne vildledende brug af modellerne til prognoseformål. Vores – det vil sige makroøkonomernes – viden om, hvad der bestemmer den samfundsøkonomiske udvikling, er stadig meget begrænset. Så derfor at foregive, at netop den i dag eksisterende struktur kan fremskrives 5, 10 for slet ikke at nævne 40 år ud i en uvis fremtid, hvor end ikke miljø- og klimaforandringer er i bedste fald et meget, meget betyget regnestykke. Ja, direkte vildledende når disse forbehold og dermed usikkerhed ikke lægges åbent frem af Finansministeren, når den økonomiske politik forhandles på baggrund af ministeriets modelberegninger.

Her gør jeg gerne den danske makroøkonometris *grand old lady* Ellen Andersens ord til mine:

“Modellerne siger det, der bliver puttet ind i den. Svaret er ikke noget, der kommer ud af den, det er noget, der kommer ind i den. Og det er ikke inde i ADAM. Det kommer fra Finansministeriet.”

(Professor Ellen Andersen, 27. februar 2017 interview i anledning af 80 årsdagen med Kristine Dons Christensen, Zetland)

Appendiks: Danske makroøkonometriske modeller

Det økonomiske Råds sekretariat (DØR):

SMEC I, (kvartals- og konjunkturmodel), 1970

SMEC II, års- og konjunkturmodel, blot 5 makroadfærdsrelationer, 1974

SMEC III, udbygget års- og konjunktur, 1976-

GESMEC, 1980 strukturmodel

Danmarks Statistik:

ADAM, 1971- , årsmodel, der løbende undergår udbygning og forandring i dialog med finansministeriet. Oprindeligt dokumenteret i Andersen, Ellen (1975a), *En model for Danmark 1949-1965*

Finansministeriet:

Egne versioner af ADAM og DREAM, der forventes afløst af MAKRO i 2022(?), en kombineret konjunktur- og strukturmodel baseret på generel ligevægtsteori

Danmarks Nationalbank,

MONA, 1988- , kvartalsmodel

Arbejderbevægelses Erhvervsråd,

Heimdal, en model, der primært skulle belyse Danmarks internationale samhandel. Den mistede aktualitet i takt med at Danmarks betalingsbalance over for udlandet bedredes markant op gennem 1990'erne og videre ind i dette århundrede

Bibliografi

- Andersen, Ellen (1975a), *En model for Danmark 1949-1965*, Studie nr. 21, København: Akademisk Forlag (1. del af disputatsen)
- Andersen, Ellen (1975b), *Træk af makroøkonometriske modellers historie og udvikling*, Studie nr. 20, København: Akademisk Forlag (2. del af disputatsen)
- Andersen, Ellen (1991), Modeludviklingen i Danmark, i Jespersen (red.), *Model og virkelighed – træk af debatten om de økonomiske modeller*, Jurist- og Økonomforbundets Forlag
- Christensen, Kristine Dons (2017), Regnemodellernes moder advarer mod at stole for meget på den model, hun selv har skabt, *Zetland*, 27. februar.
- Det økonomiske Råds sekretariat (DØR), 1974, *SMEC II*, København: egen udgivelse
- DREAM-gruppen (2019), *Ændringer i erhvervsdeltagelsen siden årtusindeskiftet*, arbejdspapir, 2019:4
- DREAM-gruppen (2021), *MAKRO: Oversigt over model og centrale modelleringsvalg*, Dokumentationsnotat, 3. dec.
- Enhedslisten (2018), *Den politiske regnemaskine*, Christiansborg, <https://enhedslisten.dk/den-politiske-regnemaskine>
- Fuglsang, N. (2021), *'The Politics of Economic Models'*, PhD-afhandling, Department of International Economics, Government and Business, Copenhagen Business School
- Jespersen, J. (red.), (1991), *Model og virkelighed – træk af debatten om de økonomiske modeller*, Jurist- og Økonomforbundets Forlag
- Jespersen, J. (2007), *Makroøkonomisk Metodologi – i et samfundsvidenskabeligt perspektiv*, Jurist- og Økonomforbundets Forlag
- Jespersen, J. (2012), 'Hvad ved vi om det "lange sigt"?' bidrag til *Festskrift til Niels Kærgård* redigeret af P. Andersen m. fl., DJØFs Forlag
- Jespersen, J. (2015), 'Når finansministeriets model får magten', *Økonomi og Politik*, 88/3, s. 41-50
- Jespersen, J. (2022), Samfundsøkonomiske modeller og prognoser: Hvad har økonomer bidraget med?, *Slagmark #84*, marts
- Kestler, A. (2013), Corydon: "Jeg tror på konkurrencestaten", *Politiken 24. august 2013*.
- Klein, L. (1980), *An Introduction to Econometric Forecasting and Forecasting Models*, New York: Lexiton Books
- Lucas, Robert E. (1976), "Econometric Policy Evaluation: A Critique," *Carnegie-Rochester Conference Series on Public Policy* 1, s. 19-46.

- Kærgård, N. (2020), 'Ellen Andersen 1. marts 1937 – 19. april 2018', *Nationaløkonomisk Tidsskrift*,
- Kærgård, N. (2020), *Finansministeriet og makroøkonomiske regnemodeller: et historisk rids*, Samfundsøkonomien, (2), s. 31-40
- Lucas, R., Jr. (1987), Models of Business Cycles, Oxford: Basil Blackwell
- Maas, H. (2014), *Economic Methodology: A historical introduction*, London: Routledge
- Rosted, J. (2021), *Genopretningspolitikken*, 1976-93, Odense: Historia
- Stahl, Rune M. (2018). *Economic theory, politics and the state in the neoliberal epoch*. Copenhagen: Department of Political Science, University of Copenhagen. Ph.D. Dissertation
- Ågerup, M. (2009), *Den retfærdige ulighed*, København, Cepos

Noter

1. (!) signalerer, at der med disse á priori-antagelser er påført modellen en egenskab, der ikke er empirisk bekræftet; men som alligevel lægges til grund for den anbefalede økonomiske politik.
2. I dag er DREAM placeret i en selvstændig administrativ enhed med en bestyrelse sammensat af repræsentanter fra universiteterne, finansministeriet og Nationalbanken dog primært finansieret af en bevilling på finansloven. Desuden er der indgået aftale med Det økonomiske Råd om brug af DREAM til mere langsigtede fremskrivninger af dansk økonomi.
3. Jeg tillader mig at sætte 'velfærdsreformer' i gáseøjne, idet der sandt at sige ikke var megen velfærd knyttet til de reformer, der primært sigtede mod at reducere de arbejdsmarkedrelaterede sociale ydelser, fjerne efterlønnen, hæve pensionsalderen og reducere og udjævne indkomstskatten. Det skete med den eneste sikre konsekvens, at uligheden ville øges. En konklusion, hvorom der herskede stort set konsensus, hvilket måske forklarer, hvorfor det ikke blev nævneværdigt diskuteret – eller måske snarere fordi, der langt ind i det dominerende økonommiljø var en opfattelse af, at der blot var tale om en mere 'retfærdig ulighed', som det blev formuleret af Cepos-direktøren Martin Ågerup, 2009
4. Det permanente årlige tab i BNP blev, ifølge Jesper Rangvid, CBS, på godt 200 mia. kr. <https://finans.dk/finans2/ECE9655598/hver-dansker-har-mistet-350000-kr-paa-finanskrisen-og-tabet-vokser-fortsat/?ctxref=ext>
5. Da Mette Frederiksens blev præsenteret for spørgsmålet om nødvendigheden af offentlige besparelser som en konsekvens af den øgede offentlige gældssætning i kølvandet COVID-19-krisen, svarede hun: »*Det er en lidt gammeldags måde at se det på,*«²
6. Den International Valutafond, IMF, <https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2016/12/31/Redistribution-Inequality-and-Growth-41291>, hvori det blev påvist, at en omfordelende politik ikke behøver at hæmme BNP-væksten, tværtimod.

² <https://www.information.dk/udland/leder/2021/01/mette-frederiksen-ret-staten-husholdning>

MAKRO-modellen

Peter Stephensen, Martin Bonde, Joao Ejarque, Grane Høegh og Anders Kronborg

8. december 2021

www.dreamgruppen.dk

1. Indledning

MAKRO er en stor empirisk funderet makroøkonomisk model, der skal kunne anvendes til flere formål. Den skal anvendes til fremskrivninger af dansk økonomi på mellemlang og lang sigt. Den skal anvendes til makroøkonomiske konsekvensvurderinger, herunder økonomiens kort- og mellemstidige tilpasning i forbindelse med økonomisk-politiske tiltag. Den skal anvendes til vurdering af den finanspolitiske holdbarhed. På sigt skal den kunne anvendes som ramme for kortsigtede konjunkturfremskrivninger.

MAKRO er udviklet af en modelgruppe i DREAM. Gruppen består af 4 økonomer og 3 studenter. MAKRO-gruppen har i skrivende stund arbejdet på udviklingen af modellen i 4 år.

MAKRO-modellen offentliggøres i første omgang som en betaversion i december 2021. Formålet med offentliggørelsen af en betaversion er at sikre åbenhed og transparens omkring modellen, herunder dens teoretiske og empiriske grundlag, samt sikre eksterne interesserens mulighed for at indgå i en faglig debat om modellen forud for ibrugtagningen af modellen i Finansministeriet.

Fokus for betaversionen af MAKRO er at give et indblik i modellens egenskaber ved politikeksperimenter og øvrige stød til økonomien. Betaversionens grundforløb er derfor stiliseret og tjener alene som et grundlag for at foretage marginaleksperimenter.

I forlængelse af offentliggørelsen af betaversionen følger en implementeringsperiode, hvor Finansministeriet, modelgruppen i Danmarks statistik, og MAKRO-gruppen i samarbejde vil gøre modellen klar til fuld ibrugtagning til fremskrivninger mv. I denne implementeringsperiode vil brugerne også have lejlighed til at opbygge praktisk erfaring med anvendelse af modellen.

I implementeringsperioden vil der frem mod fuld ibrugtagning ske tilretning og yderligere udvikling af modellen og dens empiriske grundlag. Dette videre arbejde vil ske på baggrund af de gjorte erfaringer med modellen samt eksternt input i forlængelse af offentliggørelsen af betaversionen. Betaversionen er således en foreløbig modelversion. Det grundlæggende modelkoncept er imidlertid afspejlet i betaversionen, som ligeledes giver et billede af, hvordan modellens marginalegenskaber på nuværende tidspunkt forventes at se ud, når den tages i brug.

Formålet med dette papir er at give en helt overordnet beskrivelse af betaversionen af MAKRO, herunder formålet med modellen, de centrale modelleringsvalg og den empiriske strategi. Der findes separate notater, som indeholder nærmere beskrivelser af henholdsvis de centrale modelleringsvalg (Høegh et al. 2021a) og den empiriske strategi bag modellen (Kronborg et al. 2021a). Nærværende notat kan ses som et ikke-teknisk resumé af disse. En udførlig teknisk beskrivelse af modelleringen af alle agenter og markeder kan findes i den tekniske dokumentation (Ejarque et al. 2021).

stød (Høegh et al. 2021b) samt belyser overensstemmelsen imellem kortsigtsegenskaberne af den samlede model og central empiri på dette punkt (DREAM, 2021). En række arbejdspapirer dokumenterer det øvrige empiriske grundlag for modellen (Kastrup & Kronborg, 2021, Kronborg, 2021, 2021a, Kronborg & Kastrup, 2021, Kronborg et al. 2019, 2021b, 2021c, 2021d, Kronborg & Stephensen, 2019).

Med henblik på at tydeliggøre konsekvenserne af den valgte modellering og empiriske strategi er de væsentligste forskelle i modelegenskaber mellem MAKRO og udvalgte makroøkonomiske modeller på danske data ved en række stød belyst i et særskilt papir (Finansministeriet, 2021c). Ligeledes er der foretaget en sammenligning af finanspolitiske multiplikatorer i betaversionen af MAKRO og en række internationale modeller (Røpke 2021).

Nærværende papir er opbygget som følger. I afsnit 1.1 gives en kort beskrivelse af formålet med MAKRO. I afsnit 1.2 beskrives de grundlæggende modelleringsvalg og der gives et overblik over modelleringen af forskellige hovedområder i MAKRO. Endelig gives i afsnit 1.3 en beskrivelse af den empiriske tilgang.

1.1 Kort om formål med MAKRO

MAKRO vil i Finansministeriet blive anvendt i forbindelse med

1. konjunkturvurderinger (primært som ramme, jf. nedenfor).
2. mellem- og langfristede fremskrivninger.
3. konsekvensvurderinger af økonomisk-politiske tiltag og eksogene stød.

MAKRO skal således anvendes til de samme overordnede opgaver, som makromodellen ADAM i dag anvendes til i Finansministeriet. Det er derfor vigtigt, at MAKRO fastholder centrale styrker ved ADAM, såsom detaljeringsgraden af beskrivelsen af de offentlige finanser samt en omfattende databank af høj kvalitet.

Herudover har selve formålet med udviklingen af MAKRO været at bringe modellen på omgangshøjde med nyere empiriske metoder og resultater fra den økonomiske litteratur. Dette har krævet en ny tilgang til den empiriske fundering af modellen samt til

modellering af forventningsdannelse og adfærd. Gevinsterne herved vil primært være relateret til punkt 2 og 3 ovenfor.

De modelegenskaber, der er i fokus i forbindelse med udviklingen af MAKRO, er særligt økonomiens samlede tilpasning til dens strukturelle niveauer på kort- og mellemfristet sigt, herunder ved forskellige typer af stød til økonomien, fx stigninger i arbejdsudbudet. Hertil kommer modellering af virksomhedernes investeringsbeslutninger og husholdningernes forbrugs- og opsparingsbeslutninger på baggrund af nyere økonomisk forskning mv.

På den baggrund har der i udviklingen af MAKRO været lagt vægt på en række forhold.

For det første, at modellen er baseret på en bred empirisk tilgang. I økonometriske makromodeller er de enkelte ligninger empirisk funderet på baggrund af tidsserieestimeringer. Det empiriske fundament for MAKRO er bredere end dette, således at forskellige typer af empiriske tilgange og resultater i højere grad supplerer hinanden. Herunder er modellens beskrivelse af økonomiens kortsigtstilpasning til forskellige typer af stød direkte baseret på empirisk analyse af den relevante tilpasningstid.

Det andet forhold der lægges vægt på er, at modellen tager højde for betydningen af forventninger. MAKRO er bygget til at kunne håndtere, at konsekvenserne af et stød til økonomien på kort sigt kan afhænge af, om stødet er permanent eller midlertidigt. Eksempelvis vil en regelændring der medfører en indkomstfremgang hos husholdningerne i et givet år kunne slå forskelligt igennem på forbruget, alt efter om indkomstfremgangen bortfalder igen året efter eller er varig. Dette kræver, at virksomheder og husholdninger i et vist omfang har fremadskuende forventninger.

For det tredje, at modelleringen af adfærd som udgangspunkt er teoretisk velfunderet, således at virksomhedernes og husholdningernes økonomiske adfærd afspejler deres bestræbelser på at opnå det for dem bedst mulige resultat givet de rammevilkår og eventuelle begrænsninger (fx kreditbegrænsninger, tilpasningsomkostninger mv.), de står over for.

For det fjerde, at der gives en eksplisit beskrivelse af husholdningernes forbrugs- og opsparingsbeslutninger over livsforløbet, bl.a. for at tage højde for demografiens betydning for forbrug og opsparing.

Endelig for det femte er det vigtigt, at modellen indeholder gab og strukturelle niveauer for bl.a. beskæftigelse og output. Dette bidrager til, at modellen eksplisit kan beskrive økonomiens tilpasning til det (evt. nye) strukturelle niveau efter et stød.

Et særligt fokus for anvendelsen af modellen i Finansministeriet er udviklingen i de offentlige finanser, herunder den finanspolitiske holdbarhedsindikator, og der er således

endvidere lagt vægt på en detaljeret beskrivelse af de offentlige finanser og øvrige forhold, som er væsentlige for den finanspolitiske holdbarhed. Dette omfatter de strukturelle faktorer, som har betydning for økonomiens udvikling: den aldrende befolkning, tilbagetrækning fra arbejdsmarkedet, udviklingen i uddannelsesniveauet, olie og gas i Nordsøen mv. Til MAKRO benyttes fortsat bl.a. Befolkningsregnskabet fra Finansministeriet og Pensionsmodellen fra DREAM til at fremskrive disse forhold.¹

MAKRO er en relativt stor makromodel. Med henblik på ikke at gøre opbygningen af MAKRO mere kompliceret end højst nødvendigt, har det derfor fra starten været væsentligt at holde sig for øje, hvad der ikke er en del af formålet med udviklingen af en sådan model. Det drejer sig blandt andet om, hvilke elementer i fremskrivningerne og konsekvensvurderingerne, der (fortsat) vil blive udarbejdet uden for MAKRO-modellen, fx ved anvendelse af andre – og til formålet mere velegnede – modeller.

I udviklingen af MAKRO er der derfor foretaget en række fravælg.

For det første er det ikke et mål, at MAKRO grundlæggende skal kunne mere eller anvendes anderledes i forbindelse med konjunkturvurderingerne, end det er tilfældet med ADAM i dag. Udviklingen af MAKRO er eksempelvis ikke sket med det formål, at konjunkturprognoserne i stort omfang skal baseres på modelgenererede kortsigtsfremskrivninger. MAKRO vil således primært udgøre en konsistent ramme for udarbejdelse af konjunkturprognosene. Selve skønnet over udviklingen i dansk økonomi for prognoseårene – både detaljeret og aggregeret set – vil fortsat i vid udstrækning blive foretaget eksplisit af modelbrugerne på baggrund af en lang række indikatorer og skøn foretaget uden for modellen.

For det andet er det ikke et mål, at MAKRO skal anvendes til fremskrivning af befolkningsudviklingen og strukturel beskæftigelse. Fremskrivningen af demografi og strukturel beskæftigelse i grundforløbet for de mellemfristede fremskrivninger vil fortsat ske i en separat (for)model, som er udviklet til netop denne opgave (Finansministeriets befolkningsregnskab). Denne fremskrivning vil – som det også er tilfældet med ADAM i dag – udgøre et eksogent input til MAKRO.

Endelig for det tredje er det ikke et mål, at MAKRO skal kunne anvendes til at vurdere strukturelle beskæftigelsesvirkninger af specifikke regelændringer/konkrete økonomisk-politiske tiltag. Dette er heller ikke tilfældet med andre danske makroøkonomiske modeller. Det skyldes, at fastlæggelsen af strukturelle beskæftigelsesvirkninger kræver, at der kan tages detaljeret højde for karakteren og virkningerne af det økonomisk-

¹ I forbindelse med beta-versionen indgår et stiliseret og foreløbigt grundforløb, der udelukkende tjener som grundlag for marginalstød for at illustrere modellens egenskaber. Der er ikke tale om en egentlig fremskrivning af dansk økonomi ved gældende eller vedtagen politik, og der er bl.a. ikke taget højde for corona-krisen.

politiske tiltag, herunder inddrages informationer på et detaljeringsniveau, som ikke kan inddbygges direkte i en makroøkonomisk model. Strukturelle beskæftigelsesvirkninger af økonomisk-politiske tiltag vil som udgangspunkt fortsat blive opgjort uden for MAKRO og lagt ind via ovennævnte separate formodel.

Overordnet set vil selve overgangen til MAKRO ikke repræsentere en grundlæggende omvälvning af de fremskrivninger og konsekvensvurderinger, der udarbejdes i Finansministeriet. Det skal snarere ses som et led i den udvikling af metoder og vurderinger, som løbende finder sted.

På trods af de forbedringer, som søges opnået, vil der heller ikke med MAKRO være tale om en perfekt eller entydigt rigtig model. En sådan model eksisterer ikke. Også med MAKRO vil modelbrugeren have det endelige ansvar for de beregninger, der foretages, herunder de forudsætninger, der lægges til grund. Ligeledes vil der – som det også har været tilfældet med ADAM og andre store modeller – være behov for løbende at udvikle modellen, efterse egenskaber mv.

1.2 Overordnet om centrale modelleringsvalg og modellens opbygning

Fremadskuende forventninger og teoretisk konsistent modellering udgør de helt grundlæggende modelleringsvalg i MAKRO. Derfor uddybes disse først, hvorefter der gives et overordnet overblik over modelleringen af forskellige områder af økonomien i betaver-sionen af MAKRO.

Fremadskuende forventninger

Med fremadskuende forventninger menes, at virksomheder og husholdninger inddrager information om kommende ændringer i deres rammevilkår og økonomien mere generelt i deres forventningsdannelse (i det omfang sådanne ændringer er kendt på forhånd). Det kan være ændringer i forhold såsom skattesystemet, folkepensionsalderen, finanspolitikken i øvrigt, uddannelsesniveauet, demografien mv. Fremadskuenhed i forventningsdannelsen er blandt andet afgørende for, at modellen kan give et bud på forskellen i virkningerne af midlertidige og permanente stød til økonomien.

Fremadskuende forventninger står i modsætning til såkaldte statiske eller adaptive forventninger, hvor agenterne alene danner deres forventninger til fremtiden på baggrund af den hidtil observerede udvikling i rammevilkår og økonomien i øvrigt (bagudskuende forventninger). Rent empirisk er der en række tegn på relevansen af fremadskuenhed i forventningsdannelsen, herunder at husholdninger reagerer på forhånd på kendte ind-komstændringer, at virksomhedernes ansættelses- og investeringsbeslutninger afhænger af forventninger til fremtidige rammevilkår samt at aktivpriser reagerer på nyheder

om kommende ændringer i det effektive afkast.² Fremadskuenhed i forventningsdannelsen vurderes at være et centralt element i forbindelse med modelleringen af en økonomi, som udsættes for en række forventede stød og undergår strukturelle udviklinger – eksempelvis den demografiske udvikling, indeksering af pensionsalderen mv.

I MAKRO er *afsættet* for modelleringen af virksomhedernes og forbrugernes fremadskuenende forventninger, at disse dannes modelkonsistent (eller rationelt). Dvs. at agenternes forventninger er i overensstemmelse med modellens forudsigelser, og at der i forhold hertil ikke begås systematiske forventningsfejl. Denne type af fremadskuenende forventninger er meget udbredt i moderne makroøkonomiske modeller. Der er imidlertid en række årsager til, at en fuldt modelkonsistent forventningsdannelse i sin rene form ikke er realistisk. Agenternes forventningsdannelse kan fx være underlagt forskellige typer af friktioner i forhold til tilegnelsen af præcis information om aktuelle eller fremtidige økonomiske forhold, eller agenterne kan være begrænset i deres evne til at danne modelkonsistente forventninger på baggrund af denne information. Modeller med fuldt modelkonsistent forventningsdannelse har da typisk også udfordringer med at matche empiriske resultater vedrørende økonomiens tilpasning til stød.

Derfor indeholder modelleringen af adfærdens i betaversionen af MAKRO elementer, som (eksplicit eller implicit) kan bidrage til, at agenternes adfærd afspejler mindre end fuld modelkonsistent forventningsdannelse. Det drejer sig eksempelvis om, at den aktuelle indkomst og formue har direkte indvirkning på husholdningernes forbrugsbeslutning (blandt andet via en andel såkaldte hånd-til-mund forbrugere) samt at den aktuelle indtjening har direkte betydning for virksomhedernes investeringsbeslutning (via en såkaldt finansiell accelerator). Den kvantitative betydning af disse elementer fastlægges som en del af matchingen af modellen til empiriske resultater for økonomiens samlede tilpasning til stød, jf. også Kronborg et al. (2021a).

Behandlingen af forventninger i MAKRO vil indgå i den videre udvikling af modellen og dens empiriske grundlag, herunder for at afdække et eventuelt behov for i højere grad at inddarbejde eller tage højde for mindre end fuldt modelkonsistente forventninger.³

² Der findes en lang række af danske og udenlandske studier, der indikerer at fremadskuenende forventninger er relevante. For Danmark drejer det sig blandt andet om ændringer i renteudgifter (Druedahl et al., 2021) og kapitaliseringseffekter af boligbeskatning (Høj & Schou, 2018). På amerikansk data finder studier bl.a., at annoncerede skatteændringer påvirker investeringer m.v. (Mertens & Ravn, 2012) og aktiekurser (Lang & Shackelford, 2000) efter de bliver vedtaget, men før de er implementeret.

³ Modelbrugeren kan vælge at foretage stød til betaversionen af MAKRO på en måde, som kan afspejle mindre end fuld information eller opmærksomhed omkring stødets karakter (fx dets varighed).

Teoretisk konsistens

I MAKRO tilstræbes det som nævnt ovenfor, at modelleringen af adfærd som udgangspunkt er teoretisk velfunderet, således at virksomhedernes og husholdningernes økonomiske adfærd udledes af eksplisitte optimeringsproblemer. Det vil sige, at adfærden afspejler deres bestræbelser på at opnå det for dem bedst mulige resultat givet de rammevilkår og eventuelle begrænsninger (fx kreditbegrænsninger, tilpasningsomkostninger mv.), som de står over for.

Af og til fremføres i dele af litteraturen om empiriske makromodeller, at der kan være et generelt trade-off mellem den vægt, der i modelbygningen lægges på teori henholdsvis empiri (jf. fx Wren-Lewis, 2019). I Pagan (2003) er præsenteret en afvejning mellem hensynet til empirisk grundlag og teoretisk stringent og konsistent modellering.

Seriøse policy-modeller vægter empiri højt og forsøger samtidig i videst muligt omfang at være teoretiske stringente. Vægten på de to kan dog variere. Nogle nyere makroøkonometriske modeller (fx såkaldte DSGE-modeller) har et meget stort fokus på teoretisk konsistens, mens traditionelle modeller (såkaldte makroøkonometriske modeller) har primært fokus på den historiske forklaringsgrad (fx høj R2) af de enkelte ligninger i modellen, der beskriver husholdningernes og virksomhedernes adfærd.

Et stort fokus på en høj forklaringsgrad for historiske data kan trække i retning af mange enkeltevis estimationer af forskellige adfærdsligninger, som ikke direkte udledes fra eksplisitte optimeringsproblemer men specificeres mere frit, hvorefter disse adfærdsligninger kombineres til en samlet model.

En væsentlig udfordring for denne tilgang er imidlertid, at mens forventningsdannelsen hos husholdninger og virksomheder er central for økonomiens funktionsmåde og reaktion på stød, så eksisterer der ikke gode data for husholdningernes og virksomhedernes forventninger. Disse er i det store og hele ikke observerbare. Da det ikke med rimelighed kan lægges til grund, at husholdningernes og forventninger historisk set og i alle tilfælde har været nøjagtige (det ville svare til at de aldrig er blevet overraskede over udviklingen i økonomien⁴), ”løses” problemet med mangel på gode data for forventninger typisk med en antagelse om bagudskuende (statisk eller adaptiv) forventningsdannelse, som den makroøkonometriske model efterfølgende arver. Som nævnt ovenfor vurderes fuldstændig fravær af fremadskuenhed imidlertid heller ikke at være en rimelig beskrivelse af forventningsdannelsen, og en tilsyneladende høj historisk forklaringsgrad i en adfærdsligning estimeret med denne forudsætning kan derfor være misvisende.

⁴ Eksempelvis indikerede spørgeskemabaserede forbruger- og erhvervstillsindikatorer, at forventningerne til den makroøkonometriske udvikling var alt for positiv inden finanskrisen i 2008

Derudover er tilgangen (dvs. uden adfærdsaligninger udledt fra eksplisitte optimeringsproblemer) ikke robust over for den såkaldte Lucas-kritik, dvs. at adfærdens kan ændre sig ved stød, der ligger uden for det historiske erfarsområde, eller ved strukturelle ændringer af fx politik-regimet.⁵

Omvendt kan et stort fokus på teoretisk konsistens i adfærdsmodelleringen kræve udstrakte forsimplinger, som risikerer at ende i forklaringer på historiske udviklinger, der i nogle tilfælde kan fremstå konstruerede eller over-forsimplede, hvilket der er eksempler på i DSGE-litteraturen. Dertil kommer, at empiriske metoder, som er fuldt konsistente med en helt stringent teoretisk model med fremadskuende forventninger, og som derfor tager udgangspunkt i modellens egenskaber som samlet system, i praksis ikke kan inddrage tilstrækkelig information til at sikre pålidelig estimation af alle modellens adfærdsparametre, jf. også nedenfor.

Selvom det teoretiske fundament for MAKRO er mere stringent og konsistent end traditionelle makroøkonomiske modeller, vægtes hensynet til empirien derfor højere end fuldkommen teoretisk konsistens. I den forbindelse lægges også vægt på en bred empirisk tilgang, som omfatter – men ikke begrænses til – at modellens egenskaber som samlet system holdes op imod empiri for økonomiens tilpasning til relevante stød.

Overordnet om modellens opbygning

Som nævnt ovenfor, har der i udviklingen af MAKRO været fokus på både dens kortsigtede og langsigtede egenskaber. Derfor er MAKRO opbygget som en langsigtet strukturmodel, der er udvidet med reale og nominelle friktioner, som primært har betydning for de kortsigtede egenskaber.

De langsigtede strukturelle egenskaber i MAKRO bygger på (mange) repræsentative agenter – de vigtigste er husholdninger fordelt på alder og virksomheder fordelt på brancher – med mikrofunderet adfærd samt et element af fremadskuende forventninger. Tilsvarende er kortsigtstræghederne så vidt muligt mikrofunderede og baseret på agenternes fremadskuende forventninger.

Da MAKRO er en omfattende model med megen institutionel detalje, er det valgt at se bort fra eksplisit modellering af usikkerhed (stokastik), som det ellers kendes fra nogle moderne makroøkonomiske modeller (DSGE-modeller). Af hensyn til modellens overensstemmelse med det empiriske grundlag, er dette imidlertid indarbejdet ved specifikke elementer som f.eks. risikopræmier i virksomhederne og et element af

⁵ En tilgang med enkeltvise estimationer af frit specificerede adfærdsaligninger rummer endvidere kendte risici for ad hoc tilpasninger, evt. overfitting og ustabile relationer. Erfaringerne både i Danmark og internationale har ikke sjældent været, at relationer, der oprindeligt virkede godt vurderet på historisk forklaringsgrad, efterfølgende er brutt ned og har måttet justeres, genestimeres eller helt ændres. Dette er en almindelig del af arbejdet med traditionelle makroøkonometriske modeller.

forsigtighedsopsparing i husholdninger, som kan medvirke til at husholdninger og virksomheders adfærd alligevel i et vist omfang afspejler usikkerhed om fremtiden (der kan være en vigtig faktor).

Adfærds modelleringen på hovedområderne i MAKRO kan kortfattet opsummeres som følger:

- Husholdningerne er modelleret som separate generationer (en såkaldt OLG-model), som hver især deltager på arbejdsmarkedet i overensstemmelse med fremskrivningen i finansministeriets befolkningsregnskab, arver, sparar op i både pensionsordninger og fri opsparing samt forbruger henover livsforløbet. Husholdningerne fordeler deres forbrug på boligforbrug samt 4 andre forbrugsgrupper, hvoraf boligformuen i vidt omfang gældsfansnieres. Forbrugsniveauet fastlægges delvist på baggrund af den aktuelle indkomst (via en andel hånd-til-mund husholdninger) og delvist på baggrund af den forventede fremtidige indkomst mv. (via en andel fremadskuende husholdninger). Centrale elementer i modelleringen af forbruget rettet mod modellens kortsigtsegenskaber omfatter vanedannelse, et element af forsigtighedsopsparing (via direkte nytte af formue), en træghed i belåning af friværdi samt nedvægtning af forventede boligprisændringer i valget af boligforbrug.
- Udbuddet af boliger er givet på helt kort sigt, og tilpasses via investeringer i nye boliger, som består af en kombination af land og boligkapital. Centrale elementer i modellering af boligudbuddet rettet mod modellens kortsigtsegenskaber omfatter at investeringer i boliger er underlagt kvadratiske installationsomkostninger samt at en fast andel af den samlede landmængde svarende til afskrivningsraten på boligkapitalen "friges" til nybyggeri (investeringer i boliger) hver periode. Boligprisen fastlægges således at efterspørgslen og udbuddet på boligmarkedet mødes, og kortsigtsegenskaberne af boligpriserne påvirkes dermed af træghederne på begge sider af dette marked.
- Produktionen er opdelt på otte private brancher og en offentlig branche. Virksomhederne i de private brancher producerer ved hjælp af arbejdskraft, bygnings- og maskinkapital samt materialeinput således at omkostningerne minimeres. Kapital opbygges via investeringer og prisen på kapitalinput udgøres af en såkaldt usercost, som repræsenterer omkostningerne ved at anvende yderligere kapital i produktionen på et givent tidspunkt (herunder finansiering mv.). Outputprisen fastsættes mhp. profitmaksimering under monopolistisk konkurrence. Centrale elementer i modelleringen af virksomhederne rettet mod modellens kortsigtsegenskaber omfatter for det første, at ændringer i stigningstakten af outputprisen er forbundet med tilpasningsomkostninger. For det andet, at justeringer af antallet af ansættelser eller omfanget af investeringer er forbundet med (kvadratiske) tilpasningsomkostninger. For det tredje

er virksomhedernes afkastkrav ifm. investeringer påvirket af den aktuelle indtjening (en såkaldt finansiel accelerator). For det fjerde kan virksomhederne midlertidigt justere produktionen ved at variere kapacitetsudnyttelsen af kapital og arbejdskraft.

- Arbejdsmarkedet er modelleret som en søgermodel, hvor jobsøgning kræver en indsats fra arbejdstagernes (husholdningernes) side og besættelse af ledige stillinger er forbundet med en omkostning for arbejdsgiverne (virksomhederne). Søgefriktionerne betyder at det tager tid at finde job og besætte stillinger, hvorfor der til et hvert givent tidspunkt vil være ufrivillig ledighed og ubesatte stillinger. Lønninger forhandles mellem repræsentanter for henholdsvis arbejdstagerne og arbejdsgiverne, hvor det forventede overskud ved de berørte ansættelsesforhold deles. Centrale elementer i modelleringen af arbejdsmarkedet rettet mod modellens kortsigtsegenskaber omfatter – ud over de nævnte søgerfiktioner og tilpasningsomkostningerne ved justering af antallet af ansættelser på virksomhedssiden (jf. foregående punkt) – for det første, at det kun er en andel af alle ansættelseskontrakter, der genforhandles hver periode, hvilket giver anledning til træghed i (gennemsnits)lønnen samt fremadskuenhed i løndannelsen.
- Udlandet spiller for det første en rolle på dels produktmarkederne via udenrigshandlen dels kapitalmarkederne. Modelleringen tager udgangspunkt i Danmark som en lille åben økonomi, der ikke påvirker udlandet. Eksportkonkurrerende priser, importpriser, det udenlandske renteniveau og eksportmarkedets størrelse tages således for givet. Eksport- og importefterspørgslen antages at have eksogent give priselasticiteter (en såkaldt armingtonmodellering), dog påvirkes eksportefterspørgslen af den strukturelle produktionskapacitet (i betaversionen konkret strukturel beskæftigelse) i den indenlandske private sektor (en såkaldt skalaeffekt i eksporten). Centrale elementer i modelleringen af udenrigshandlen rettet mod modellens kortsigtsegenskaber omfatter en træghed i skalaeffekten i eksporten samt trægheder på gennemslaget af ændringer i størrelsen af eksportmarkedet samt træghed i gennemslaget af de relative priser. Dette betyder eksempelvis at de effektive eksportpriselasticiteter er lavere på kort sigt end på langt sigt. På kapitalmarkederne er udlandet den marginale investor, der stiller kapital og investeringsmuligheder til rådighed for henholdsvis virksomheder og husholdninger til et givet afkastkrav (knyttet til renteniveauet).

Modelleringen af de forskellige områder er nærmere beskrevet i Høegh et al. (2021a) og Ejarque et al. (2021).

1.3 Empirisk tilgang

Den empiriske strategi for MAKRO har til formål at sikre, at både kortsigtsegenskaberne for modellen som samlet system (især tilpasningshastigheder) og langsigtsegenskaberne er empirisk velunderbyggede. Som det fremgår af ovenstående, er det et centralt

hensyn, at de anvendte estimationsmetoder i praksis kan anvendes til solid empirisk fundering af en omfattende model med et udbygget teoretisk fundament og fremadskuende adfærd.

Med MAKRO tilstræbes derfor bedst mulig overensstemmelse med tilgængelig empiri ud fra en bred tilgang. I formuleringen "bred tilgang" ligger, at der både trækkes på anerkendte resultater fra litteraturen, egne særskilte estimationer af bestemte adfærdsparametre og en empirisk fundering af de samlede modelegenskaber vedrørende økonomiens tilpasningstid ved centrale stød. Der er således hverken tale om, at hele MAKRO estimeres som et samlet simultant system, eller at modellen er opstillet som en kombination af enkeltvist estimerede adfærdsligninger.

En bred empirisk tilgang

Modellens kortsigtsegenskaber sikres ved at fastsætte modellens kortsigts- eller træghedsparametre, så økonomiens reaktioner på en række stød i MAKRO bedst muligt matcher tilsvarende empirisk estimerede reaktioner.⁶ De estimerede reaktioner stammer fra empiriske modeller (primært såkaldte SVAR-modeller), der i høj grad er ateoretiske og derfor tillader data at tale relativt frit i forhold til økonomiens reaktion på og tilpasning efter stød.

Denne tilgang til empirisk fundering af MAKRO's samlede kortsigtsegenskaber vurderes at have en række fordele fsva. transparens og fortolkning. Den tydelige skillelinje mellem modellen (MAKRO) og de empiriske resultater (primært fra SVAR-modeller), der matches til, gør valget af den centrale empiri transparent. Metoden tillader endvidere umiddelbar visuel inspektion af, hvor godt matchingen af de empiriske resultater lykkes – herunder på hvilke områder den lykkes godt, og på hvilke den lykkes mindre godt. Endeligt kan man fortolke de empiriske resultater og modellen både i sammenhæng og hver for sig med henblik på at vurdere om afvigelser primært kan henføres til udfordringer i estimationen eller modelleringen.

Konkret er økonomiens reaktioner estimeret for efterspørgselsstød (offentligt forbrug og udenlandsk efterspørgsel), udbudsstød (arbejdsudbud) samt stød, der indeholder både udbuds- og efterspørgselselementer (oliepris og den pengepolitiske rente i euroområdet), jf. også MAKRO (2021). Modellens kortsigtsparametre tilpasses, således at MAKRO har tilpasningsegenskaber, der svarer så godt som muligt til de empirisk estimerede reaktioner for dansk økonomi ved disse stød, jf. også MAKRO (2021). Modellens egenskaber verificeres herudover ved at holde dem op mod eksisterende mikroøkonometrisk litteratur om blandt andet kortsigtede marginale forbrugstilbøjeligheder ved

⁶ Dette kaldes også for impuls-respons matching og er en velkendt metode i DSGE-litteraturen, hvor den bruges som alternativ til såkaldt bayesiansk maximum likelihood estimation.

midlertidige indkomststød mv. (Kronborg et al. 2021) samt såkaldte fortrængningsefekter ved tvungen pension (Høegh, 2021).

Modellens langsigtssegenskaber sikres blandt andet via særskilte estimationer af en lang række substitutionselasticiteter i virksomhedernes og husholdningernes efterspørgsels-systemer samt udenrigshandelselasticiteter og den ovennævnte skalaeffekt i eksporten (Kastrup & Kronborg 2021). Ud over resultaterne fra disse konkrete estimationer indrages generelt også relevante resultater og indsigt fra den eksisterende empiriske litteratur i fastlæggelsen af modellens parametre. Når substitutionselasticiteterne er fastlagt, tages de for givet i den empiriske fundering af kortsigtssegenskaberne omtalt ovenfor.

Afvejninger i den empiriske strategi for MAKRO

Den empiriske tilgang i MAKRO står i modsætning til traditionelle makroøkonometriske modeller, hvor langsigtselfasticiteter og tilpasningshastigheder ofte estimeres ved én samlet estimation. Det er vurderingen, at kombinationen af forskellige estimationsmetoder i MAKRO er hensigtsmæssig af flere årsager. For det første er det et kendt problem, at kvantitet og kvalitet af data kan gøre det vanskeligt at bestemme både langsigtselfasticiteter og tilpasningshastigheder i samme estimation. Løsningen er relativt ofte at fastsætte enten tilpasningshastigheden eller en langsigtselfasticitet eksogen, hvilket ikke nødvendigvis er tilfredsstillende i en model, der skal have empirisk velunderbyggede egenskaber på kort og langt sigt. For det andet vil kombinationen af enkeltvist estimerede adfærdsligninger ikke nødvendigvis give anledning til en retvisende beskrivelse af økonomiens samlede egenskaber (herunder pga. simultanitetsbias), og der har med denne tilgang sjældent været et sideløbende, udbygget grundlag for at evaluere de samlede modelegenskaber. For det tredje, er der et element af fremadskuende forventninger i mange af MAKRO's adfærdsligninger, som er et vigtigt element i moderne makroøkonomi. Det er som nævnt ovenfor vanskeligt at estimere sådanne relationer baseret på klassisk enkeltligningestimation.

Den empiriske tilgang i MAKRO er ligeledes bredere end i typiske DSGE-modeller, hvor der er overvejende fokus på at estimere modellen i én samlet systemestimation. Sådan systemestimation kan håndtere fremadskuenhed i forventningsdannelsen – i det mindste rent teknisk – og modvirke eventuel forurening af parameterestimater fra simultanitetsbias eller udeladte variable, som kan være et problem for modeller baseret på enkeltligningestimationer. Systemestimation er dog typisk ikke tilstrækkelig informativ til at fastlægge alle parametre i (selv forholdsvis små) modeller, hvorfor tilgangen under alle omstændigheder ofte må suppleres med anden information (fx via såkaldte priors) – og er afhængig af kvaliteten heraf. Herunder kan det i estimationen være nødvendigt

at lægge nogle relativt strenge begrænsninger på værdierne af nogle af parametrene (via såkaldte priors), hvilket gør estimationen af disse mindre datadrevet.

På tilsvarende vis er der også både fordele og ulemper ved den valgte tilgang til empirisk fundering af MAKRO. De relationer og parametre, der estimeres via enkeltligningsestimation, kan vise sig utsat for nogle af de ovennævnte problemer, eksempelvis med tendens til relationer, der viser sig at være ustabile over tid. De parametre, der fastlægges på baggrund den generelle litteratur, kan risikere at være påvirket af modelbyggerens subjektive holdninger og kvaliteten af de metoder, der er anvendt i de pågældende studier. Endeligt vil parametre, der fastsættes ud fra matching af modellens kortsigtsegen-skaber til empiriske resultater for økonomiens tilpasning til stød (hvilket kan betragtes som en form for delvis systemestimation) naturligt være utsat for den usikkerhed, der knytter sig til estimationerne heraf.⁷ Endvidere kan den historiske forklaringskraft af de relationer, der indeholder parametre fastlagt på denne måde, typisk ikke evalueres en-keltvist på samme måde som det kan lade sig gøre med de enkeltvist estimerede ad-færdsrelationer i traditionelle makroøkonometriske modeller.

Samlet gælder, at mens der i MAKRO generelt lægges et eksplisit teorigrundlag til grund – baseret på, hvad der anses for relativt bredt accepteret og gængs teori – så accepteres et trade-off mellem forskellige empiriske ønsker eller muligheder: Der accepteres således på visse stræk en mindre gennemskuelig historisk dataforklaring for en række centrale adfærdsrelationer til gengæld for en tilgang, som kan håndtere eksplisit adfærds-modellering med et element af fremadskuenhed i forventningerne og i højere grad er baseret på en direkte empirisk fundering af modellens tilpasningsegenskaber ved en række efterspørgsels- og udbudsstød. Hertil kommer en bred læsning af den økonomiske litteratur og empiri (fx i fastsættelsen af dybe parametre og modellering af adfærds-mekanismer).

På den baggrund er det en vigtig del af modeludviklingen, at de foretagne valg og konsekvenserne heraf lægges åbent frem og begrundes, hvilket offentliggørelsen af beta-versionen af MAKRO er et centralt element i.

2. Referencer

- J. Druedahl, E. B. Jensen and S. Leth-Petersen. The Intertemporal Marginal Propensity to Consume out of Future Persistent Cash-Flows: Evidence from Transaction Data. Arbejdspapir, 2021.
- J. Ejnarque, M. Bonde, G. Høegh, A. Kronborg, and P. Stephensen. MAKRO model documentation. DREAM-arbejdspapir, 2021.

⁷ Denne usikkerhed knytter sig – lige som det er tilfældet ved andre estimationer – bl.a. til den generelle vari-ans i data, valget af variable (og indflydelsen af udeladte variable), der indgår i de anvendte SVAR, de antagel-ser, der ligger til grund for identifikationen og fortolkningen af forskellige stød til økonomien, samt de ændrin-ger i resultater, der kan fremkomme, når nye data bliver tilgængelige.

- Finansministeriet. Marginal forbrugstilbøjelighed ud af ændringer i boligformue i betaversionen af MAKRO. Finansministeriet notat, 2021a.
- Finansministeriet. Husholdningernes rentefølsomhed i betaversionen af MAKRO. Finansministeriet notat, 2021b.
- Finansministeriet. Sammenligning af stiliserede stød i betaversionen af MAKRO, ADAM og Nationalbankens DSGE-model. Finansministeriet notat, 2021c.
- G. Høegh. Analyse af stød til pensionsindbetalinger - fortrængningseffekter i MAKRO. DREAM-arbejdspapir, 2021.
- G. Høegh, M. Bonde, J. Ejarque, A. Kronborg, and P. Stephensen. MAKRO: Oversigt over model og centrale modelleringsvalg. DREAM-arbejdspapir, 2021.
- G. Høegh, J. Røpke, A. Kronborg, and P. Stephensen. Grundlæggende stødanalyser i MAKRO. DREAM-arbejdspapir, 2021.
- A. K. Høj, M. R. Jørgensen and P. Schou. Lax tax changes and full capitalisation. *Fiscal Studies*, 2018.
- C. Kastrup and A. Kronborg. Udbudseffekter i dansk eksport? DREAM-arbejdspapir, 2021.
- A. Kronborg. Estimating foreign shocks in a VAR model. DREAM-arbejdspapir, 2021a.
- A. Kronborg. Estimating government spending shocks in a VAR model. DREAM-arbejdspapir, 2021b.
- A. Kronborg, G. Høegh, P. Stephensen, M. Bonde, and J. Ejarque. Det empiriske grundland for MAKRO. DREAM-arbejdspapir, 2021a.
- A. Kronborg and C. Kastrup. Estimering af forbrugssystemet i MAKRO. DREAM-arbejdspapir, 2021.
- A. Kronborg, C. Kastrup, and P. Stephensen. Estimating the Constant Elasticity of Substitution when Technical Change is Time-Varying: A Kalman Filtering Approach. DREAM-arbejdspapir, 2019.
- A. Kronborg and K. Poulsen. Estimater for elasticiteterne i MAKROs produktionsfunktion. DREAM-arbejdspapir, 2021b.
- A. Kronborg, K. Poulsen, and C. Kastrup. Estimering af udenrigshandelselasticiteter i MAKRO. DREAM-arbejdspapir, 2020c.
- A. Kronborg, J. Røpke, M. Bonde, and P. Stephensen. Marginal propensity to consume. DREAM-arbejdspapir, 2021d.
- A. Kronborg and P. Stephensen. Decomposing the Ins and Outs of Unemployment: Cyclical, Structural, and Demographic Trends in the Danish Labor Market. DREAM-arbejdspapir, 2019.
- M. H. Lang and D. Shactman. Capitalization of capital gains taxes: evidence from stock price reactions to the 1997 rate reduction. *Journal of Public Economics*, 2000.
- MAKRO. Matching af impuls responser og øvrige kortsigtsmomenter: MAKRO ift. empirien. DREAM-arbejdspapir, 2021.
- K. Mertens and M. O. Ravn. Empirical evidence of the aggregate effects of anticipated and unanticipated us tax policy shocks. *American Economic Journal: Economic Policy*, 2012.
- A. Pagan. Report on modelling and forecasting at the Bank of England. *Bank of England Quarterly Bulletin*, 2003 (1-29).
- J. Røpke, G. Høegh, and A. Kronborg. Finanspolitiske multiplikatorer i MAKRO. DREAM-arbejdspapir, 2021.
- P. Stephensen, M. Bonde, G. Høegh, J. Ejarque, and A. Kronborg. MAKRO-modellen. DREAMarbejdspapir, 2021.
- S. Wren-Lewis. Ending the microfoundations hegemony. *Oxford Review of Economic Policy*, 2018.

*Niels Kærgård
Professor emeritus
Institut for Fødevare- og Ressourceøkonomi
Københavns Universitet
Rolighedsvej 23
1958 Frederiksberg C
E-mail: nik@ifro.ku.dk*

Data og statistiske metoder ved økonomiske modelberegning: Udviklingen fra 1970'erne til i dag

1. Indledning

Dansk økonomisk modelbygning startede for alvor i slutningen af 1960'erne og blev synlig for bredere kredse med offentliggørelsen af Ellen Andersens arbejde med det, der blev til ADAM (Annual Danish Aggregate Model) i 1972, af Det Økonomiske Råds modelarbejde med SMEC I (Simulation Model of the Economic Council) og SMEC II i 1973 og 1974 og Nationalbankens NATAN i 1975, se Andersen (1972), Hansen og Paldam (1973), Rosted, Schaumann & Sørensen (1974 og 1974a) og Blomgren-Hansen (1975).

Det betyder ikke, at økonometrien ikke allerede på det tidspunkt havde en lang historie internationalt og i Danmark. Men i Danmark var det kun spredte tilløb. Den første, der kvantificerede økonomiske sammenhænge med brug af matematisk-statistiske metoder var faktisk danskeren Edvard Ph. Mackeprang, der i 1906 forsvarede en disputats med en omhyggelig estimation af efterspørgselskurver, se Kærgård (1984). Men han fik hverken i Danmark eller internationalt elever eller betydning. Disputatsen var på dansk og kritiske opponenter forhindrede, at den fik nogen udbredelse.

Internationalt kom der imidlertid i 1910'erne og 1920'erne en forskning i gang, og det resulterede i starten af *Econometric Society* i 1932, der fra 1933 udgav tidsskriftet *Econometrica* med nordmannen Ragnar Frisch som redaktør fra starten til 1954. Dette selskab og dets tidsskrift blev hurtigt den ledende strømning i økonomisk forskning, og Frisch fik fuldt berettiget i 1969 den første Nobelpris i økonomi (delt med den hollandske modelbygger Jan Tinbergen).

Dels fordi transportmulighederne var begrænsede før flyvning blev almindelig, og dels fordi økonomer i høj grad skrev på deres nationale sprog og danskere, nordmænd, og svenskere forstod hinandens, var der en tæt kontakt mellem økonomerne i de nordiske

lande. Og i dette nordiske netværk var Frisch central. Der var derfor adskillige danskere med i *Econometric Society* fra starten. Men det var ikke økonometri i snæver forstand, men matematisk økonomi og statistiske metoder mere generelt, og ingen af de tidlige danske medlemmer ville vi i dag kalde økonometrikere.

Det store økonomistudium i Danmark var Det Statsvidenskabelige Studium (cand.polit.) ved Københavns Universitet. Her var der indtil 1962 valgfrihed mellem ”lille” og ”stor” statistik. Det var Anders Hald med lærebogen ”Statistical theory with engineering applications”, der stod for ”stor statistik”, og det gav mange økonomer et solidt fundament i statisk metode, selv om det var tekniske anvendelser, der var Halds hovedinteresse. En række cand.polit.’er med baggrund i ”stor statistik” med tilhørende valgfag fik en karriere i statistik, Ellen Andersen (professor i empirisk økonomi i København), Ebbe Yndgaard (professor på Aarhus Universitet), Ulf Christiansen (ekstern lektor i statistik på KU), Severin Olesen Larsen (statistiker på Seruminstittutet), m.fl.

Dette ændrede sig med Anders Halds flytning til det naturvidenskabelige fakultet i 1962 og Georg Rasch’s udnævnelse til hans efterfølger som professor ved Det Rets- og Statsvidenskabelige (det samfundsvidenskabelige) fakultet. Han var en inddæt modstander af regressionsanalyse, normalfordelingen og alle de metoder økonomerne anvendte, og Statistisk Institut og fagstatistiskerne kom derfor ikke til at spille nogen rolle ved udviklingen af økonomiske modeller.

Det blev derfor det valgfri fag ”økonometri”, der sammen med studieophold i udlandet blev stedet, hvor modelbyggerne blev uddannet. Der havde været undervisning i dette fag tilbage i tiden. Senere professor i driftsøkonomi Sven Danø oversatte og udgav således Jan Tinbergens hollandske lærebog i økonometri fra 1937, og Hector Estrup indførte i begyndelsen af 1960’erne Johnstons økonometrilærebog som pensum, men underviste i den uden at bruge matrix regning.

Det var Ellen Andersen, der blev lektor ved Økonomisk Institut på Københavns Universitet i 1965 og var på et halvt års studieophold i Oslo, der indførte moderen økonometriundervisning på internationalt niveau. Det blev hende og hendes elever, der kom til at præge modelbygningen i 1970’erne og de følgende årtier, se Kærgård (2020).

I Aarhus var det Ebbe Yndgaard, der startede økonometriundervisningen, men hans elever – specielt Svend Hylleberg og Henning Bunzel – interesserede sig mere for teori og metodeudvikling end bygning af anvendte økonometriske modeller.

2. Modelbygningen i 1970’erne og dens begrænsninger

Ellen Andersen startede altså sidst i 1960erne med at bygge det, der blev til ADAM. Den endelige offentliggørelse forsinkedes af, at sigtet med arbejdet var en disputats, og det krævede et bind 2 med en oversigt over teorier og ideer i verdens andre modeller

samt en længere periode til bedømmelse, så offentliggørelsen og forsvaret kom først i 1975 flere år efter første udgave af modellen var færdig, se Andersen (1975 og 1975a).

Allerede i 1969 oprettede Danmarks Statistik et prognosekontor, der blev til ADAM-gruppen, og som overtog udbygning og vedligeholdelse af modellen. Denne models praktiske brug til prognoser og konsekvensberegninger fik en flyvende start, da den første chef for Danmarks Statistikks modelkontor, Erling Jørgensen, i 1975 blev departementschef i Finansministeriet.

Ellen Andersens elever Jørgen Rosted, Christian Sørensen og Alexander Schaumann byggede i Det Økonomiske Råds sekretariat SMEC II, der var den første model, der fra 1974 blev brugt i praksis. Det blev også Jørgen Rosted, Erling Jørgensen ansatte som modelmand i Finansministeriet, se Rosted (2021).

Forbindelsen til det internationale modelmiljø var der dog også. Niels Blomgreen-Hansen, der stod for Nationalbankens første model, NATAN, udviklede modellen som ph.d.-afhandling hos Lawrence Klein i Pennsylvania. Poul Uffe Dam, der fra 1970 blev Danmarks Statistikks modelmand var i studietiden også et år i USA.

Når man ser på de tidlige modelbyggerses problemer, så var de ikke økonometrisk-statistiske. Både maximum likelihood estimation og mindste kvadraters metode var allerede for længst udviklede og grundigt analyserede estimationsmetoder. Specielt for simultane modeller var full information maximum likelihood en velkendt og i teorien ideel estimationsmetode. Der var også udviklet en lang række andre metoder til at estimere simultane modeller; two stage least square, limit information maximum likelihood osv.

Jens Warming diskuterede allerede i 1906 identifikationsproblemet, altså at når man f.eks. forklarer den omsatte mængde ved hjælp af priser, ved man ikke, om det er en udbuds- eller en efterspørgselskurve, man estimerer. Trygve Haavelmo havde allerede i 1944 udviklet problemerne for simultane modeller; det forhold, at man ikke bare kan estimere f.eks. en forbrugsfunktion, hvor man bruge indkomsten som forklarende variabel, fordi forbruget også påvirker indkomsten. Der er altså estimationsproblemer i simultane modeller.

Det, der var brug for, for at forbedre modellerne var ikke ny teori, men mere regnekraft og bedre data. Da Ellen Andersen konstruerede den første udgave af ADAM var datagrundlaget nationalregnskabstal for årene 1949-1965, dvs. 17 observationer. Beregningerne foregik på en mekanisk facit bordregnemaskine. At estimere en regressionsanalyse med de 17 observationer og to forklarende variable kunne hun ifølge eget udsagn klare på 20 minutter. Her måtte der derfor lægges afgørende vægt på forenkling af både modelbygningen og beregningerne.

Da 1960'ernes og 1970'ernes økonomiske problemer i høj grad vedrørte konjunkturregulering var det naturligt at koncentrere sig om *de kortsigtede sammenhænge* og se bort fra de langsigtede; det vil først og fremmest sige at udelade de langsigtede ligevægtsskabende mekanismér.

En anden central forenkling var en intens brug af *økonomisk teori* til at specificere de indgående variable og funktionsformen. Man kan f.eks. antage eller udlede, at produktionsfunktionen er en Cobb-Douglas funktion:

$$Y = A L^\alpha K^{(1 - \alpha)}$$

hvor Y er produktionen, L mængden af arbejdskraft og K mængden af kapital. Her er der kun to parametre, A og α , og funktionen kan let gøres til en lineær funktion med en forklarende variabel:

$$\ln Y/K = \ln A + \alpha \ln L/K$$

Sådanne praktiske løsninger blev til en metodeholdning. Ren empiri uden et økonomisk-teoretisk fundament blev betragtet som ”ad hoc’eri”, og Frisch formulerede det skarpt i sin Nobelforelæsning: ”*Facts that speak for themselves talk in a very naive language*”.

Problemer med manglende data kan også illustreres ved den blomstrende forskning i estimationsmetoder ved ”undersized semples”, se f.eks. Swamy & Holmes (1971), Problemet var, at der i de økonomiske modeller indgik et stort antal variable, mens man altså havde relativt få observationer. Man kunne så i situationer, hvor mange af de variable indgik samtidigt, f.eks. i første trin af 2SLS, komme i den situation, at der var flere forklarende variable end observationer. Et utal af løsningsforslag blev foreslæbt. I stedet for de mange variable kunne bruges nogle få ”principal component” i variabelmatricen. Eller man kunne ud fra økonomisk teori begrænse sig til de variable, der måtte antages at være vigtigst.

3. Udviklingen i de følgende årtier

Allerede tidligt i 1970’erne begyndte man at overføre data og beregningsprogrammer til hulkort for at fragte dem til NEUCC i Lyngby, og der gennemfører beregningerne elektronisk. Hurtigt udvikledes mere og mere effektive computere og mere og mere brugervenlige programmer.

Samtidigt kom der flere og flere data. Nationalregnskaberne blev udbygget, og der kom automatisk et nyt år og dermed en ny observation til for hvert år. Hvor knappe data var, og hvor vigtigt det var med flere, kan illustreres med en tradition i økonomisk instituts empiriske afdeling, hvor Ellen Andersen arbejdede med ADAM’s relationer, og jeg skrev ph.d.-afhandling om billefterspørgslen. Hvert år, når der kom et nyt år i nationalregnskabsstatistikken, gav Ellen wienerbrød til kaffen, og når der kom den årlige opgørelse af bilbestanden, gav jeg wienerbrød. En ekstra observation var noget, der betød noget.

1960’ernes og 1970’ernes problemer med data og regnekraft blev gradvist afhjulpet, og det blev i høj grad det, der bestemte udviklingen i økonometriens. Men både udviklingen i computernes regnekraft, programmernes brugervenlighed og i statistikproduktionen er naturlige emner for selvstændige analyser og falder uden for denne artikels rammer.

Dermed falder også det store arbejde Henning Bunzel og hans medarbejdere gjorde i Aarhus bl.a. med programsystemet ISISMA uden for denne artikels emne.

Et centralt tema blev, da data og regnekraft tillod mere komplicerede modeller at kombinere det korte og det lange sigt. De kortsigtede konjunkturbevægelser fører jo ikke til, at økonomien kører helt ad sporet. Der optræder på længere sigt nogen ligevægtsskabende mekanismer. Bliver arbejdsløsheden meget høj, resulterer det i lave lønninger, og lave lønninger giver lave priser og god konkurrenceevne, og så stiger efterspørgslen og beskæftigelsen.

Introduktionen af det lange sigt og ligevægtsskabende mekanismer skete på flere måder. Rent praktisk introducerede Finansministeriet sidst i 1980'erne en række ligevægtsskabende mekanismer. En Phillips-kurve gjorde lønnen afhængig af arbejdsløsheden, og man kæmpede intenst for at gøre eksporten mere følsom overfor konkurrenceevnen. Det har været en bestræbelse på gennem årtier at få eksportens priselasticitet gjort større (i økonomisk teori antages ofte, at der eksisterer en fast verdensmarkedspris, som en lille åben økonomi er nødt til at holde sig til, dvs. priselasticiteten er uendelig, men i praksis er det svært at estimere rimeligt store priselasticiteter). Endelig indførte man mål for kapacitetsudnyttelsen som forklarende variabel. Alt dette gav anledning til heftig debat, jf. Kærgård (2020a).

Teoretisk blev det lange og det korte sigt kombineret i de såkaldte error correction modeller. En langsigtet ligevægtsrelation (hvor alle de variable kan være matricer)

$$Y = A + BX$$

blev indlagt i den kortsigtede relation mellem ΔY og ΔZ

$$\Delta Y = a + b \Delta Z + c (Y - (A + BX))$$

Som forklarende variabel i kortsigtsrelationen indgår altså residualerne i langsigtsrelationen (der kan være et betydeligt overlap mellem de variabler, der indgår i X og i Z).

Sådanne kombinationer blev fra sidst i 1980'erne udviklet til meget avancerede co-integrationsmodeller, der på en meget elegant måde kombinerer det lange og det korte sigt, se f.eks. Johansen (1988) og Engle & Granger (1991).

Et andet nyt område, der har præget de seneste årtiers økonometri er mikroøkonometrien. Der er ved siden af de gamle nationalregnskaber kommet en righoldig kortlægning af danskernes økonomiske adfærd ved hjælp af registerdata. Der er desuden kommet et stort antal spørgeskemaundersøgelser af alle mulige slags.

Paneldata, hvor man følger en gruppe agenters adfærd over tiden er derved blevet almindelige. Det har givet anledning til ny teoriudvikling f.eks. i form af de såkaldte Fixed Effect modeller, se f.eks. Allison (2009) og til et væld af empiriske undersøgelser. Her i landet er specielt værd at nævne arbejdsmarkedsgruppen i Aarhus med Peder J. Pedersen, Henning Bunzel og Nina Smith og deres elever.

Den righoldige mængde af data og computerkraft til at håndtere dem har også for mange forskere ændret holdningen til kombinationen mellem data og økonomisk teori. Nu taler data, der taler for sig selv, ikke længere nødvendigvis et naivt sprog. ”Mine data taler så tydeligt, at jeg ikke behøver at give dem mikrofon”, som det blev sagt ved et ph.d.-forsvar, da opponenterne efterlyste mere teori. Modelbygning var ikke længere bare estimation af en fra den økonomiske teori given model. Det er tydelig formuleret allerede i 1986 i Spanos’ lærebog i økonometri:

Econometrics is concerned with the systematic study of economic phenomena using observed data. --- This definition is much broader than certain textbook definitions narrowing the subject matter of econometrics to the “measurement” of theoretical relationships as suggested by economic theory. It is argued in the epilogue that the latter definition of econometrics constitutes a relic of an outdated methodology. (Spanos, 1986, p. 3)

Slagordet blev nu ”allowed the data to speak freely” (Hoover, Johansen & Juselius, 2008).

4. Fremtidens modeller

Denne sætten økonomisk teori i skammekrogen blev ikke accepteret af alle. Der kom en ny type af modeller i form af de anvendte generelle ligevægtsmodeller, AGE- eller CGE-modellerne. Alle modeller er ligevægtsmodeller i den forstand, at de indeholder ligevægtsbetingelser. Selv den simple Keynes-model har jo en betingelse om, at den udbudte og den efterspurgte mængde er den samme. Det, der gør ”ligevægtsmodeller” til ligevægtsmodeller, er, at udgangspunktet er den Walrasianske ligevægtsmodel, hvor forbrugerne nyttemaksimerer, producenterne profitmaksimerer, og der er ligevægt på alle markeder (i Keynes-modellen er der ligevægt på varemarkedet, men ikke på arbejdsmarkedet).

Også denne modeltradition bygger på udviklingen i regnekraft og dataindsamlingen. Det er ofte meget store modeller f.eks. med mange generationer af agent og modelkonsistente forventninger. Løsningen af sådanne modeller ville være helt umulig uden moderne computerkraft.

I disse modeller ”kalibreres” parametrerne, dvs. de tages i høj grad fra andre undersøgelser, måske fra andre lande:

The procedure most commonly used to select parameter values has come to be labelled “calibration” --- The economy under consideration is assumed to be in equilibrium, a so-called “benchmark” equilibrium. The parameters of the model are chosen such that the model can reproduce this data set as an equilibrium solution -- Typically, calibration involves only one year’s data, or a single observation represented as an average over a number of years. A crucial point in using calibration is that because of the reliance on a single observation, the benchmark data typically do not identify a unique set of values for the parameters

in any model. Particular values for the relevant elasticities are usually specified on the basis of other research, and these serve to identify uniquely the other parameters of the model along with the equilibrium observation. This typically places a lot of reliance on literature surveys of elasticities and, as many of the modellers have observed in discussing their own work, it is surprising how sparse (and sometimes contradictory) the literature is on some elasticity values. (Shoven & Whalley, 1984, pp. 1018-1020).

En sådan kalibrering kan forekomme som en naturlig udvikling. Når en stor del af den økonomiske viden kommer fra analyser af mange forskellige datakilder, nationalregnskabet, registerdata, spørgeskemaundersøgelser osv. er det unødvendigt restriktivt at holde sig så meget til nationalregnskabet, som man gjorde i 1970'ernes modeller. Det kan f.eks. være naturligt at tage nogen parametre fra arbejdsmarkedet fra register- og spørgeskemaundersøgelser. Estimater kan også kombinere flere forskellige datakilder.

Det rejser dog to problemer. For det *første* er det et problem, hvis "lige vægtsstrukturen" slår for hårdt igennem, dvs. hvis de lige vægte, der er en usikker langsigtet attraktion, kommer til at dominere beskrivelsen af den kort- og mellemfristede udvikling for meget. Virkelighedens økonomier er jo langtfra altid præget af en hurtig tilpasning til lige vægt-

For det *andet* er der en større frihedsgrad ved valg af parametre ved kalibrering end ved estimering. Når man vælger parameterværdier fra andre undersøgelser, er der meget at vælge imellem, og fastlæggelse af parametrene bliver derfor i højere grad et politisk valg, end hvis de estimeres ud fra et samlet datasæt f.eks. med udgangspunkt i nationalregnskabet. Muligheden for politiske betingede valg stiger derfor, og det bliver centralt, at modelbygningen foretages af en institution med et etableret ry for at være politisk uafhængig. Her var den gamle opdeling, hvor ADAM blev bygget og vedligeholdt i Danmarks Statistik, men brugt ved politikformuleringen af Finansministeriet ret ideel.

Litteratur

- Allison, P.D. (2009), *Fixed effects regression models*, SAGE publications.
- Andersen, Ellen (1972), Forudsigelser af den økonomiske udvikling, *Nationaløkonomisk Tidsskrift, bind 110, 1-2, side 27-35*.
- Andersen, Ellen (1975), *En model for Danmark 1949-1965*, Studie nr. 21, Akademisk Forlag, København, 311 sider. (1. del af disputatsen).
- Andersen, Ellen (1975a), *Træk af makroøkonometriske modellers historie og udvikling*, studie nr. 20, Akademisk Forlag, København, 143 sider (Tidligere udgivet som Cykelafdelingens gule memoserie nr. 15), (2. del af disputatsen).

- Blomgren-Hansen, Niels (1975), Strukturen i Nationalbankens finansielle model, *Nationaløkonomisk Tidsskrift*, Bind 113.3, side 181-195
- Engle, R. & C. Granger (1991), *Long-run economic relationships: Readings in cointegration*, Oxford University Press, Oxford.
- Hansen, Jørgen & Martin Paldam (1973), *SMEC. En kvartalsmodel af den danske økonom*, Københavns Universitetets Økonomiske Institut og Det Økonomiske Råds sekretariat, København.
- Hoover, Kevin D., Soren Johansen & Katarina Juselius (2008), Allowing the Data to Speak Freely: The Macro econometrics of the Cointegrated Vector Auto regression. *American Economic Review*, vol. 98(2), pp. 251-55.
- Johansen, Søren (1988), Statistical analysis of cointegration vektors, *Journal of economic dynamics and control*, vol. 12. 2-3, side 231-254.
- Kærgård, N. (1984), The Earliest History of Econometrics: Some neglected Danish contributions, *History of Political Economy*, Vol. 16, pp. 437-444.
- Kærgård, Niels (2020), Ellen Andersen, 1. marts 1937 – 19. april 2018, *Nationaløkonomisk Tidsskrift*, https://www.djoef-forlag.dk/publications/noet/files/2020/articles/artikel_2.pdf.
- Kærgård, Niels (2020a), Finansministeriet og makroøkonomiske regnemodeller: Et historisk rids, *Samfundsøkonomen*, 2020.1, side 31-40-
- Rosted, Jørgen, Alexander Schaumann og Christen Sørensen (1974), *SMEC II. Måling af finanspolitikkens aktivitetsvirkning*, Det Økonomiske Råds sekretariat, København.
- Rosted, Jørgen, Alexander Schaumann og Christen Sørensen (1974), Finanseffekt og multiplikatorer i SMEC II, *Nationaløkonomisk Tidsskrift*, Bind 112.2, side 267-297.
- Rosted, Jørgen (2021), *Den økonomiske genopretning 1976-1993*, Historia, Odense
- Shoven, J. & J. Whalley (1984), Applied General Equilibrium Models of Taxation and International Trade: An Introduction and Survey, *Journal of Economic Literature*, vol. 23, pp. 1007-1051.
- Spanos, Aris (1986), *Statistical foundations of Econometric Modelling*, (with Foreword by David Hendry), Cambridge University Press, Cambridge.
- Swamy, P.A.V.B. & Janus Holmes (1971), The use of undersized samples in the Estimation oaf Simultaneous Equation Systems, *Econometrica*, vol. 39.3, side 455-459.

Estimering af input-output koefficienter fra aggregeret data på virksomhedsniveau*

Rasmus Seneberg Zitthen, Arne Henningsen,
Simon Alexander Andreasen, Mads Frandsen
og Mathias Struck Jürgensen

Institut for Fødevare- og Ressourceøkonomi, Københavns Universitet

Januar 2022

Abstrakt: Virksomheder producerer ofte multiple output, mens inputmængder i regnskabsdata ofte er aggregeret til virksomhedsniveau og andellen af hvert inputmængde som er anvendt i produktionen af hvert output er derfor ukendt. Eftersom mange analyser kræver data på inputmængderne for hvert af de producerede output, i særdeleshed analyser som er anvendt i både driftøkonomisk og politisk beslutningstagen. Dette kapitel opsummerer de væsentlige resultater i Zitthen et al. (2021), som vurderer og sammenligner forskellige metoder til estimering af input-output koefficienter for virksomheder som producerer multiple output, hvilket kan benyttes til at opnå output-specifikke inputmængder. Vi skelner mellem fire økonometriske metoder: (1) klassisk økonometriske metoder, (2) Random Coefficient Regression, (3) Entropy-baseret metoder og (4) Bayesiansk økonometri. Klassisk økonometrisk metode er ofte mangelfuld og forårsager urealistiske og upålidelige estimater. Random Coefficient Regression tager højde for at virksomheder i realiteten ikke har de samme input-output koefficienter. Entropy-baseret metoder og Bayesiansk økonometri kan leve mere pålidelige estimater, eftersom disse metoder tillader mere fleksible modelspecifikationer og kan inkorporere prior viden. Sammenlignet med Entropy-baseret metoder, har Bayesiansk økonometri den fordel, at kunne inkorporere prior viden på en mere transparent måde. Vi konkluderer at en kombination af Random Coefficient Regression og Bayesiansk økonometri ser ud som den mest egnede metode til at estimere input-output koefficienter og output-specifikke inputmængder.

*Nærværende kapitel er et uddrag af Zitthen et al. (2021), hvori vi præsenterer den grundlæggende økonometriske teori bag estimering af output-specifikke inputallokeringer for landbrugsbedrifter. Herudover præsenterer Zitthen et al. (2021) den eksisterende litteratur på området og diskuterer styrkerne og svaghederne ved de mest brugte estimeringsmetoder.

1 Introduktion

Virksomheder i mange sektorer, herunder landbrugsbedrifter, bruger ofte multiple input til at producere multiple output. Ydermere går samme input ofte igen i produktionen af flere forskellige output. Dette ses f.eks. med et input som gødning, der både bruges i produktionen af byg, hvede, havre m.fl. De aggregerede inputmængder samt outputmængder opgøres i regnsskabsdata, mens de outputspecifikke inputallokeringer ofte udelades (Just et al., 1983). Viden om, hvor meget af et givet input kræves til produktion af et specifik output, er imidlertid essentiel for både udarbejdelsen af troværdige budgetkalkyler til driftøkonomiske beslutningstagen, men også ift. hvordan ændringer i de politiske rammevilkår påvirker erhvervet (Just et al., 1990; Louhichi et al., 2012). Sidstnævnte er især relevant i landbrugssektoren, da denne sektor ofte er yderst reguleret. Den fornødne indsigt kunne givetvis opnås gennem spørgeskemaer, i hvilke man spørger den enkelte landmand, hvilke og hvor store mængder af input, der er brugt til et givet output. Denne metode er dog både dyr og yderst tidskrævende. Herudover er det ej heller sikkert, at det enkelte landbrug har præcise informationer om inputallokeringen (Just et al., 1990). Gennem tiden, er der derfor blevet forslået en række simplere og mere hensigtsmæssige metoder til at finde de outputspecifikke inputallokeringer, hvorfaf mange tager deres udgangspunkt i økonometriske estimeringer (f.eks. Errington, 1989; Heckelei et al., 2008; Just et al., 1983; Léon et al., 1999; Louhichi et al., 2012). Ofte benævnes information om input mængder, der bruges i produktionen af en enhed output eller en enheds produktionsaktivitet som henholdsvis input-output koefficienter, inputallokeringskoefficienter eller inputaktivitetskoefficienter (f.eks. Errington, 1989; Gocht, 2008).

2 Økonometriske metoder

Dette afsnit gennemgår økonometriske metoder, der er blevet brugt til at estimere input-output koefficienter og/eller outputspecifikke inputmængder. Vi udelader detaljerede udledninger og beskrivelser af de gennemgående estimeringsmetoder, men præsenterer de generelle ideer om disse metoder og vurderer og diskuterer deres styrke og svagheder.

I de sidste fire årtier er flere forskellige metoder blevet foreslået til at estimere input-output koefficienter og/eller outputspecifikke inputmængder (f.eks. Just et al., 1983; Errington, 1989; Just et al., 1990; Moxey and Tiffin, 1994; Lence and Miller, 1998; Léon et al., 1999; Louhichi et al., 2012; Fragoso and Carvalho, 2013; Lips, 2014). Vi kategoriserer de foreslæde metoder i fire kategorier: klassiske økonometriske metoder, Random Coefficient Regression, Entropi-baserede metoder og Bayesiansk økonometri.

Disse fire kategorier af estimeringsmetoder er kort forklaret, vurderet og diskuteret i henholdsvis afsnit 2.1, 2.2, 2.3 og 2.4.

Vores gennemgang af metoder omfatter ikke metoden til at estimere outputspecifikke inputmængder foreslæbt af De Loecker et al. (2016), som er baseret på et økonometrisk skøn der tager hensyn til bedrifters observationer som kun producerer et enkelt output. Metoden kan derfor kun anvendes, hvis der for hver af de betragtede output er et tilstrækkeligt stort antal bedrifter, som kun producerer dette output (og ingen andre output), hvilket sjældent er tilfældet i empiriske anvendelser. En anden ulempe ved denne metode er, at den forudsætter at fordelingen af input til output er den samme for alle input,¹ hvilket er en usandsynlig antagelse i mange empiriske anvendelser.²

Ikke-økonometriske metoder, især positiv matematisk programmering, kan også bruges til at opnå output-specifikke input-mængder (f.eks. Howitt, 1995), men i denne gennemgang overvejer vi kun økonometriske tilgange.

2.1 Klassiske Økonometriske Metoder

I Zitthen et al. (2021) diskuterer vi, om 'loven om én pris' gælder, dvs. om alle bedrifter står over for de samme priser for de samme input og output, når disse input og output har samme karakteristika på tværs af bedrifterne (dvs. input og output med forskellige karakteristika, f.eks. godtning med forskelligt næringsindhold eller hvede med forskelligt proteinindhold, kan have forskellige priser), og om priser tager højde for kvalitetsforskelle i input og output mellem bedrifter. Vi konkluderer at det ofte er rimeligt at antage, at 'loven om én pris' er opfyldt i vid udstrækning, hvorfor vi kan estimere regressioner i monetære værdier, som kan anses som mængdeindeksler, der tager højde for kvalitetsforskelle.

Griliches (1963), Errington (1989), Hallam et al. (1999) m.fl. foreslår, at anvende regressionsmetoden Ordinary Least Squares (OLS) til at estimere f.eks. de samlede omkostninger for hvert input:

$$x_{li} = \sum_{j=1}^J a_{lj} y_{ji} + \varepsilon_{li} \quad \forall l = 1, \dots, L, \quad (1)$$

¹Forfatterne er Emir Malikov taknemmelige for at gøre dem opmærksomme på metoden foreslæbt af De Loecker et al. (2016) samt for at have påpeget denne ulempe.

²F.eks. for en bedrift, der producerer korn og husdyr, vil det f.eks. betyde, at procentdelen af dens samlede godtningssinput, der bruges til kornproduktion, er den samme som procentdelen af dens samlede foderstofinput, der bruges til kornproduktion, og som følge heraf, at den procentdel af dens samlede godtningssinput, der bruges til husdyrproduktion, er den samme som den procentdel af dens samlede foderstofindsats, der bruges til husdyrproduktion.

hvor subscript $i = 1, \dots, N$ angiver bedriften, x_{li} angiver omkostningerne ved input l for bedrift i , a_{lj} angiver den input-output koefficient for input l og output j , y_{ji} angiver omsætningen fra output j for bedrift i , L er det samlede antal af input, J er det samlede antal af output og ε_{li} er tilfældige fejlertermer, der antages at være uafhængigt og identisk fordelt (iid).

De vigtigste fordele ved at anvende klassiske økonometriske metoder, såsom OLS, til at estimere input-output koefficienter er, at (1) de er enkle, (2) de kan give konfidensintervaller for input-output koefficienterne og (3) de er velegnet til at teste mere generelle økonometriske antagelser (Midmore, 1990). De to første punkter hænger sammen, og de gør, at man (med lidt baggrundsviden i økonometri) nemt kan opnå estimerater af input-output koefficienterne. De opnåede estimerater er dog kun gyldige, hvis visse antagelser er opfyldt, f.eks. en passende funktionel form af regressionsligningen. Dette fører os til punkt (3), som angiver, at forskellige standard testprocedurer kan bruges til at teste forskellige antagelser, der er nødvendige for at opnå middelrette og efficiente estimerater med OLS og andre klassiske økonometriske metoder.

Anvendelsen af OLS og andre klassiske økonometriske metoder til at estimere inputallokering er imidlertid blevet stærkt kritiseret (f.eks. af Mittelhammer et al., 1981; Just et al., 1983; Midmore, 1990; Lence and Miller, 1998), f.eks. på grund af heteroskedasticitet, høj multikollinearitet, endogenitetsproblemer, ikke-homogenitet på tværs af observationer, potentiel ikke-lineæritet og korrelation af fejlled mellem regressionsligninger.

F.eks. regressionsligning (1) er ofte plaget af meget høj multikollinearitet, fordi mindre bedrifter ofte har lave omsætninger af alle de output de producerer, mens at større bedrifter ofte har høje omsætninger for de fleste output, de producerer. For at reducere problemet med multikollinearitet skelner Errington (1989) ikke mellem forskellige typer afgrøder, men bruger kun ét aggregeret "afgrøde"-output. Denne tilgang er dog ofte utilstrækkelig, fordi mange efterfølgende analyser kræver input-output koefficienter for individuelle afgrøder. En anden måde at håndtere multikollinearitet på er brugen af større datasæt, men denne mulighed er normalt umulig.

Frekventistiske økonometriske metoder, der forsøger at løse multikollinearitet, såsom Ridge-regression og Lasso-regression, synes at være uegnede til at estimere input-output koefficienter, fordi disse metoder giver estimerater, der er biased mod nul. De eneste egnede metoder til at adressere multikollinearitet i estimeringen af input-output koefficienter synes at være entropi-baserede metoder og Bayesianske metoder, som er introduceret i henholdsvis afsnit 2.3 og 2.4.

Ud over multikollinearitet er regressionsligning (1) ofte plaget af væsentlig heteroske-

dasticitet, fordi variationen i inputmængderne eller af omkostningerne ved input stiger med bedriftens produktionsvolumen (Midmore, 1990; Léon et al., 1999). Man kunne adressere heteroskedasticiteten ved at estimere regressionsligningerne med metoden (Feasible) Weighted Least Squares ((F)WLS) regression. Men vi fandt ikke nogen undersøgelse, der bruger denne metode til at estimere input-output koefficienter. En anden metode, der adresserer heteroskedasticitet, er Random Coefficient Regression. Da vi ikke anser denne metode for at være en klassisk økonometrisk metode, diskuterer vi denne metode i afsnit 2.2.

I betragtning af at regressionsligningerne ofte er plaget af betydelig multikollinearitet og heteroskedasticitet, bliver estimatorne ofte meget upræcise. Den høje unøjagtighed medfører, at nogle estimerede input-output koefficienter er alt for høje, og nogle andre estimerede input-output koefficienter er alt for lave, f.eks. negative (Errington, 1989). Fra et praktisk synspunkt er meget upræcise estimatorer, f.eks. negative koefficienter eller alt for store estimatorer, ikke realistiske og ugyldiggør derfor fortolkningen af estimatorne. For at løse problemet med at have negative koefficienter kan man pålægge koefficienterne ikke-negativitetsbegrænsninger, når man estimerer modellen, men — som Moxey and Tiffin (1994) viser i en empirisk undersøgelse — dette fører blot til de koefficienter, der er negativ i ubegrænsede estimeringer til at være lig med nul i den begrænsede estimering.

Udover de økonometriske problemer, er der blevet stillet spørgsmålstege ved estimering af ligning (1) med OLS-metoden, fordi antagelserne om nonjointness mellem produktionen af de forskellige output sandsynligvis er urealistiske i mange empiriske applikationer (Lence and Miller, 1998; Gocht, 2008).

Sammenfattende har klassiske økonometriske metoder vist sig at have mange svagheder (f.eks. Mittelhammer et al., 1981; Just et al., 1983; Errington, 1989; Just et al., 1990; Midmore, 1990; Moxey and Tiffin, 1994; Lence and Miller, 1998; Gocht, 2008).

2.2 Random Coefficient Regression

Modellen nævnt i afsnit 2.1 antager at der er ens input-output koefficienter på tværs af samtlige virksomheder. Den antagelse synes imidlertid relativt urealistisk, da forholdet mellem input og output ofte påvirkes af bedriftsspecifikke forhold, såsom produktionsstørrelse, produktionsgrenene på bedriften, jordtype og -bonitet, vejrforhold m.v. f.eks. i landbrugsbedrifter. Flere studier foreslår derfor at lemppe denne antagelse og lade input-output koefficienterne varierer på tværs af bedrifterne (f.eks. Dixon et al., 1984; Hornbaker et al., 1989; Scandizzo, 1990; Dixon and Hornbaker, 1992; Wikström et al., 2011). Specifikt foreslår Dixon et al. (1984), Hornbaker et al. (1989) og Dixon and

Hornbaker (1992) brugen af Random Coefficient Regression (RCR). RCR tillader hver bedrift at have deres eget sæt af input-output koefficenter, som individuelt kan afhænge af de førnævnte observerede faktorer, såvel som ikke-observerede faktorer. Brugen af RCR kan implementeres i mange forskellige regressionsmodeller, herunder også dem nævnt i Zitthen et al. (2021). I det følgende beskrives RCR med udgangspunkt i ligning (1). Da regressionens underliggende funktioner er identiteter³, er det inkonsekvent at inkludere fejleddet ε_{li} , og fejleddet udelades ofte i estimeringen af RCR. For at tillade, at input-output koefficenterne a_{lj} varierer mellem bedrifterne tilføjes subscript i . Derved opnår vi følgende ligning:

$$x_{li} = \sum_{j=1}^J a_{lji} y_{ji} \quad \forall l = 1, \dots, L, \quad (2)$$

hvor input-output koefficenterne er estimeret som:

$$a_{lji} = B_{lj0} + \sum_{m=1}^M B_{ljm} \eta_{mi} + \kappa_{lji} \quad \forall l = 1, \dots, L, j = 1, \dots, J, \quad (3)$$

hvor M er antallet af variable, der forklarer input-output koefficenterne, η_{mi} indikerer den m 'de variabel, der forklarer input-output koefficenterne, B_{lj0} og B_{ljm} er koefficenterne, som skal estimeres og κ_{lji} er fejlledet, der fanger effekten af ikke-observerede variable på input-output koefficenterne. I ligning (3), kan vi se at samtlige input-output koefficenter afhænger af præcis de samme forklarende variable. Det er imidlertid muligt at bruge forskellige sæt af forklarende variable i estimeringen af de individuelle input-output koefficenter:

$$a_{lji} = B_{lj0} + \sum_{m \in \mathcal{M}_{lj}} B_{ljm} \eta_{mi} + \kappa_{lji} \quad \forall l = 1, \dots, L, j = 1, \dots, J, \quad (4)$$

hvor \mathcal{M}_{lj} indikerer sæt af variabler, som input-output koefficienten a_{lj} afhænger af.

Indsætter vi ligning (4) i ligning (2), får vi:

$$x_{li} = \sum_{j=1}^J \left(B_{lj0} + \sum_{m \in \mathcal{M}_{lj}} B_{ljm} \eta_{mi} + \kappa_{lji} \right) y_{ji} \quad \forall l = 1, \dots, L \quad (5)$$

$$= \sum_{j=1}^J B_{lj0} y_{ji} + \sum_{j=1}^J \sum_{m \in \mathcal{M}_{lj}} B_{ljm} \eta_{mi} y_{ji} + \sum_{j=1}^J \kappa_{lji} y_{ji} \quad \forall l = 1, \dots, L. \quad (6)$$

³Se Zitthen et al. (2021), for yderligere uddybning.

Foruden at lempe den ofte urealistiske antagelse om ens input-output koeficienter på tværs af bedrifter, har RCR også den store fordel, at metoden i høj grad også løser de OLS og andre klassiske økonometriske metoders problemer ift. heteroskedasticitet, fordi variansen af det sammesatte fejled, $\sum_{j=1}^J \kappa_{lji} y_{ji}$ kan samvarierer med produktionsomfanget y_{ji} . RCR lider dog under, at metoden øger problemerne med multikolinaritet, da de mange interaktionsled som bliver brugt som forklarende variable ($\eta_{mi} y_{ji}$) er korreleret med hinanden og med de ‘enkelte’ forklarende variable (y_{ji}).

2.3 Entropi-baserede Metoder

I jagten på mere plausible og pålidelige resultater på trods af multikolinaritet foreslås flere alternativer til de klassiske økonometriske metoder, herunder både entropi-baserede og bayesianske metoder (se afsnit 2.4). Idéen er, at inkorporere såkaldt *non-data information* (ofte kaldet prior information). Hermed forstår altså viden, som undersøgeren måtte have om emnet, inden studiet startes. Før har de entropi-baserede metoder været set som en selvstændig metodisk ramme, men det er senere vist, at de entropi-baserede også falder inden for den bayesianske økonometri. Metoderne er anseeligt mere komplekse end de klassiske metoder (OLS m.fl.), men har fået indpas, da de generelt giver væsentligt mere pålidelige resultater (f.eks. Moxey and Tiffin, 1994; Lence and Miller, 1998; Léon et al., 1999; Fragoso and Carvalho, 2013; Louhichi et al., 2018). I Zitthen et al. (2021) tager vi udgangspunkt i *Generalised Maximum Entropy (GME)*. Metoden blev først introduceret af Shannon (1948), og starter med de ikke-observerede sandsynligheder, $\rho = [\rho_1, \rho_2, \dots, \rho_K]$:

$$H(\rho) = - \sum_{k=1}^K \rho_k \ln(\rho_k), \quad (7)$$

hvor $\rho_k \ln(\rho_k) \equiv 0$ for $\rho_k = 0$ og K er antallet af datapunkter, ofte også kaldet støttepunkter (Léon et al., 1999). Mere specifikt angiver, ρ_k sandsynligheden for at observere datapunkt k (Fragoso and Carvalho, 2013). Givet dette, når $H(\rho)$ sit maksimum når $\rho_1 = \rho_2 = \dots = \rho_k = \frac{1}{K}$, hvilket er ækvivalent til, at sandsynlighederne er uniformt fordelt. Målet er nu, at maksimere $H(p)$, i hvilken der (hvis relevant) kan implementeres yderligere restriktioner. Idéen er derved, at de estimerede sandsynligheder for ρ_k vil trækkes mod en uniform fordeling under bibetingelse af de valgte restriktioner. Disse restriktioner vil her være en kombination af pågældende data samt undersøgerens prior information (Léon et al., 1999). I Zitthen et al. (2021) går vi yderligere i dybden med GME, samt alternativet Generalised Cross Entropy (GCE). Vi finder, at GME har mange fordele,

men også sine svagheder. Blandt fordelene er, at (1) der ikke kræves nogen antagelser ift. fejlleddets fordeling, (2) metoden løser udfordringen med multikolinaritet, (3) metoden kan benyttes på mindre datasæt og (4) metoden tillader ikke-lineære restriktioner og ulighedsrestriktioner (Louhichi et al., 2012; Fragoso and Carvalho, 2013). Den største udfordring ved GME fremkommer ved måden, hvorpå støttepunkterne er konstrueret. Dette fordi, at støttepunkterne har stor betydning for entropi-modellens resultater. Der kan tages højde for de nævnte udfordring ved brug af prior information (Louhichi et al., 2012) eller i stedet ved bruge af GCE (Lence and Miller, 1998). Ved brugen af prior information skal undersøgeren dog stadig bestemme såkaldte *støttebånd* og bredden herpå, hvilket igen kan have stor betydning for de endelige resultater.

2.4 Bayesiansk Økonometri

For at imødekomme udfordringerne fundet ved entropi-baserede metoder, undersøger vi i Zitthen et al. (2021) også brugen af bayesiansk økonometri til bestemmelse af input-output koefficienter. Herud over undersøges fordele og ulemper ved den Bayesianske metode. Den Bayesianske metode er baseret på ideen om at kombinere en prior forventet fordeling af de ukendte α -parametre med informationen givet af data baseret på Bayes' teorem (Moxey and Tiffin, 1994; Coelli et al., 2005, p. 231-234; Heckelei et al., 2008). I analysen behandler vi alle parametre i estimeringen som stokastiske variable (Gocht, 2008; Heckelei et al., 2008). Ydermere skelner vi mellem prior density for koefficienterne i form af en prior probability density function (pdf) angivet ved $p(\alpha)$, stikprøve informationen i form af en likelihood function⁴ baseret på dataet, $L(\alpha, y)$, og posterior pdf, $h(\alpha | y)$. På baggrund af Bayes' teorem finder vi, at posterior pdf er proportional til prior pdf multipliceret med likelihood funktionen (Geweke, 1999; Coelli et al., 2005, p. 231-234; Gocht, 2008):

$$h(\alpha | y) \propto L(\alpha, y) \cdot p(\alpha). \quad (8)$$

Prior fordelingen kan enten specificeres som informativ eller som ikke-informativ prior. Som navnet indikerer bidrager en ikke-informativ prior ikke med forudgående viden om parametrene. I dette tilfælde er det kun data (og ikke den forudgående viden) der påvirker parameterestimaterne. I modsætning til indragelse af informativ prior, som bidrager med forudgående information om parameterestimaterne. En informativ prior er ofte en forudgående fordeling omkring en forventet middelværdi og en passende varians.

⁴I Heckelei et al. (2008) og Louhichi et al. (2018) er likelihood funktionen angivet som $L(\alpha | y)$, hvorimod Geweke (1999) og Gocht (2008) anvender $L(y | \alpha)$.

Oplysninger om en prior fordeling (f.eks. hvilken type af fordeling, middelværdi, varians) kan komme fra faglitteratur, undersøgelser eller tidlige studier (Louhichi et al., 2018).

I en empirisk undersøgelse fandt Moxey and Tiffin (1994), at den Bayesianske metode var meget god sammenlignet med den inequality-constrained least-squares (ICLS) metode. Moxey and Tiffin (1994) fremmer brugen af Bayesianske priors, når de arbejder med begrænsningsestimationer. Udover dette er det velkendt, at GME/GCE og den Bayesianske metode ligner hinanden meget og indeholder mange af de samme fordele såvel som ulemper (Heckelei et al., 2008; Louhichi et al., 2012). Nogle yderligere fordele ved den Bayesianske metode er blevet påpeget af Heckelei et al. (2008). For det første kan den Bayesianske metode angives således, at den svarer til GME/GCE-metoden⁵, for det andet er de implementerede priors mere transparente, fordi prior pdf' er tildeles direkte til de ukendte. Endelig har denne metode færre variabler og kan derfor være mindre krævende i form af regnekraft (Heckelei et al., 2008).

I betragtning af disse fordele ser den Bayesianske metode ud til at være meget attraktiv til at estimere input-output koefficienter. Men da litteraturen kun indeholder en begrænset mængde studier, der bruger Bayesiansk økonometri på dette område, skal der udføres yderligere forskning i, hvordan man gør dette i praksis.

2.5 Prior

Som sagt kan oplysningerne om prior information komme fra faglitteratur, undersøgelser eller tidlige studier (Louhichi et al., 2018). For en empirisk anvendelse for danske landbrugsbedrifter forslår vi, at prior fordelinger af input-output koefficienterne specificeres således, at den forventede værdi er et vægtet gennemsnit af tidlige års estimerede input-output koefficienter:

$$a_{lj}^* \sim \Psi(\mu_{lj}, \sigma_{lj}^2) \quad (9)$$

$$\mu_{lj} = \sum_{t=1}^T \tau_t \hat{a}_{lj|t}^*, \quad (10)$$

hvor a_{lj}^* angiver input-output koefficienten for input l og output j for det gældende år, $\hat{a}_{lj|t}^*$ er input-output koefficienten for input l og output j estimeret for t år siden, τ_t er et vægt for input-output koefficienten estimeret for t år siden, T er antallet af år som er anvendt til at beregne det vægtede gennemsnit, og $\Psi(\mu_{lj}, \sigma_{lj}^2)$ er en passende fordeling med en forventet værdig μ_{lj} og variansen σ_{lj}^2 . Vi specificerer det vægtede gennemsnit

⁵Dette er også tilfældet ved brugen af faste input (Gocht, 2008).

således, at de seneste år vægtes højere end år længere tilbage:

$$\tau_t = \tau_1 (1 - r)^{t-1} \quad \forall t = 2, \dots, T \quad (11)$$

$$\tau_1 = \left(\sum_{t=1}^{T-1} (1 - r)^{t-1} \right)^{-1}, \quad (12)$$

hvor $r \in [0, 1]$ er den relative sats, hvormed vægten falder, når man går et år tilbage i tiden, og ligning (12) normaliserer vægtene, så de summerer op til én. Vi tror, at et årligt fald i vægten med 10%, dvs. $r = 0,1$, er tilstrækkelig lille til at lade det vægtede gennemsnit være robust over for ændringer fra år til år og samtidig være tilstrækkeligt stort til at tage højde for tendenser over tid.

Vi beregner variansen af prior fordelingerne baseret på vægtede gennemsnits værdier af de kvadrerede residualer fra middelværdierne μ_{lj} :

$$\sigma_{lj}^2 = \psi_{lj} \sum_{t=1}^T \tau_t \left(\hat{a}_{lj|t}^* - \mu_{lj} \right)^2, \quad (13)$$

hvor ψ_{lj} er en tuning parameter der specificerer, hvor informativ prioren skal være, hvor en lille værdi angiver en meget informativ prior og en meget stor værdi indikerer en næsten ikke-informativ prior.

Vi forslår at bruge en log-normal fordeling til prior fordelingerne for input-output koefficienterne $\Psi(\mu_{lj}, \sigma_{lj})$. En fordel ved log-normal fordelingen er at det udelukker negative estimerter af input-output koefficienter. For at opnå en log-normal fordeling med forventet værdi μ_{lj} og varians σ_{lj}^2 sættes placeringsparametren og skalaparametren for log-normal fordelingen til:

$$\tilde{\mu}_{lj} = \ln \left(\frac{\mu_{lj}^2}{\sqrt{\mu_{lj}^2 + \sigma_{lj}^2}} \right) \quad (14)$$

$$\tilde{\sigma}_{lj}^2 = \ln \left(1 + \frac{\sigma_{lj}^2}{\mu_{lj}^2} \right), \quad (15)$$

således at:

$$\ln a_{lj}^* \sim \mathcal{N}(\tilde{\mu}_{lj}, \tilde{\sigma}_{lj}^2), \quad (16)$$

hvor $\mathcal{N}(\tilde{\mu}_{lj}, \tilde{\sigma}_{lj}^2)$ indikerer en normal fordeling med middel værdi $\tilde{\mu}_{lj}$ og variancen $\tilde{\sigma}_{lj}^2$.

3 Afsluttende bemærkninger

De økonometriske metoder som anvendes til estimering af input-output koeficienter og output-specifikke input mængder kan kategoriseres i: (1) klassisk økonometrisk metoder, (2) random coefficient regression, (3) entropi-baseret metoder, og (4) Bayesiansk økonometri.

Klassiske økonometriske metoder er meget restriktive og giver ofte upålidelige og usandsynlige estimerater, som er forårsaget af multikollinearitet og andre statistiske og økonometriske problemer.

Random coefficient regression metoder har den enorme fordel, at de slækker på den restriktive antagelse om ens input-output koeficienter på tværs af bedrifter, som kræves af andre økonometriske estimeringsmetoder. Men givet de yderligere koeficienter i en random coefficient regression med observerede faktorer, der påvirker input-output koeficienterne, kan problemer med multikollinearitet blive endnu værre sammenlignet med klassiske økonometriske metoder.

Entropi-baserede og Bayesianske metoder har mange ligheder, og det blev vist, at entropi-baserede metoder er et særtilfælde af Bayesianske metoder. Begge typer metoder kan bruges til at løse ulempene ved de klassiske økonometriske metoder, især ved at bruge forudgående viden til at løse problemer forårsaget af multikollinearitet. Den forudgående viden kan for eksempel komme fra faglitteraturen, eksisterende forskning eller undersøgelser. I tilfælde af årlige analyser kan priors være baseret på resultaterne opnået i de foregående år. Sammenlignet med entropi-baserede tilgange har Bayesianske tilgange den fordel, at de tager højde for priors på en mere gennemsigtig måde. Hverken eksisterende entropi-baserede metoder eller eksisterende Bayesianske metoder til at estimere input-output koeficienter slækker dog på den restriktive antagelse om lige input-output koeficienter på tværs af virksomheder.

Vi anbefaler på den baggrund, at estimere en random coefficient model med en Bayesiansk metode, da dette vil tillade for forskelle input-output koeficienter på tværs af bedrifter og for at bruge prior information for at opnå pålidelige og plausible estimerater på trods af høj multikollinearitet og et stort antal koeficienter at estimere.

Litteratur

Coelli, T. J., Prasada Rao, D., O'Donnell, C. J., and Battese, G. E. (2005). *Econometric Estimation of Production Technologies*. Springer.

- De Loecker, J., Goldberg, P. K., Khandelwal, A. K., and Pavcnik, N. (2016). Prices, markups, and trade reform. *Econometrica*, 84(2):445–510.
- Dixon, B. L., Batte, M. T., and Sonka, S. T. (1984). Random coefficients estimation of average total product costs for multiproduct firms. *Journal of Business & Economic Statistics*, 2(4):360–366.
- Dixon, B. L. and Hornbaker, R. H. (1992). Estimating the technology coefficients in linear programming models. *American Journal of Agricultural Economics*, 74(4):1029–1039.
- Errington, A. (1989). Estimating enterprise input-output coefficients from regional farm data. *Journal of Agricultural Economics*, 40:52–56.
- Fragoso, R. and Carvalho, M. L. d. S. (2013). Estimation of cost allocation coefficients at the farm level using an entropy approach. *Journal of Applied Statistics*, 40(9):1893–1906.
- Geweke, J. (1999). Using simulation methods for bayesian econometric models: inference, development, and communication. *Econometric reviews*, 18(1):1–73.
- Gocht, A. (2008). Estimating input allocation for farm supply models. Contributed paper presented at the 107th EAAE Seminar “Modelling of Agricultural and Rural Development Policies” in Sevilla, Spain, January 29th – February 1st, 2008.
- Griliches, Z. (1963). Estimates of the aggregate agricultural production function from cross-sectional data. *Journal of Farm Economics*, 45(2):419–428.
- Hallam, D., Bailey, A., Jones, P., and Errington, A. (1999). Estimating input use and production costs from farm survey panel data. *Journal of Agricultural Economics*, 50(3):440–449.
- Heckelei, T., Mittelhammer, R. C., and Jansson, T. (2008). A bayesian alternative to generalized cross entropy solutions for underdetermined econometric models. Technical report.
- Hornbaker, R. H., Dixon, B. L., and Sonka, S. T. (1989). Estimating production activity costs for multioutput firms with a random coefficient regression model. *American Journal of Agricultural Economics*, 71(1):167–177.
- Howitt, R. E. (1995). Positive mathematical programming. *American Journal of Agricultural Economics*, 77(2):329–342.

- Just, R. E., Zilberman, D., and Hochman, E. (1983). Estimation of multicrop production functions. *American Journal of Agricultural Economics*, 65(4):770–780.
- Just, R. E., Zilberman, D., Hochman, E., and Bar-Shira, Z. (1990). Input allocation in multicrop systems. *American Journal of Agricultural Economics*, 72(1):200–209.
- Lence, S. H. and Miller, D. J. (1998). Estimation of multi-output production functions with incomplete data: A generalised maximum entropy approach. *European Review of Agricultural Economics*, 25(2):188–209.
- Leon, Y., Peeters, L., Quinqu, M., and Surry, Y. (1999). The use of maximum entropy to estimate input-output coefficients from regional farm accounting data. *Journal of Agricultural Economics*, 50(3):425–439.
- Lips, M. (2014). Calculating full costs for Swiss dairy farms in the mountain region using a maximum entropy approach for joint-cost allocation. *International Journal of Agricultural Management*, 3(3).
- Louhichi, K., Espinosa, M., Ciaian, P., Perni, A., Ahmadi, B. V., Colen, L., and y Paloma, S. G. (2018). The eu-wide individual farm model for common agricultural policy analysis (ifm-cap v. 1): Economic impacts of cap greening. Technical report.
- Louhichi, K., Jacquet, F., and Butault, J. P. (2012). Estimating input allocation from heterogeneous data sources: A comparison of alternative estimation approaches. *Agricultural Economics Review*, 13(389-2016-23472):83–102.
- Midmore, P. (1990). Estimating input-output coefficients from regional farm data - a comment. *Journal of Agricultural Economics*, 41(1):108–111.
- Mittelhammer, R. C., Matulich, S. C., and Bushaw, D. (1981). On implicit forms of multiproduct-multifactor production functions. *American Journal of Agricultural Economics*, 63(1):164–168.
- Moxey, A. and Tiffin, R. (1994). Estimating linear production coefficients from farm business survey data: A note. *Journal of Agricultural Economics*, 45(3):381–385.
- Scandizzo, P. L. (1990). The estimation of input-output coefficients: Methods and problems. *Ricerche Economiche*, XLIV(4):455–474.
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Wikström, D., Peeters, L., and Surry, Y. R. (2011). Semiparametric cost allocation estimation. Paper presented at the XIIth International Congress of the European Association of Agricultural Economists (EAAE) in Zürich, Switzerland. <http://purl.umn.edu/115742>.

Zitthen, R. S., Henningsen, A., Andreasen, S. A., Frandsen, M., and Jürgensen, M. S. (2021). Estimating input allocation coefficients from aggregate firm-level data: A review of various econometric methods. *Department of Food and Resource Economics: Unpublished paper.*

Is stress regionally persistent?

Jørgen T. Lauridsen, Department of Economics, University of Southern Denmark,
jtl@sam.sdu.dk

United Nations Sustainable Development Goal 8:

Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all (<https://sdgs.un.org/>)

Abstract

Stress is known to be a widespread disease, which is caused by several reasons, and which is known to be intriguing to remedy or prevent. The present paper applies spatial regression based on spatially aggregated data for 98 Danish municipalities with repeated observation for the three years 2010, 2013 and 2017. The present paper confirms what is known by demonstrating that unhealthy lifestyle and social burden increase the local level of stress, that stress is intriguing to prevent, and that stress is increasing over time from 2010 to 2017. Furthermore, the paper adds new knowledge by showing a strong local persistency, in the sense that the local level of stress is increased by high stress level in the surrounding neighborhood. Apart from indicating persistency, such spatial patterns are furthermore indicative of an increasing geographical inequality in stress, in the sense that deprived neighborhoods with high stress levels are even further deprived. Complexity of stress as a persistent and growing problems is demonstrated, which leads to a policy recommendation of not only considering stress as a micro-level phenomenon between individuals and their micro-environment, but also consider implications of the macro-level in terms of regions and neighborhoods.

Keywords:

Stress; spatial spillover; spatio-temporal models

JEL classifications:

C13, C21, C23, I10, I12, I14

1. Introduction

The incidence and prevalence of work and life stress and derived diseases is increasing and among the major and increasing threats against health and quality of life. Thus, from being number 10 on the 2001 list of such threats, stress and related diseases is expected to jump to a second place by 2020 (WHO, 2001).

For the case of Denmark, the proportion with a high stress level rose from 2010 to 2017, except for males above 75 and females above 65 (DHA, 2018). In 2017, the proportion

was 29 percent for female and 21 percent for male, while the highest figures were 40.5 percent for females aged 16-24, 47 percent for unemployed, 55.6 percent for disability retired, and 54.6 percent for other groups outside the labor market (op. cit.).

Stress is obviously related to workplace conditions like lack of control over own work situation and workload, and it is further strengthened by a number of other negative features including job uncertainty; uncertainty about roles and responsibilities; burnout; lack of fairness and broadness; lack of collegial support, recognition and collaboration; lack of work and life balance; and negative experience alike sexual harassment, bullying, violence and unwanted attention (NFA, 2018).

However, stress is not only caused by workplace conditions. To the contrary, stress is well known to be related to several socioeconomic and lifestyle factors including unemployment, job loss and job uncertainty (Rugulies et al. 2010). Poor neighborhood conditions like criminality, highly trafficked roads, poor school conditions, lack of opportunities for leisure activities and good social interaction causes stress (Maggi et al. 2005). Lifestyle factors like alcohol consumption, smoking habits and drug abuse are known to be related to stress (Kriegbaum et al. 2011a-b, Kjøller and Rasmussen 2006) alike physical inactivity and obesity (NHA, 2018). Childhood poverty is known as a determinant for stress later in life (Klein et al. 2007, Yanos et al. 2007, Evans and Kim 2007). Lack of education is related to stress, as the proportion of stressed among those with only basic school is almost double the corresponding figure for high educated (DHA, 2018). Civil status matters, as married and cohabitated feel less stressed than divorced and widowed (DHA, 2018). Although stress is not in itself a disease, it is known to be strongly related to several diseases like hypertension (Liu et al. 2017), cardiovascular diseases (Kivimaki et al. 2012, Nabi et al. 2013, Rosengren et al. 2004), depression (Rugulies et al. 2006) and longstanding illness (NHA, 2018).

Although the preventive healthcare system should play an increasing active role, it is surprising to see indications of the opposite. Thus, for the case of Denmark, the proportion of stressed who consulted a GP and were advised to take it easy dropped slightly from 2010 to 2017 (NHA, 2018).

Given that several of the above conditions obviously relates to local environment and neighborhood, one should readily expect local environment and neighborhood characteristics to matter as sources for regional variation in stress. However, apart from simple descriptive studies merely just summarizing regional variation in work and life related stress, the author was not able to find studies of relevance. As a single exception, Wang et al. (2015) relates socioeconomic inequality to stress level for 21 Chinese cities in a multilevel study. For the case of Denmark, descriptive figures by 2017 showed that more

inhabitants of the regions of Zealand and Southern Denmark than of the Capitol Region were stressed, while the opposite were the case for the regions of Central Denmark and Northern Jutland (DHA, 2018).

2. Methods

The point of departure is a linear regression model defined for the $N=98$ Danish municipalities in a single year by

$$(1) \quad y_t = X_t \beta + v_t, \quad v_t \sim N(0, \sigma^2 I)$$

where X_t is an N by K dimensional matrix of the K explanatory variables, y_t an N dimensional vector of the stress rates in the municipalities, and β a K dimensional coefficient vector measuring the effects of the explanatory variables on the stress rate. The term v_t is a residual term, which represents the fertility rates when controlled for the explanatory factors of X_t , and may be denoted the residual stress rate.

Operationally, endogenous (learning) spatial spillover is controlled for by adding the average of y_t in the neighbourhood municipalities (denoted by y_t^W) as an explanatory variable in (1) to obtain the *spatially autoregressive* (SAR) specification (Anselin, 1988)

$$(2) \quad y_t = y_t^W \lambda + X_t \beta + v_t,$$

where λ is a parameter specifying the magnitude of spill-over, formally restricted to the interval between (-1) and (+1), but for most practical purposes restricted to be positive.

Alternatively, any kind of spatial clustering, including observed as well as unobserved exogenous spatial spillover, may be controlled for by applying the spatially autocorrelated (SAC) specification (Anselin 1988)

$$(3) \quad y_t = X_t \beta + \varepsilon_t, \quad \varepsilon_t = \lambda W \varepsilon_t + v_t.$$

The SAC approach will be applied to investigate whether spatial spillover of stress rates is merely ascribed to exogenous structures rather than being of an endogenous learning nature.

One further methodological problem needs attention for the SAR as well as the SAC specification. While pooled data for $T=2$ years are applied, the residual stress rates across years for any municipality are correlated. Also, the variance of the residual stress rate within each year may potentially vary across years. Thus, between any two years, the covariance of the residual stress rate reads as

$$(4) \quad E(v_t' v_s) = \sigma_{ts}^2 \quad t, s = 1, \dots, T.$$

To obtain efficient estimates of β , we apply Feasible Generalised Least Squares (F-GLS) estimation as suggested by Zellner (1962) to obtain Seemingly Unrelated Regression (SUR) estimates for β . By integrating (4) into any of (1) through (3), SUR, SAR-SUR and SAC-SUR specifications are obtained.

3. Data

Table 1 provides an overview of the data used for the study. Data are collected for 98 municipalities for the years 2010, 2013 and 2017 from three different sources, The Danish Health Profile, The Danish Ministry of Taxation, and the Key Figure Base at The Ministry of the Interior. The outcome variable is percentage feeling highly stressed, defined as belonging to the upper 20 percent on the Cohen's Perceived Stress Scale (PSS) with 10 items (Cohen et al. 1983, Eskildsen et al. 1015). According to theory, literature and availability, a variety of explanatory variables affecting stress were included. As an indication of change over time, dummies for 2013 and 2017 are defined. Health behaviour includes physical activity, alcohol, unhealthy food habits and obesity, while health is captured by Self-Assessed Health (SAH), ill physical health and ill mental health. Contact to primary healthcare is measured by percentage reporting been in contact with a general practitioner (GP) within last 12 months. Social relations are represented by one variable (being unwanted alone), and age structure by percentages of 25-64 and 65+ year. Quality of workplace and control over own work situation is measured by educational level. Income level is measured by tax deductible income per inhabitant. Finally, presence or absence of social distress and neighbourhood deprivation is captured by a variety of variables (urbanisation, social housing, unemployment, non-Western inhabitants, criminal activity, social benefit receivers, and peripheral area).

Table 1. Description of data.

Variable	Description	Mean	SD
Stress ¹	% feeling highly stressed (upper 20% on the PSS-10 scale)	13.29	2.76
PhysAct ¹	% reporting being physically active in leisure time	27.19	3.30
DailySmoker ¹	Percentage reporting being daily smokers	18.71	3.58
Alcohol ¹	% reporting drinking too much (21/14 glasses weekly for M/F)	8.47	2.33
UheFood ¹	% reporting eating unhealthy food daily	15.08	3.77
Obese ¹	% reporting being obese (BMI>30)	51.22	5.66
SAH ¹	% reporting good / very good Self Assessed Health	83.85	2.80
IllPhysHealth ¹	% reporting ill physical health	11.39	2.45
IllMentalHealth ¹	% reporting ill mental health	10.68	2.18
GP	% reporting been in contact with GP within last 12 months		
UnwantedAlone ¹	% reporting being unwanted alone daily	5.62	1.09
P25_64 ³	% population between 25 and 64	51.19	1.88
P_65 ³	% population 65 and above	19.69	3.87
Educ ³	% of population with higher education	24.48	8.76
Income ³	Tax deductible income per inhabitant (100,000 DKK, 2010 level)	1.72	0.34
Urban ³	% population living in urban area	83.22	12.77
SocHous ³	% of population living in social housing	18.05	12.01
Unemp ³	% of population unemployed	4.01	1.12
NonWest ³	% of population being from non-Western country	3.10	1.92
SiCrime ³	Simple crimes reported per 1,000 inhabitants	48.48	19.82
ViCrime ³	Violence crimes reported per 1,000 inhabitants	1.57	0.62
SocBen ³	% population receiving social benefits	4.12	1.25
Peripheral ²	Indicator for peripheral municipality	0.15	0.36
year13	Indicator for year 2013 (reference year 2010)	0.33	0.47
year17	Indicator for year 2017 (reference year 2010)	0.33	0.47

Source: ¹ The Danish Health Profile (www.danskernesundhed.dk), ² The Danish Ministry of Taxation (www.skm.dk), and ³ the Key Figure Base (www.im.dk)

4. Results

Table 2 shows results from two different spatial models. First and foremost, the significantly positive spillover parameters demonstrates a strong local persistency, in the sense that the local level of stress is increased by high stress level in the surrounding neighborhood. Apart from indicating persistency, such spatial patterns are furthermore indicative of an increasing geographical inequality in stress, in the sense that deprived neighborhoods with high stress levels are even further deprived.

Table 2. Estimated spatial models

Variable	SUR			SAC-SUR			SAR-SUR		
	Coef	Stderr	P	Coef	Stderr	P	Coef	Stderr	P
Constant	18.037	0.041	0.04	17.902	8.731	0.04	19.257	8.564	0.02
PhysAct	0.035	0.433	0.43	0.020	0.041	0.62	0.03	0.043	0.49
DailySmoker	0.192	<0.001	<0.01	0.169	0.046	<0.01	0.185	0.046	<0.01
Alcohol	0.132	0.009	0.01	0.096	0.052	0.06	0.136	0.050	0.01
UheFood	0.002	0.949	0.94	-0.015	0.037	0.68	0.026	0.035	0.46
Obese	0.005	0.872	0.87	0.003	0.034	0.92	0.013	0.032	0.68
SAH	-0.145	0.042	0.04	-0.147	0.067	0.02	-0.134	0.069	0.05
IllPhysHealth	-0.131	0.086	0.08	-0.121	0.073	0.09	-0.125	0.074	0.09
IllMentalHealth	0.592	<0.001	<0.01	0.567	0.057	<0.01	0.577	0.060	<0.01
GP	-0.013	0.699	0.69	-0.015	0.035	0.67	-0.013	0.033	0.69
UnwantedAlone	0.210	0.010	0.01	0.182	0.078	0.02	0.192	0.080	0.02
P25_64	-0.103	0.171	0.17	-0.057	0.078	0.46	-0.178	0.075	0.02
P_65	-0.209	<0.001	<0.01	-0.192	0.051	<0.01	-0.239	0.049	<0.01
Educ	0.017	0.494	0.49	0.015	0.025	0.55	0.028	0.024	0.24
Income	1.186	0.017	0.01	0.960	0.504	0.05	1.063	0.484	0.03
Urban	-0.002	0.878	0.87	0.001	0.011	0.93	-0.008	0.011	0.48
SocHous	0.009	0.358	0.35	0.015	0.010	0.13	0.011	0.009	0.25
Unemp	0.256	0.007	0.01	0.221	0.100	0.02	0.253	0.093	0.01
NonWest	-0.076	0.910	0.91	-0.042	0.713	0.95	-0.266	0.659	0.69
SiCrime	0.002	0.715	0.71	0.002	0.006	0.72	0.004	0.006	0.55
ViCrime	0.333	0.010	0.01	0.379	0.125	<0.01	0.408	0.127	<0.01
SocBen	0.029	0.785	0.78	-0.034	0.105	0.74	0.003	0.102	0.97
Peripheral	-0.415	0.061	0.06	-0.364	0.212	0.08	-0.459	0.218	0.03
year13	1.422	<0.001	<0.01	1.405	0.376	<0.01	1.286	0.328	<0.01
year17	3.583	<0.001	<0.01	3.62	0.574	<0.01	2.771	0.586	<0.01
Spillover				0.359	0.123	<0.01	0.195	0.044	<0.01
LogL	-144.0			-133.2			-134.3		

Furthermore, the results confirm – with small variations across specifications – what is known by demonstrating that unhealthy lifestyle and social burden increase the local level of stress, that stress is intriguing to prevent, and that stress is increasing over time from 2010 to 2017. Complexity of stress as a persistent and growing problems is thus demonstrated, which leads to a policy recommendation of not only considering stress as a micro-level phenomenon between individuals and their micro-environment, but also consider implications of the macro-level in terms of regions and neighborhoods.

5. Conclusion

The study confirms what is known by demonstrating that unhealthy lifestyle and social burden increase the local level of stress, that stress is intriguing to prevent, and that stress is increasing over time from 2010 to 2017. Furthermore, the paper adds new knowledge by showing a strong local persistency, in the sense that the local level of stress is increased by high stress level in the surrounding neighborhood. Apart from indicating persistency, such spatial patterns are furthermore indicative of an increasing geographical inequality in stress, in the sense that deprived neighborhoods with high stress levels are even further deprived. Complexity of stress as a persistent and growing problems is demonstrated, which leads to a policy recommendation of not only considering stress as a micro-level phenomenon between individuals and their micro-environment, but also consider implications of the macro-level in terms of regions and neighborhoods.

References

- Anselin L. 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic, Dordrecht
- Kjøller M, Rasmussen NK. 2006. *Sundhed og sygelighed i Danmark 2006 og udviklingen siden 1987* [Danish Health and Morbidity Survey 2005 and trends since 1987]. Copenhagen: National Institute of Public Health.
- Maggi S, Irwin LG, Siddiqi A, Poureslami I, Hertzman E, Hertzman C. 2005. *Knowledge Network for Early Child Development. Analytic and Strategic Review Paper: International Perspectives on Early Child Development*. World Health Organization's Commission on the Social Determinants of Health.
- Klein H, Elifson KW, Sterk CE. 2007. Childhood neglect and adulthood involvement in HIV-related risk behaviors. *Child Abuse Negl* 31(1):39-53.
- Yanos PT, Czaja SJ, Widom CS. 2010. A prospective examination of service use by abused and neglected children followed up into adulthood. *Psychiatr Serv* 61(8):796-802.
- Evans GW, Kim P. 2007. Childhood poverty and health: cumulative risk exposure and stress dysregulation. *Psychol Sci* 18(11):953-957.
- Rugulies R, Thielen K, Nygaard E, Diderichsen F. 2010. Job insecurity and the use of antidepressant medication among Danish employees with and without a history of prolonged unemployment: a 3.5-year follow-up study. *J Epidemiol Community Health* 64(1):75-81.
- Kriegbaum M, Christensen U, Osler M, Lund R. 2011a. Excessive drinking and history of unemployment and cohabitation in Danish men born in 1953. *Eur J Public Health* 21(4):444-8 <https://doi.org/10.1093/eurpub/ckq152>.
- Kriegbaum M, Larsen AM, Christensen U, Lund R, Osler M. 2011b. Reduced probability of smoking cessation in men with increasing number of job losses and partnership breakdowns. *J Epidemiol Community Health* 65(6):511-6, <https://doi.org/10.1136/jech.2009.100446>.
- DHA (The Danish Health Authority). 2018. *Danskernes Sundhed – Den Nationale Sundhedsprofil 2017*. Copenhagen: Danish Health Authority,

- <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=2ahUKEwiNg-IxSg6vnAhUx-uqQKHY0NBQsQFjABegQIBhAB&url=https%3A%2F%2Fwww.sst.dk%2Fd%2Fudgivelser%2F2018%2F~%2Fmedia%2F73EADC242CDB46BD8ABF9DE895A6132C.ashx&usg=AOvVaw33zxXFg-SlcH9dixAITGpAp>
- Liu MY, Li N, Li WA, Khan H. 2017. Association between psychosocial stress and hypertension: a systematic review and metaanalysis. *Neurol Res* 39: 573-580.
- Kivimaki M, Nyberg ST, Batty GD, Fransson EI, Heikkila K, Alfredsson L, et al. 2012. Job strain as a risk factor for coronary heart disease: a collaborative meta-analysis of individual participant data. *Lancet* 380: 1491-7.
- Nabi H, Kivimaki M, Batty GD, Shipley MJ, Britton A, Brunner EJ, et al. 2013. Increased risk of coronary heart disease among individuals reporting adverse impact of stress on their health: the Whitehall II prospective cohort study. *Eur Heart J* 34: 2697-705.
- Rosengren A, Hawken S, Ounpuu S, Sliwa K, Zubaid M, Almahmeed WA, et al. 2004. Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13 648 controls from 52 countries (the INTERHEART study): casecontrol study. *Lancet* 364: 953-62.
- Rugulies R, Bultmann U, Aust B, Burr H. 2006. Psychosocial work environment and incidence of severe depressive symptoms: Prospective findings from a 5-year follow-up of the Danish work environment cohort study. *Am J Epidemiol* 163: 877-87.
- Cohen S, Kamarck T, Mermelstein R. 1983. A global measure of perceived stress. *J Health Soc Behav* 24: 385-96.
- Eskildsen A, Dalgaard VL, Nielsen KJ, Andersen JH, Zachariae R, Olsen LR, et al. 2015. Cross-cultural adaptation and validation of the Danish consensus version of the 10-item Perceived Stress Scale. *Scand J Work Env Health* 41: 486-90.
- WHO. 2001. *Mental Health - A Call for Action by World Health Ministers*. Geneva: WHO. <https://www.mhinnovation.net/resources/mental-health-call-action-world-health-ministers>.
- NFA - The National Research Center for the Working Environment. 2018. *Fakta om Arbejdsmiljø og Helbred 2018*. Copenhagen: NFA, <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=2ahUKEwjruclOn6vnAhWFoFwKHZTIDm4QFjABegQIBBAB&url=https%3A%2F%2Fnfa.dk%2F%2Fmedia%2FNFA%2FArbejdsmiljodata%2FFakta-om-Arbejdsmiljo-og-Helbred-2018.ashx%3Fla%3Dda&usg=AOvVaw1VK33REKD0fz3pmxwPloqW>.
- Wang H, Yang XY, Yang T, Cottrell RR, Yu L, Feng Y, JiangS. 2015. Socioeconomic inequalities and mental stress in individual and regional level: a twenty one cities study in China. *International Journal for Equity in Health* 14:25, <https://doi.org/10.1186/s12939-015-0152-4>.
- Zellner A. 1962. An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias. *Journal of the American Statistical Association* 58: 977-992.

Anders Milhøj

anders.milhøj@econ.ku.dk

Blev vi danskere mere syndige under coronaen?

Abstrakt

Under coronaen var mange hjemsendt uden muligheder for et normalt festligt dansk liv. Udlandsrejser var ligeledes begrænset til kun de mest nødvendige. Spørgsmålet er, hvad vi så brugte tiden til hjemme eller i fx. sommerhuset? I dette indlæg diskuteres, om tiden blev brugt til druk, røg og sex.

Metode

Tidsrækker med trend

I tidsrækker med en synlig trend anvendes i dette indlæg eksponentiel udglatning med Proc ESM i SAS. I denne procedure anvendes Winters metode (også kaldet Holt-Winters metode). Denne metode går ud på at dekomponere tidsrækken i tre komponenter:

En niveau komponent, der afspejler tidsrækvens niveau.

En trend komponent, der varierer omkring tidsrækvens hældning, der tillades at variere over tid.

En sæson komponent, der også tillades at variere

De tre komponenter ganges sammen for at få udglattede værdier af tidsrækken. Denne multiplikation giver mulighed for en relativ sæsoneffekt, der ofte ses i tidsrækker med en synlig trend. Metoden kan dog også anvendes i en additiv version, hvis fx sæsoneffekten ser ud til at være af konstant størrelsesorden.

Winters metode indeholder tre tuningsparametre, en for hver komponent, der afspejler i hvor høj grad de tre komponenter varierer over tid. De tre komponenter estimeres med en algoritme, der successivt udglatter en fortidig værdi med den aktuelle værdi, således at der på ethvert tidspunkt opnås en værdi af hver komponent, der indeholder al information fra fortiden til nu. Disse aktuelle værdier af de tre komponenter benyttes så til at forudsige fremtiden ud fra de sidst opnåede værdier.

Tidsrækker uden trend

Sæsonudjævningsmetoden Census X11 er baseret på samme grundidé som eksponentiel udglatning, men estimationsmetoden er anderledes. I Proc X11, der ligesom Proc X12

og X13 i SAS, indeholder X11-metoden, estimeres komponenterne ved en udglatning på begge sider af en observation, hvorved der opnås en mere sikker bestemmelse af komponenterne. Desuden indeholder X11-metoden er meget grundig undersøgelse for outliers, der i høj grad kompenseres for i beregningerne, hvilket også konsoliderer de fundne komponenter.

De fundne komponenter er ikke i lige så høj grad som i Proc ESM tiltænkt prædiktion, men det kan gøres 'manuelt' ved at kombinere de sidste relevante estimerede komponentværdier ud i fremtiden. Da metoden ikke danner en separat trend, er denne metode bedst til tidsrækker uden trend og ellers kun til kortsigtede prædiktioner.

Kønssygdomme

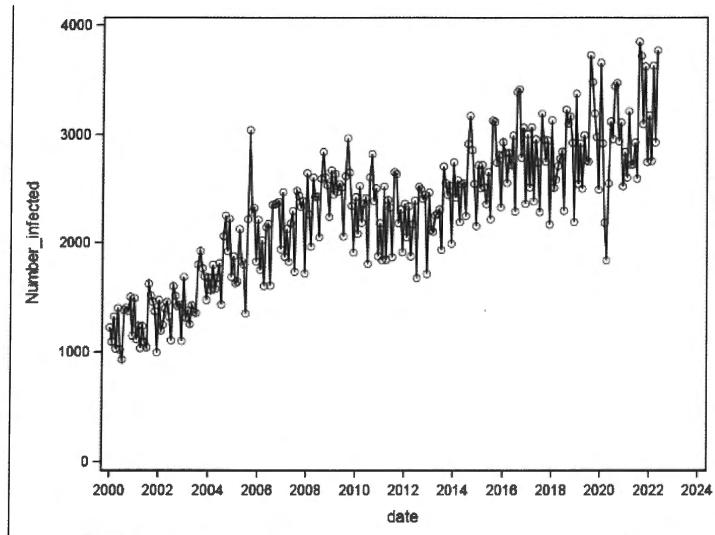
Klamydia

På pressekonferencen 20. april 2020 kom Søren Brostrøm med den markante udtalelse 'Sex er godt. Sex er sundt. Sundhedsstyrelsen går ind for sex. Vi er seksuelle væsener, og selvfølgelig skal man dyrke sex også som single i en tid med corona'. Men inden da udtalte Kåre Mølbak sig også. Det oprindelige spørøsmål fra journalisten forstod jeg som et spørøsmål om, hvordan enlige kunne finde nye partnere, når nu alt natteliv var lukket ned. Kåre Mølbak svarede, at det var et problem at finde nye kærester, hvilket fx. kunne ses ved, at klamydiaforekomsten var dalet, hvorefter Søren Brostrøm kom med sine meget citerede bemærkninger.

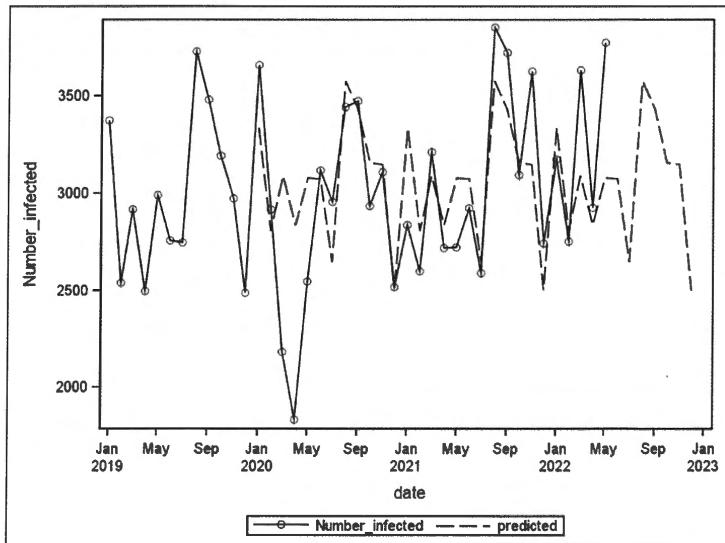
Statens Serum Institut offentliggør regelmæssigt månedlige tal for antal positive laboratorieprøver for klamydia. Der er et marked for hjemmetest for klamydia, men da behandling af faktisk smitte kræver lægebesøg, må det forventes, at lægen også tager en prøve, hvor et positivt svar indberettes. Figur 1 viser det månedlige antal positive tests fra og med januar 2000 til og med maj 2022. Det overordnede billede er en støt stigning i hele perioden dog med en stagnation eller et mindre fald under finanskrisen.

Figur 2 viser det faktiske antal smittede samt en forudsigelse af antallet givet ud fra den omtalte metode med at fremskrive ud fra data til og med december 2019 og dernæst bruge den estimerede trendcykliske komponent for december 2019 med den estimerede sæsonkomponent for 2019 lagt til. Det ses tydeligt, at Kåre Mølbak havde ret i sin udtalelse, for tallene for marts, april og maj 2020 var langt lavere end forudsigelsen. Men allerede i juni var de faktiske antal igen oppe på niveau med forudsigelserne - ja endda lidt højere.

Figur 1 Månedlige antal positive klamydiatests

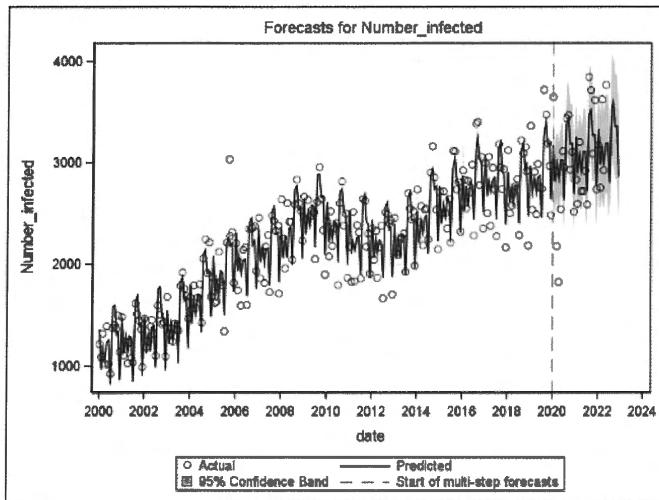


Figur 2 Faktiske antal klamydiatilfælde med en fremskrivning ud fra data til og med december 2019



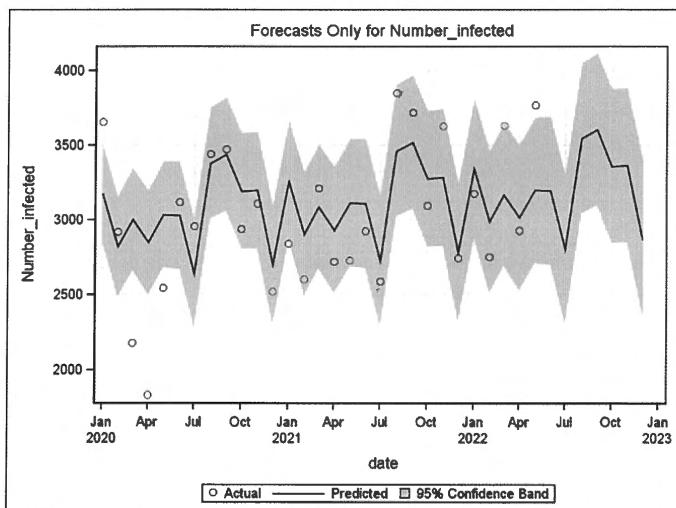
At de faktiske antal smittede er stigende ses ved at benytte en forudsigtelse, hvor en trend er inkorporeret, jf metodeafsnittet. Figur 3 viser resultatet af en kørsel med Proc ESM, Winters metode anvendes i sin additive form, da sæsonudsvingene ser konstante ud.

Figur 3 Forudsigelse af antal klamydiatilfælde fra og med januar 2020 med en estimeret opadgående trend



I Figur 4 zoomes ind på månederne fra og med januar 2020. Stigningen ses selvfølgelig ikke så tydeligt på denne figur, men det bemærkes, at antal smittede i januar 2020 var højere end forventet og at antallene i marts, april og til dels maj 2020 var langt lavere end forventet.

Figur 4 Forudsigelse af antal klamydiatilfælde fra og med januar 2020 med en estimeret opadgående trend

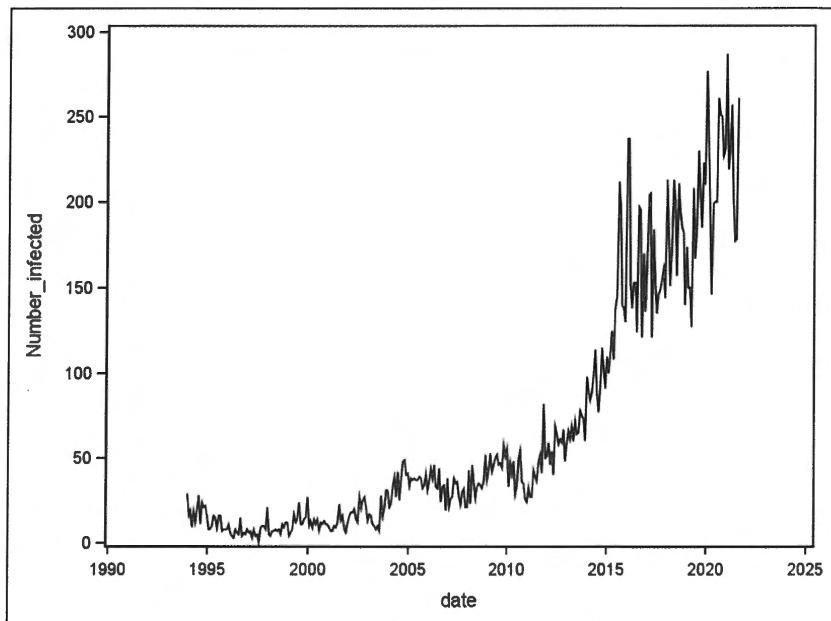


Trendkomponenten er ca 7, dvs. en støt stigning på 7 smittede pr. måned, hvilket svarer til ca 84 pr. år, hvilket kan sammenholdes med, at antallet de senere år har været omkring 3000 på månedsbasis og i alt 36633 smittede i hele 2021. Selvom denne trend ikke er stor, giver denne metode en bedre forudsigelse af det stigende antal smittede fra og med sommeren 2020.

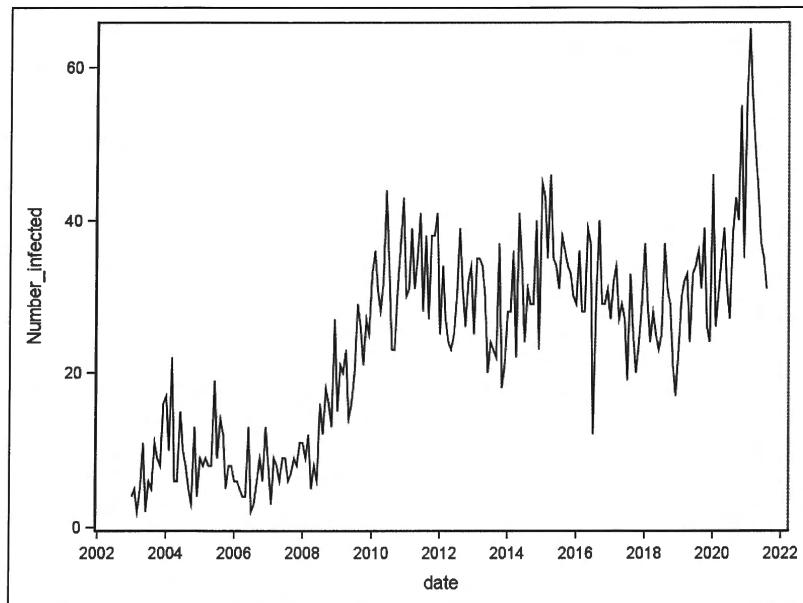
Gonorré og Syphilis

For disse sygdomme har jeg desværre kun kunnet skaffe data til og med august 2021 på Seruminstittutts hjemmeside. Det skyldes, at metoden for indberetningen fra lægerne af disse indberetningspligtige sygdomme er omlagt fra papirblanketter til en elektronisk indberetning, hvorved den automatiske publicering på hjemmesiden blev forpurret. For begge sygdomme gælder, at antal smittede er steget voldsomt siden de første observationer i midten af halvfemserne. For gonorré fra et niveau under 10 i 1990-erne til nu omkring 250 pr. måned; se Figur 5. For syphilis fra et niveau også under 10 til nu omkring 40 pr. måned; se Figur 6.

Figur 5 Antal gonorrétilfælde fra og med januar 1994 til og med august 2021.



Figur 6 Antal syfilistilfælde fra og med januar 2003 til og med august 2021.



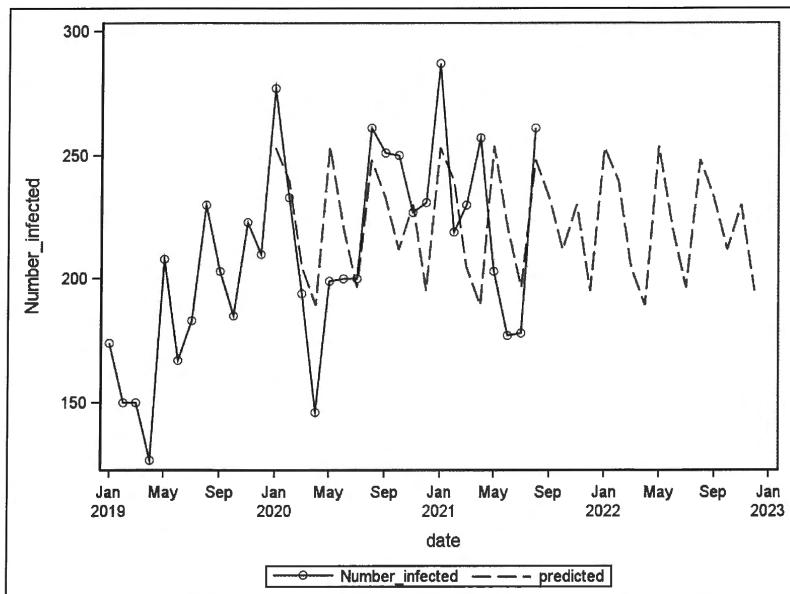
Dette voldsomme niveauskift gør det nødvendigt at benytte multiplikative metoder, svarende til at logaritmetransformere, hvilket betyder, at metoderne ikke tillader observationer på nul smittede i en måned. Det betyder, at tidsrækkerne med henblik på analyserne er skåret af ved årsskiftet efter det seneste observerede nul. For gonorré er der brugt data fra og januar 1994 og for syphilis er der brugt data fra og med januar 2003.

Figur 7 og 8 viser de faktiske observationer for 2020 og 2021 sammenholdt med forudsigelser dannet ud fra sæsondekompliceret med X11-metoden ved brug af data frem til og med december 2019. For gonorré gælder, at der i april og maj 2020 var væsentlig færre tilfælde end forventet, hvilket helt svarer til observationerne for klamydia. For syphilis ses derimod ikke nogen effekt af coronanedlukningerne i foråret 2020.

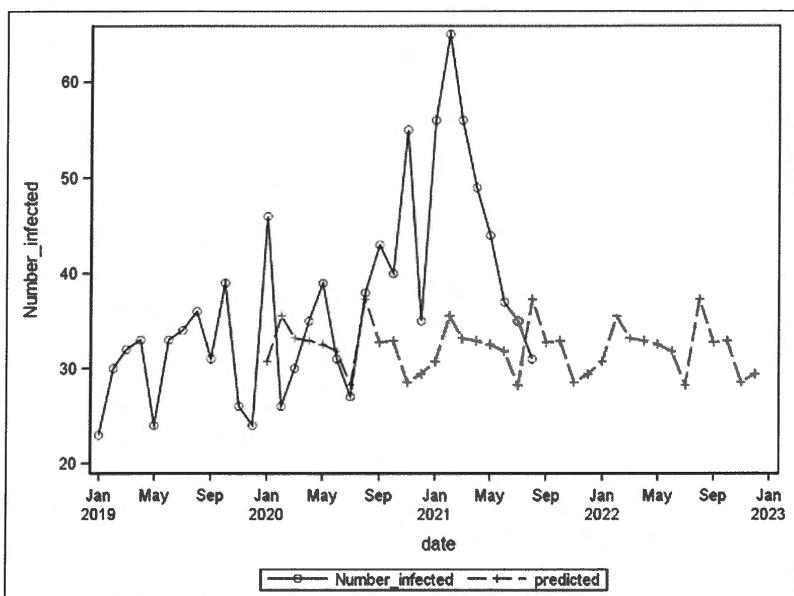
Begge sygdomme viser en forøgelse af smitteniveauet i efteråret 2021 og ligeledes for de måneder i 2021, der er data for. Det virker altså som om den stigning, som blev set før coronaen er fortsat efter kun en midlertidig nedgang, eller stagnation, i coronaens første tid. Det ses også ved brug af en metode, Winters metode, se Figur 9 for gonorré og Figur 10 for syphilis.

For syphilis er det markant, at antal tilfælde i forår og forsommert 2021 er langt højere end forventet. Denne højere prævalens i Danmark kan skyldes at færre pådrog sig syphilis i udlandet, fordi mange populære rejsemål var mere eller mindre utilgængelige.

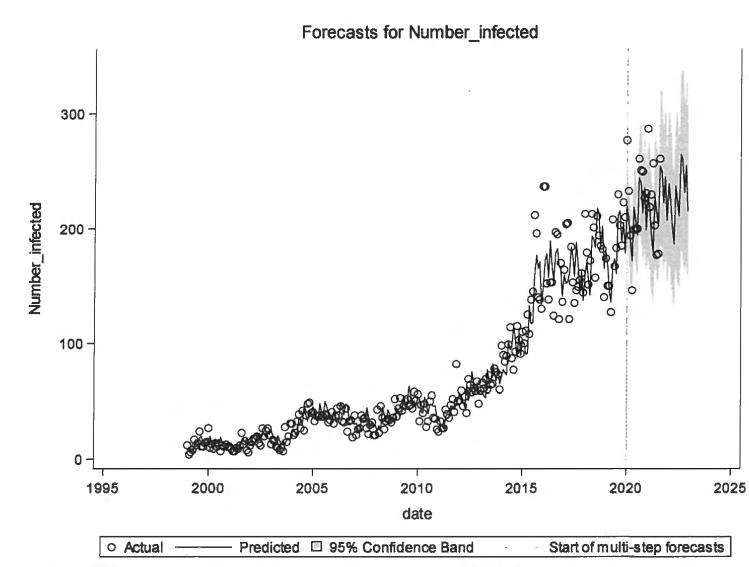
Figur 7 Forudsigelse af antal gonorrétilfælde fra og med januar 2020



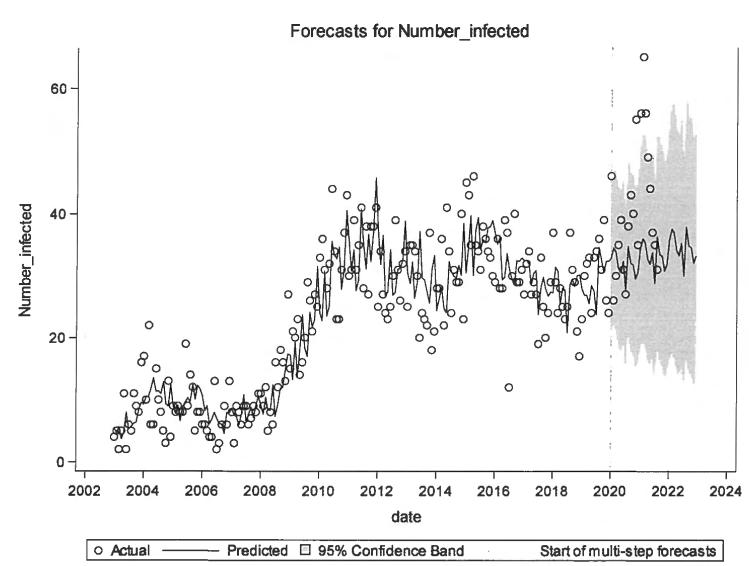
Figur 8 Forudsigelse af antal syfilistilfælde fra og med januar 2020



Figur 9 ESM Forudsigelse af antal gonorrétilfælde fra og med januar 2000 med en estimeret opadgående trend; forudsigelser fra og med 2020



Figur 10 ESM Forudsigelse af antal syfilistilfælde fra og med januar 1998 med en estimeret opadgående trend; forudsigelser fra og med 2020



Selvom der er tale om en væsentlig stigning i prævalensen af de to kønssygdomme siden 1990-erne er niveaueret i dag langt lavere end tidligere. I Peter Toft-Nielsen (1978) gen gives mange data fra Medicinalberetningerne i tresserne og halvfjerdserne; i dette indlæg vises blot i Tabel 1 de årlige antal registrerede tilfælde i årene 1901, 1938 og 1974 sammenholdt med de årlige antal tilfælde i 1997 og 2020, der er det sidste år med fuldstændige data.

Tabel 1 Årlige antal tilfælde af gonorré og syphilis

Sygdom	1901	1938	1974	1997	2020
Gonorré	6.634	9.276	14.915	78	2.669
Syphilis	2.561	470	460	17	445

Det er påfaldende, at gonorré er langt mindre udbredt nu end for 50 -100 år siden, mens forekomsten af syphilis ikke længere er tilbage til niveaueret før nedgangen, i firserne og halvfemserne, der sandsynligvis skyldtes AIDS epidemien. Prævalensen af både gonorré og syphilis er heldigvis langt lavere end i begyndelsen af 1900-tallet.

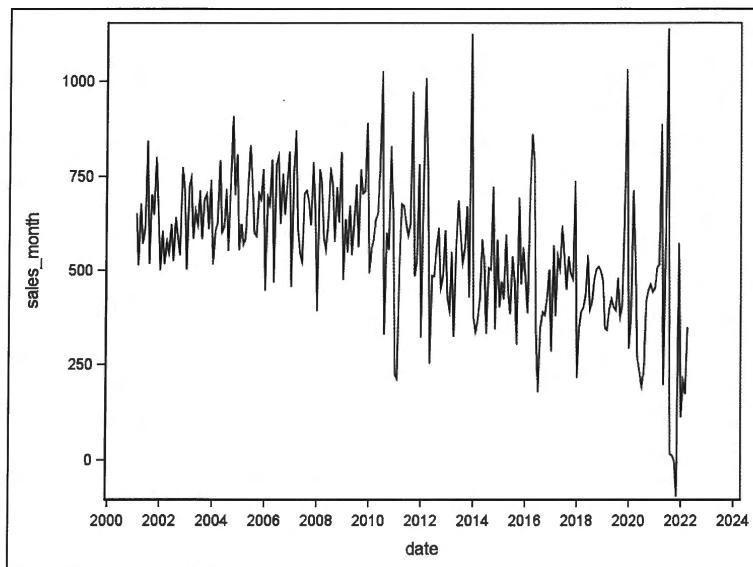
Cigaretter

Salget af cigaretter er generelt faldende i disse år, da rygning fra politisk side ønskes reduceret mest muligt, se Figur 11 for udviklingen siden 2001. Tallene dækker altså over cigaretter solgt med dansk beskatning, men da afgiften betales lidt uregelmæssigt og endda kan tilbagebetales kan der være store forskydninger i salget fra måned til måned. Da der pr 1. januar 2020 og igen 1. april 2022 var afgiftsforhøjelser med hamstring til følge både privat og i detailhandlen (afgiften betales ved levering til detailhandlen) er problemet ekstra stort i den interessant coronaperiode. Der forekommer endda måneder med negativ afgiftsindbetaling så multiplikative metoder kan ikke anvendes.

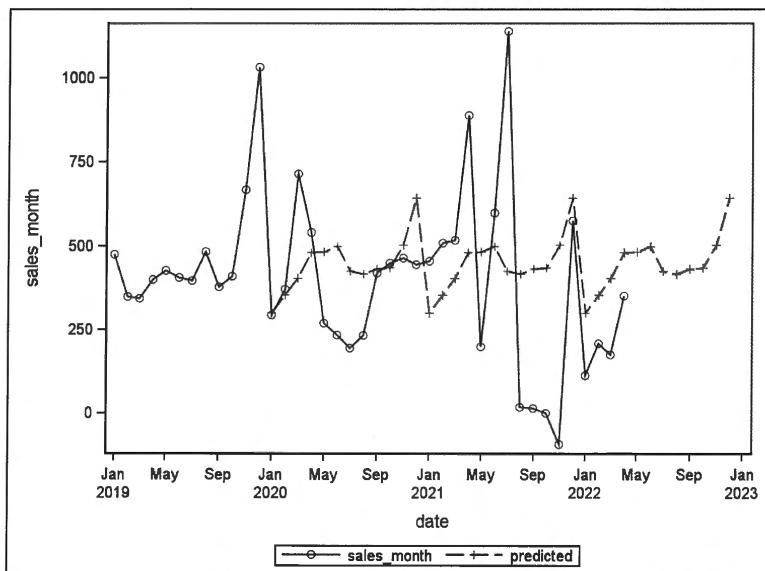
Figur 12 viser forudsigelserne dannet ud fra sæsondekomponeringen med X11-metiden i dens multiplikative form af salget fra og med januar 2020 sammenholdt med de faktiske observationer. Hamstringen op til afgiftsforhøjelsen januar 2020 ses tydeligt i form af en markant top i december 2019, så det lavere salg end forventet i marts april 2020 kan skyldes brug af hamstringslagre.

I de første måneder af 2021 var salget højere end forventet, hvilket kan skyldes at forbruget af udenlandske cigaretter enten via 'privat import' eller fordi færre danskere kunne rejste ud af landet. Bunden i maj 2021 opfattes i denne forbindelse som negativ autokorrelation, der modvirkes af toppene før og efter. Men fra sommeren 2021 har slaget værre lavere end forventet på nær julefrokosttrygningen, er var som forventet. Mulighederne for 'hamstring' i detailledet er indskrænket siden 2020, så afgiftsforhøjelsen april 2022 ser ikke ud fra de sidste observationer til at have medført hamstring.

Figur 11 Cigarettsalget fra og med januar 2001 til og med april 2022



Figur 12 Forudsigelser af cigarettesalget fra og med januar 2020

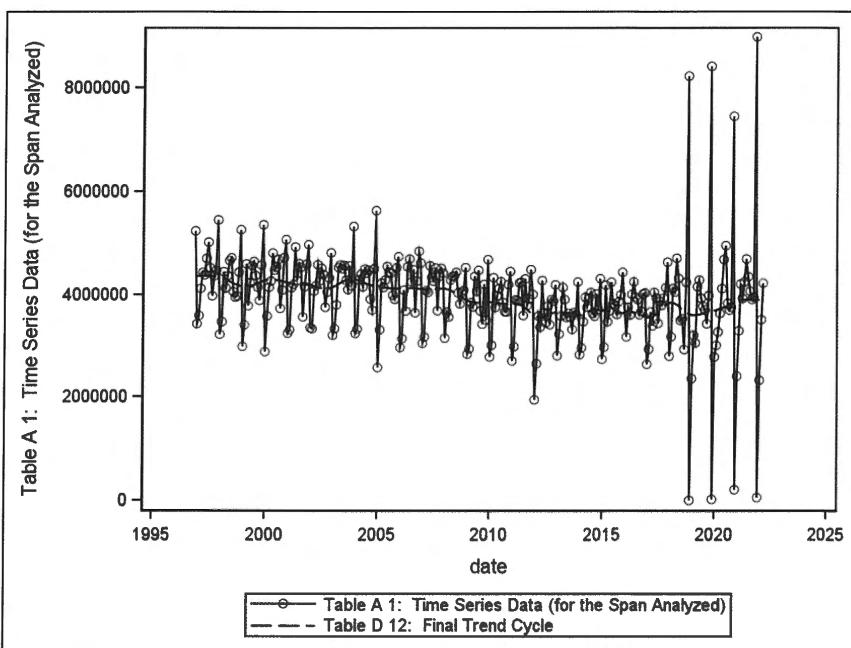


Alkohol

Figur 13 viser det totale salg af alkohol (mål i liter 100% alkohol for at sammenlægge bidragene fra øl, vin og spiritus omregnet til 100% alkohol) for alle måneder fra januar 1996 til april 2022. Disse tal er som cigaretforbruget beregnet ud fra afgiftsindbetalingerne, så de er også behæftet med usikkerheder. På figuren er der også indtegnet en udglattet kurve for de fire tidsrækker, beregnet så eventuelle sæsoneffekter og outliere er fjernet ved hjælp af X11-metoden.

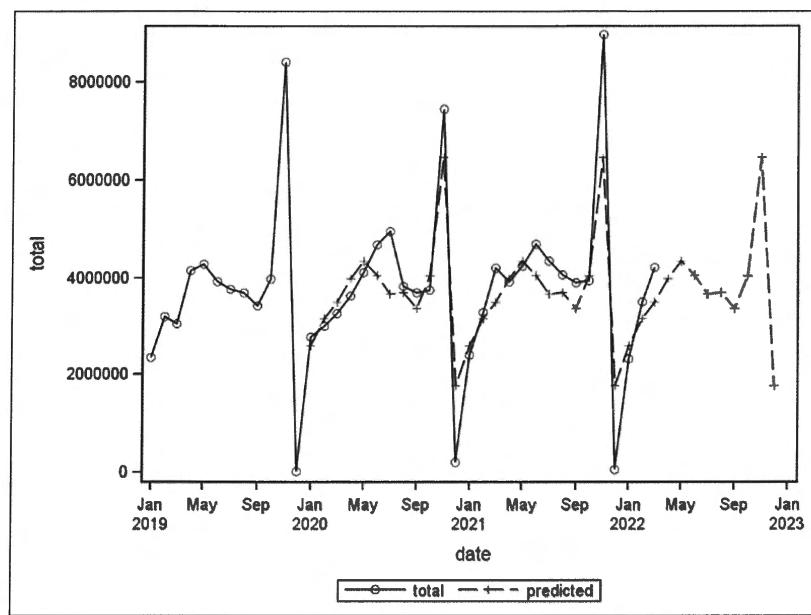
Især er det bemærkelsesværdigt, at sæsoneffekten omkring nytår er meget markante siden årsskiftet 2016/17, så afgiften tilsyneladende indbetales i december og ikke i januar. Det gør alle beregninger endnu mere usikre.

Figur 13 Det totale salg af alkohol for alle måneder fra 1996 til og med april 2022 med en kurve for niveauet indtegnet



Figur 14 viser igen det totale alkoholsalg, idet der dog er zoomet ind på årene fra og med 2019. På figuren er der også indtegnet forudsigelserne for salget beregnet ud fra dekompositionen med X11-metoden. Der ses ikke noget øget alkoholsalg under corona-nedlukningerne i foråret 2020, men der ses et tydeligt mersalg i sommeren juni 2020 og igen i sommeren 2021.

Figur 14 Det totale alkoholsalg med forudsigelser fra og med januar 2020 indtegnet



Referencer

Peter Toft-Nielsen (1978) Tidsrækkeanalyse af gonorréens epidemiologi, Komiteen for Sundhedsoplysning.

Anders Milhøj har haft indlæg om danskernes alkoholforbrug og markedsandele for øl, vin og spiritus i mange års udgaver af "Symposium i Anvendt Statistik"-serien - det ene sammen med Lisbeth la Cour. Tilsvarende var der et indlæg om cigaretsalg ved symposiet i 2019.

PTSD in school-age children: A nationwide prospective birth cohort study

Mogens Nygaard Christoffersen and Anne Amalie Elgaard Thorup

(Revision 10th December 2021)

Abstract

Traumatic childhood events are one of the few identifiable and to some extent preventable causes of psychiatric illness. Children exposed to severely stressful events may react with PTSD and this may impact their daily life level of function, their future development and mental health.

The traumatic stress model suggests that child maltreatment, community violence, accidents, and other traumas are regarded as an additive environmental factor, which can over-weight protective, compensatory factors and interact with predisposition for complex developmental disorders.

The study is based on prospective panel data including the whole population of children born in Denmark from 1984 to 1994 followed from age 7 to age 18 (N=679,000). Risk factors (child maltreatment, community violence, individual vulnerabilities) and compensatory factors (social support) are analyzed by the discrete time-Cox-model.

We found a lifetime prevalence in the range of 2.3 % PTSD in school-age children (n=15,636). Child maltreatment, victims of violent crime, parental violence and individual vulnerabilities especially autism (OR 7.08) and ADHD (OR 10.7) were predicative of PTSD. Our results were consistent with the traumatic stress model. Substance abuse was associated with PTSD.

Many potentially traumatic events were not reported in hospital or administrative records, consequently, PTSD may be underestimated in this study.

Introduction

Traumatic childhood events are one of the few identifiable causes of psychiatric illness (Kerns, Newschaffer, & Berkowitz, 2015). When children and adolescents are exposed to traumatic events considered as extreme life stressors outside the range of normal human experience, the societal, interpersonal, and psychological consequences can be considerable (Keane & Barlow, 2004; Koss, Koss, & Woodruff, 1991; Kulka et al., 1990; Toth, Gravener-Davis, Guild, & Cicchetti, 2013). Post-traumatic stress disorder (PTSD) in school-age children acquire that children are exposed to or witness a stressor that the individual perceives as threatening to physical and/or psychological integrity of self (Blacker, Frye, Morava, Kozicic, & Veldic, 2019). PTSD is defined from a specific set of well described symptoms (e.g. flashbacks, nightmares, avoidance, memory lapses, emotional numbing and hypervigilance, DSM-5: 309.81 or WHO: F43.1) that persists for more than a month and impacts the individual's functioning (American Psychiatric Association, 2013; Kerns et al., 2015).

The purpose of this study is to investigate both factors of resilience and constraining factors' influence on PTSD symptoms in school-age children. Anxiety is very common as a comorbid condition to neurodevelopmental disorders like autism and Attention-Deficit Hyperactivity Disorder (ADHD) (Green & Ben-Sasson, 2010; Tannock, 2009) and is also very common in combination with e.g. depression, eating disorders and psychosis.

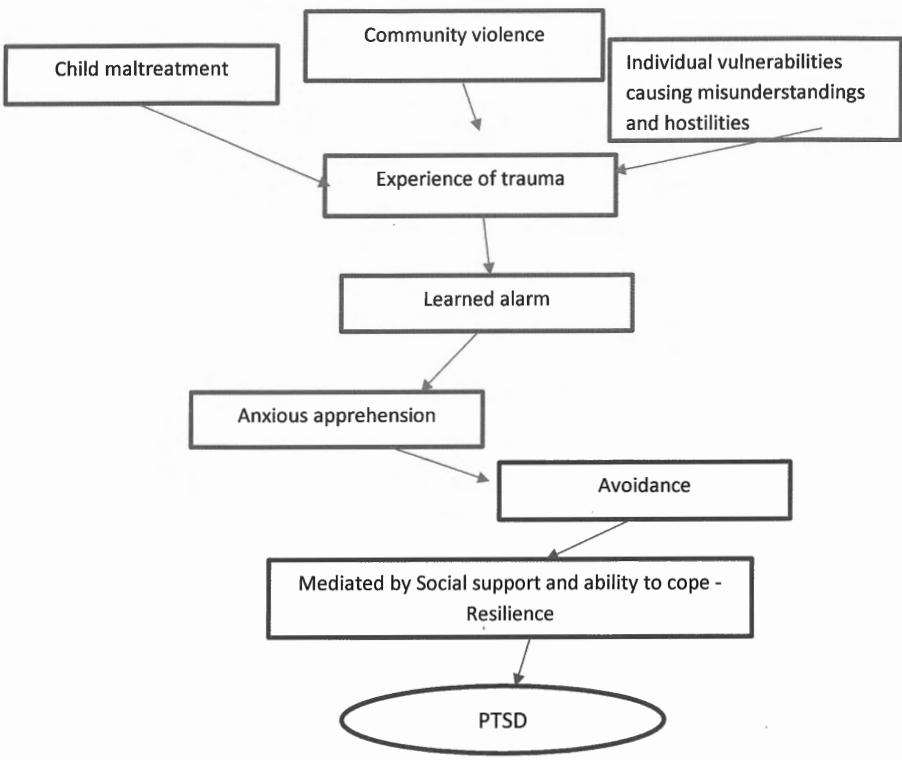
Methodological challenges

- (a) A relatively low prevalence of potential risk factors such as child maltreatment and family violence.
- (b) A low prevalence of individual extreme stressors such as vehicular accidents, industrial accidents, war, rape, torture, terrorism, natural disasters.
- (c) The presence of other confounding factors (e.g. poverty, hostile environmental factors, bullying in school, and individual vulnerability).
- (d) Unknown potential risk factors (e.g. genetic risk factors or childhood traumas that are not registered) and unknown protective factors (e.g. social support).

Theoretical issues

The aim of the study is to contribute to the understanding of the etiology of PTSD in seven to eighteen years old children and adolescents. We want to investigate the prevalence of PTSD and identify risk and protective factors for, and underlying causes of incident PTSD. The present study is inspired by Mowrer's two-factor learning theory (see Figure 1) which posits that fear is learned through associations and individuals will do whatever is necessary to escape and avoid cues that stimulate these aversive emotional memories. Learned alarms emerge from one special chain of events and occur during exposure to situations that resemble the traumatic event. The anxiety is moderated by adequate coping skills and social support. Identification of the precipitating event or proximal cause of PTSD is relatively simple, based on the theoretical descriptions of this form of anxiety (Keane & Barlow, 2004).

Figure 1. A model of the etiology of PTSD



This hypothetical model is informed by prior theoretical models proposed by Keane, Barlow, Mowrer and De Bellis (Keane & Barlow, 2004; Mowrer, 1960)

Many factors affect the development of PTSD: genetic vulnerabilities; frequency and intensity of trauma; developmental stage at which trauma occurs; and comorbid psychiatric and substance use disorders (Blacker et al., 2019).

Experiences of child maltreatment

Child maltreatment is a severe form of interpersonal trauma to which the child is exposed on a daily basis or in periods (De Bellis, Michael D., 2001; Ethier, Lemelin, & Lacharité, 2004).

Early preadoptive child maltreatment

Before being adopted, many children have been exposed to some kind of adversity, although most of the adopted children show psychological adjustment within the normative range (Bimmel, Juffer, Van IJzendoorn, & Bakermans-Kranenburg, 2003; Juffer & Van IJzendoorn, 2005).

Social support as a mediating factor for creating resilience

The ecological maltreatment model suggests that maltreatment can emerge when multiple risk factors out-weight possible protective, compensatory, and buffering factors where applicable (Cicchetti & Lynch, 1993) in the family.

Living in a disadvantaged area

Three primary types of trauma are often used: *community violence* (e.g. war, inner city gang-related violence), *accidents* (e.g. transportation accidents, natural disasters), and *family or individual trauma* (e.g. interpersonal trauma, child abuse and neglect) (Foy et al., 1996). In some inner-city areas in USA children have a high risk of being exposed to community violence. Parents and other adults can provide valuable support, but are limited in how much they can offset the effects (Luthar & Goldstein, 2004).

Individual vulnerability

One of the basic assumptions of developmental traumatology research is that the biological stress system's regulation will be based on the nature of the stressors, frequencies, chronicity of the stressors, and individual differences i.e. genetic vulnerabilities (De Bellis, M. D., 2001). For example, individual genetic and biologic differences that lead to neurodevelopmental deviations such as Attention-Deficit Hyperactivity Disorder (ADHD) and Autism Spectrum Disorder (ASD) can produce an overwhelming amount of stress or conflicts with peers and family members, the immediate neighborhood and other social settings. It is well-known that certain disabilities cause higher prevalence of social misunderstandings and environmental conflicts in schoolchildren i.e. in children with ADHD, ASD or mild mental retardation. Police reports on violence and sexual assaults are also seen more often in these vulnerable groups, when controlled for other potential risk factors (Christoffersen, Mogens N., 2019; Christoffersen, Mogens Nygaard, 2020). Social isolation, family stress and poor communication skills increase risk of maltreatment in children with disabilities (Sullivan & Knutson, 2000). Therefore, children with ASD and ADHD are at an increased risk of experiencing trauma-related symptoms due to daily stressors such as social confusion, peer rejection, punishment, being reprimanded, difficulties coping with changes, and due to their difficulties in regulating of emotions and coping with stress (Kerns et al., 2015).

Children with ASD have profound social cognitive disabilities which make it difficult for them to deal with symbols, including language symbols of thoughts and feelings. Some children with ASD describe that they live in an environment in which they are unable to make themselves understood (Peeters & Gillberg, 1999). Aggressiveness and hostilities from environment may cause anxiety in the children and adolescents who find it incomprehensible. We hypothesize that ADHD and ASD may increase the risk for encountering traumatic events and for developing PTSD.

Methods: Research design

The study is prospective and based on longitudinal panel data from Denmark including the whole population sample, which provides information about the chronological sequence of potential causes before first-time registration of PTSD. The primary outcome

is incident cases diagnosed PTSD in the secondary sector, (i.e. hospital based mental health care). The outcome is binary – either it occurs, or it does not occur.

The study examines which risk factors preceded the first-time recorded PTSD diagnosis in registers set up by hospitals. We define a risk factor as a correlate that is shown to precede the outcome of interest according to Kraemer and colleagues. Risk is as a probability for an event (or an outcome) within a specified population (Kraemer, Lowe, & Kupfer, 2005). Data consists of administrative records with indicators of child maltreatment, pre-adoptive risk, environmental stressors in general, and indicators of various individual vulnerabilities. Administrative data are derived from numerous independent agencies such as police records, hospital records, and educational records, records from the criminal statistic register, records from social assistance register, and records from unemployment statistics.

Study population

National birth cohorts of children born 1984-1994 age 7-18 are followed (N=679,000) until a first time PTSD confirmed in a psychiatric ward.

The data are used for indication of experiences of child maltreatment, pre-adoptive risks or individual vulnerabilities, traumatic stress in the family, family lack of support, disadvantaged neighborhood, and demographic variables. Data are available from 1980 and forward for both children and their parents. Data on victims of sexual assault or violent assault are only available after 2001 by way of an administrative register of police records. This was decisive for the decision to limit the study to include the window from 2001 to 2012 where potentially victims are tracked.

Risk factors (covariates): Indicators of traumatic stress in the family

Children's developmental disabilities and other disabilities are considered individual vulnerabilities. The types of disability are based on a database mandated, compiled and maintained by Danish hospitals in accordance with the international statistical classification (ICD-10) of diseases and health-related problems (WHO, 1994). We classified disabilities into 14 main groups, which did not cover all disabilities. The categories did not include disabilities, which could be consequences of maltreatment such as internalizing disorders, depression, and other emotional disorders.

For the sake of our analysis, some disabilities that may affect the adolescent's ability to communicate, were grouped under speech disability (i.e. developmental disorders of speech and language), while other disabilities such as Cerebral Palsy was kept in a separate category.

Data analysis

Our model allows that individuals can have multiple disabilities. When reviewing the effect of a specific type of vulnerability in the regression analysis, the reference group would be the person-years without that specific type of disability. The model allows for changing covariates and changing disability over time. The study takes advantage of

analysis of covariance and multiple regression statistical analysis methods so that inter-relationship between several predicative variables and first-time PTSD can be examined simultaneously.

Potential demographic risk factors such as living in a disadvantaged area, long-term parental unemployment, parental teenage-motherhood, non-Danish citizens, and gender are included.

The data is analyzed by the discrete time-Cox-model (Allison, 1982). A discrete-time model treats each individual history as a set of independent observations. Only incident cases are included into the model in order to characterise the sequence of events characterising the causes and circumstances of the event. Individuals' history is broken up into 12 sets of discrete time-units (age 7 to 18 years) in which an event either did or did not occur. Each individual is observed until either an event occurs, or the observation is censored by reaching the age limit, because of death, or because the individual is lost to observation for other reasons e.g. immigration. Consequently, individuals are excluded from the case group and controls after the first event.

The person-years at risk were constructed for the total birth cohort. Pooling the non-censored years of all individuals, the person-years, made the numbers at risk ($N=4,917,535$). It has been shown that the maximum likelihood estimator can be obtained by treating all the time units for all individuals as though they were independent, when studying first-time events (Allison, 1982).

The model allows for changing covariates over time. The risk covariates such as experiences of psychosocial maltreatment, stressful life events or individual differences are divided into three types for the purpose of this study. The Type I covariates are those that are taken to be indicative throughout the risk period, irrespective of when the covariate was notified e.g. parental substance abuse or child's diagnosis of autism. Covariates of Type II, in contrast, identify the presence of that factor in the year prior to the event e.g. parental long-term unemployment during a calendar year, or moving into a disadvantaged housing area, until moving out. Finally, the Type III covariates act on the following year and all the subsequent years when observed the first time e.g. family separation or brain injury.

Finally, for each age-groups (age=8, 9,..,17) a constant (Dummy) is estimated. This allows for standardization relating to age. Maximum likelihood estimators for the regression models are then calculated on the basis of pooling all the person-years.

Mediator analysis

Social support is hypothesized as a mediator (Figure 1) between traumatic events and PTSD. The method uses three regression equations to test for the statistical significance of a mediator effect (Baron & Kenny, 1986).

[Fig. 2 presents the three regression equations. In case of a simple linear regression, the regression coefficient (or the total effect) C' will be the sum of the indirect effect $A \times B$ and the direct effect C . It can be proven algebraically without any reference to time

sequence or cause and effect (Wonnacott & Wonnacott, 1990). Sobel (1982) provided an approximate significance test for this purpose but used a bootstrap approach to obtaining confidence intervals, according to Preacher and Hayes (Iacobucci, 2008; Preacher & Hayes, 2004; Sobel, 1982). We assume that the independent variable (childhood trauma) is active at t1, the mediator (social support/lack of social support) is effective at t2 while the outcome variable (PTSD) is measured at t3. The sequential ordering ($t_1 < t_2 < t_3$) can be tested.]

Statistics

We will describe schoolchildren with a PTSD diagnosis in comparison with their contemporaries. In order to evaluate the risk factors' contribution to the number of persons diagnosed with PTSD, attributable fractions (AF) are calculated (Greenland, 2008).

Attributable fractions express the reduction in incidence of PTSD that would be achieved if the population had not been exposed at all compared with the current exposure pattern (Greenland & Drescher, 1993). The estimated AF of a certain risk factor depends on two parameters, only. One parameter is the strength of the risk factor measured by OR or Relative Risk (RR). The other parameter is the current exposure of the risk factor in the population. The estimated AF is calculated solely on the basis on these two parameters (Levin, 1953; Woodward, 1999). AF is only defined when OR and Relative Risk is more than 1.

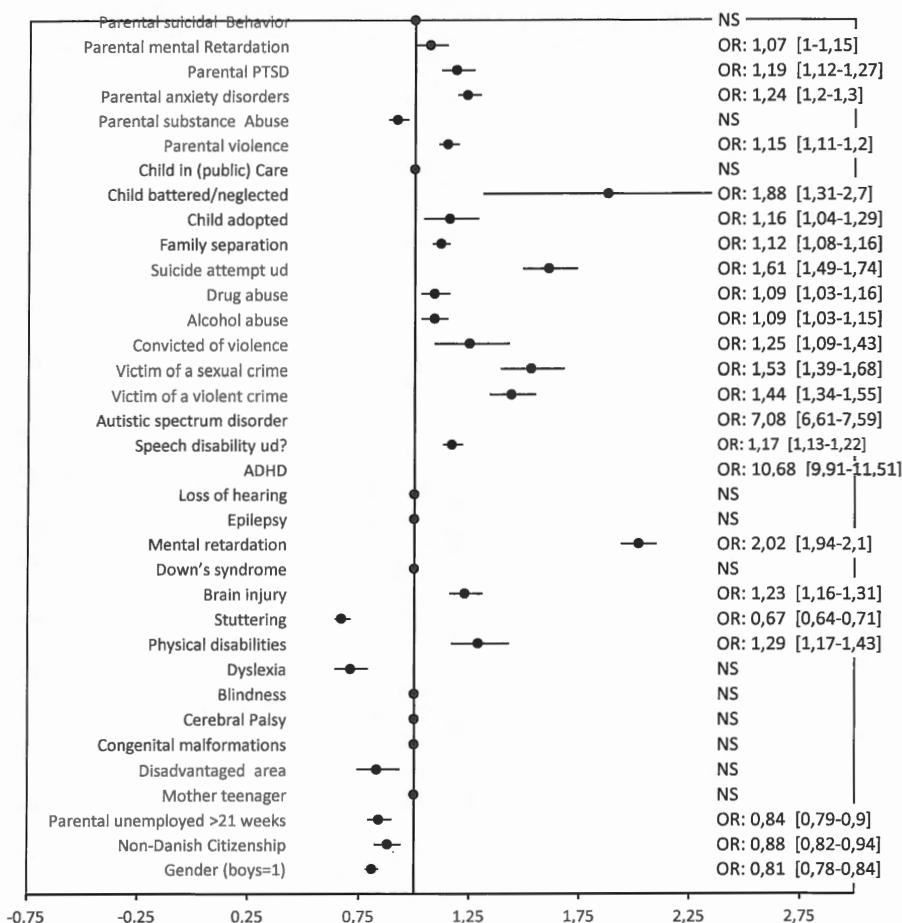
Results: Test of the trauma stress model

The lifetime prevalence of PTSD in school-age children diagnosed according to ICD-10 in a hospital ward was 2.3 per cent (N=15,636 following the eleven birth cohorts from seven to eighteen years).

The aim of the study was to illuminate the etiology of PTSD in children and adolescent by using Danish Register data to evaluate associations between adverse environmental factors and traumatic events and later PTSD in school-age children. We found a lifetime estimation of PTSD in school-age children in the range of 2.3%. The PTSD diagnosis is confirmed only in psychiatric wards, and it is likely that less severe cases will not be included in the present study.

We found an increased risk for PTSD among children diagnosed with neurodevelopmental disorders like ASD, ADHD, children with low levels of social support. Children exposed to child maltreatment, and children being adopted showed an increased risk of PTSD. These estimates are most likely in the lower end due to hidden data on anxiety disorders, and the hazards such as domestic violence, child maltreatment, conflicts and bullying between school children and other potential risk factors that are underreported or not reported to the registers. The results have similarities with previous studies but results also indicated structural divergences.

Figure 2. Forrest Plot of subgroups, post-traumatic stress disorder, adjusted Odds Ratio (OR).



Discussion

We conclude that the high prevalence of PTSD among school-age children with ADSH or ASD must be understood both as a co-morbid phenomenon and as a secondary condition that develop more easily due to a low threshold for PTSD in school-age children with ASD or ADHD because of their problems with understanding and interpreting all kinds of stimuli from the surroundings.

Our study of disadvantaged areas did not corroborate the American findings, maybe because disadvantaged areas are smaller and the society is more supportive? Rates of exposure to violence are high among children in United States (Luthar & Goldstein, 2004). These studies have methodological difficulties separating the individual risk factors in the family from characteristics of local areas. When other potential risk factors such as family violence, being a victim of violence, were taken into account, disadvantaged areas in isolation were not associated with school-age children's PTSD. This may be explained by aspects of the Danish welfare system like higher levels of security and smaller societal differences between rich and poor.

To obtain data for all important outcomes of severe constraints on school-age children, it is necessary to examine long-term follow up studies based on observational data. Non-randomized population studies might be the only way to study effects of for example maltreatment, environmental reactions to individual vulnerabilities and anxiety.

Our results were consistent with the described traumatic stress model. Children exposed to maltreatment or other kinds of traumatic stress (e.g. pre-adoptive deprivation, sexual abuse, violence in the family, community violence) will almost always react with fear, re-experiencing of the trauma and will be likely to avoid stimuli associated with the trauma. Interpersonal trauma where the perpetrator is a family member increases the risk of PTSD and erodes social support, since the child lives with the person who represents the trauma. Parents' own anxiety symptoms or other mental health problems may reduce their capacity to support children. The lack of available social support for resilience and the exposure to traumatic stress were associated with diagnosed PTSD in school-children.

Limits of the study

The results indicate that children with ASD and children with ADHD have a seven to ten times higher risk of being diagnosed with PTSD, when other potential constraints were taken into account. We suppose that these figures estimate the association because Berkson's bias may influence data (Berkson, 1946; Schwartzbaum, Ahlbom, & Feychtig, 2003). We have a reason to believe that many children with PTSD do not get sufficient treatment for the disorder, that may become chronic, from psychiatric ward, municipality early interventions or general practitioners, and PTSD may in some cases only be reported in connection with elucidating decreasing function or increasing of symptoms of ADHD or ASD.

One drawback to the present study is that the data only provide information on diagnoses on personal vulnerabilities and lack information on existing and received help and social support. We use family dissolution as an indicator on lack of social support, but family cohesion is not the same as social support. There is a need for further studies because it is crucial to understand the possible positive impact of social support and adapt treatment strategies to counterstrike adverse outcomes.

Cognitive Behavioral Therapy (CBT) is recommended as first choice treatment of schoolchildren, Trauma-focused cognitive-behavioral therapy has demonstrated positive outcomes in reducing symptoms of PTSD (de Arellano, Michael A Ramirez et al., 2014). Individually information on these measures are not made available in administrative databases for the use of research.

Prospective longitudinal studies on large probability samples offer the best way to study predictors of environmental stressors causing trauma and possible PTSD in school-age children. The present study is based on a huge sample with comprehensive information about potential risk factors for all individuals. This allows for disentangling of the predictors' influence on the risk of developing PTSD. Unknown potential risk factors are the Achilles' heel of the strategy. The administrative data-system gives little or no knowledge of individual measures and local initiatives in primary sector taken to address potential traumatic life-events and anxiety disorders in school-age children.

Another key psychosocial factor is individual vulnerabilities (e.g. ADHD, ASD) causing misunderstandings, hostilities or just negative feedback and experiences from the surroundings. The positive role of psychoeducation and other social and educational interventions in children with neurodevelopmental disorders is corroborated by a systematic review (Montoya, Colom, & Ferrin, 2011), but ongoing research is needed to find means of changing the psychosocial environmental conditions by way of psychoeducation.

Extreme group analysis of patient cost of antibiotic prescribing among general practitioners

Bootstrapping of confidence intervals in subgroups of a sample of General Practitioners

Work in progress: Please do not cite or circulate without permission

Troels Kristensen¹, Charlotte Ejersted², Jens Søndergaard³

¹ Danish Centre for Health Economics (DaCHE), University of Southern Denmark

² Department of Endocrinology, Odense University Hospital

³ Research Unit of General Practice, University of Southern Denmark

Background: Both high and low empathy among GPs may influence patient care. For instance, low empathy may have cost due to decreased patient satisfaction and patients frequenting the GPs less often or even switching GP. Nevertheless, very little is known about the magnitude and variation in antibiotic prescribing-profiles among GPs with high versus low empathy levels.

Aim: To make profiles of antibiotic prescribing for GPs with high versus low empathy and estimate uncertainties in statistic metrics in these groups of GPs. In addition, to explore alternative approaches to bootstrapping of confidence intervals (CIs) for descriptive statistic metrics.

Methods: This study applies extreme group analyses (EGA) to explore patient costs of antibiotic prescribing among subgroups of GPs with high and low empathy from a stratified random sample of 464 Danish GPs. The dataset includes combined survey-data on GP empathy and drug register-data merged via the GP's authorization number. Antibiotics were divided into subcategories of penicillin, non-penicillin antibiotics, antifungals as well as broad and narrow spectrum antibiotics based on the Anatomical Therapeutic Chemical classification (ATC). The GPs in the top decile and bottom decile of the empathy score distribution, were identified to make profiles of their antibiotic prescribing in terms of patient cost. The profiles included the means, coefficient of variation (CV), variation index (VI) and mean differences (DI) in cost between the subgroups. Next, the uncertainty of the costs was estimated via CIs for all antibiotic categories. These CIs were estimated in 3 alternative ways: 1) A bootstrap procedure using the predefined extreme groups samples (n_1, n_2) only, but still for all antibiotic categories, 2) One single bootstrap procedure using the full sample of GPs ($n=464$) estimating CIs for all antibiotic categories and 3) A serial bootstrap procedure for each separate antibiotic category. This allows us to compare the CIs for the three alternative approaches.

Results: There were relatively few differences in patient cost of antibiotic prescribing across GPs with extreme empathy levels in this sample of GPs. However, it was a trend in the data that the high empathy group had lower patient cost of antibiotic prescribing for most categories of penicillins than the low empathy group. Bootstrapping of CIs based on predefined subsamples rather than the entire sample resulted in larger CIs. Restricted application of data from predefined extreme groups may lead to wider and biased CIs in extreme group analysis.

Conclusion: This study shows that extreme group analysis and related GP profiles can be used to explore antibiotic prescribing behavior among GPs. One generated hypothesis is that high empathy GPs prescribe some penicillin categories in a different way that results in lower patient costs than low empathy GPs. Included CIs based on alternative approaches reveal the level of uncertainty in the estimated patient cost across the included antibiotic categories.

1. Introduction:

It is well known that antibiotic prescribing behavior is very important for patient in both primary and secondary care as well as the society in general¹. One central reason is that unwarranted variation in terms of so called “under prescribing” and “over prescribing” may be harmful for patients and society. “Under prescribing” occur when patients receive no or too little of one or more categories of drugs such as penicillins or other non-penicillin antibiotics. “Over prescribing” means that the patient receives more antibiotic prescriptions than needed from a clinical or biomedical point of view^{2,3}. In practice, it is usually best to use specific antibiotics or narrow spectrum antibiotics to avoid use of broad spectrum antibiotics if a narrow spectrum substitute exists. Another best way is to avoid antibiotics if other better or equivalent options exists. In cases where no antibiotic bacterial resistance measures have been done, antibiotics may be avoided by careful examination of patients as symptoms may be related to non-infectious disease or virus. To understand and address unwarranted variation in antibiotic prescribing it is important to explore circumstances that may influence prescribing of antibiotics and by using microbial diagnostic methods. For instance, it has been hypothesized that variation in GP empathy may influence GP behavior and thus prescribing of antibiotics⁶. The reason is that both high and low empathy have been shown to directly influence other elements of patient care. Both directly in terms of unwarranted variation for the individual patient and indirectly, in terms of patient opportunity cost and negative externalities in terms of antibiotic resistance for society. Therefore, it is relevant to understand the association between GP empathy and their antibiotic prescribing profiles. However, before this type of confirmatory hypothesis can be tested, there is a need to undertake hypothesis generating exploratory analysis, which calls for the latter analysis. So far, very little empirical research has been conducted around empathy levels and drug prescribing⁶. Subgroup analysis such as extreme group analysis has been stated to be well suited to undertake exploratory analysis^{5,7}. Besides it helps us understand the profiles of antibiotic prescribing among GPs with high and low empathy levels. The present study focuses on direct patient prescribing cost and not least the approaches to estimation of uncertainty related to these patient costs. From a resource perspective cost or expenditure to patients is standard way to measure resource use.

Accordingly, the aim of this study is to make profiles of antibiotic prescribing patient costs for GPs with high versus low empathy and quantify uncertainties in statistic metrics based on the available sample of GP survey and register data⁸. Furthermore, it is an aim of this symposium paper to explore alternative approaches to bootstrapping of confidence intervals (CIs) for descriptive statistic metrics.

2. Method

Extreme groups analysis was employed to explore the nature of antibiotic prescribing patient costs among subgroups of GPs with high and low empathy in a sample of 464 general practitioners^{4,5}. The patient cost of antibiotic prescribing was measured in terms

of the pharmacy retail pris (abbreviated AUP in Danish) aggregated over all fillings. The price is set as the cost price (AIP) plus profit margin and value added tax (TVA) and represents the direct cost to consumers.

The GPs empathy was measured via the Jefferson Scale of empathy for Health care professionals⁹. The dataset includes combined survey-data on GP empathy and drug register-data merged via the GP's authorization number for the year 2017. Antibiotics were divided into categories of penicillin, non-penicillin antibiotics, antifungals as well as broad and narrow spectrum antibiotics. The GPs in the top decile and bottom decile of the empathy score distribution, were identified to make profiles of their costs of antibiotic prescribing. The profiles included the mean costs, coefficient of variation (CV), variation index (VI) and mean differences (DIs) between the groups. Next, the uncertainty of the costs was estimated via CIs for all cost categories.

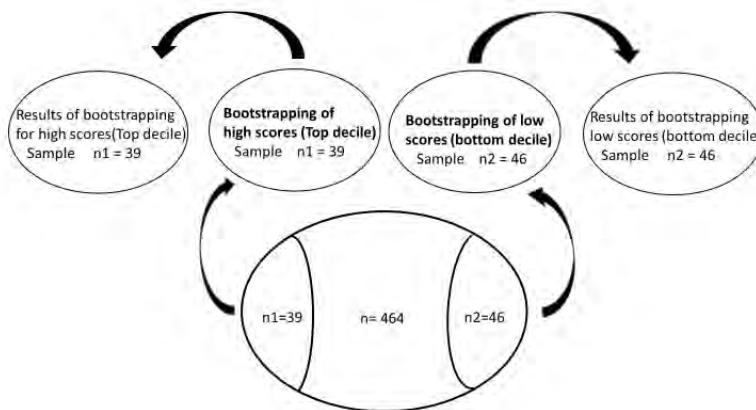
2.1 Estimation of confidence intervals in the bottom and top decile:

This study has used bootstrapping to estimate uncertainty around cost of antibiotic prescription metrics in terms of CIs similar to other health economic studies^{10,11}. The idea is that the bootstrapped empirical cumulative density function based on our data sample, gives a good sense of what the true unknown (population) distribution is¹². In addition to extreme group analysis (EGA) of GPs with high and low empathy, the idea was to quantify and explore the effect of three alternative bootstrapping approaches on CIs:

1. Bootstrapping based on subgroups of GPs with high and low empathy¹³. Here, the two extreme groups were predefined as top and bottom deciles and identified before the bootstrap procedure was implemented as one simultaneous or parallel procedure for estimation of CIs for all variables, see illustration in Figure 1.
2. Bootstrapping using the entire sample to calculate extreme groups including parallel estimation on related statistic metrics and their confidence intervals, see Figure 2. This was based on a program adopted in Stata. The idea was to calculate all subgroup statistic metrics in one bootstrap procedure rather than several independent standard procedures for each variable or antibiotic category. The parallel estimation of all statistic metrics will restrict variation to the variation created by the specific bootstrap procedure. This is different from a procedure where the bootstrap is performed in a serial way for each statistic metric. The latter is expected to increase variation because the results for the statistic metrics will be based on multiple bootstrap procedure – in particular when the start seed option for the bootstrap is set to random.
3. Independent serial bootstrapping of each individual variable according to approach 1 (for the pre-calculated extreme top and bottom decile groups) and approach 2 (using the entire sample for bootstrapping). In these cases, the statistic metrics for each variable will be based on separate bootstrapping procedures where the variation will be less restricted and CIs for metrics expected to be wider compared to approach 1 & 2.

The bootstrap replicates, which increases accuracy was set to 1000. The seed for the bootstrap was set to zero (random) to allow for differences between 1 and 3 as well as 2 and 3.

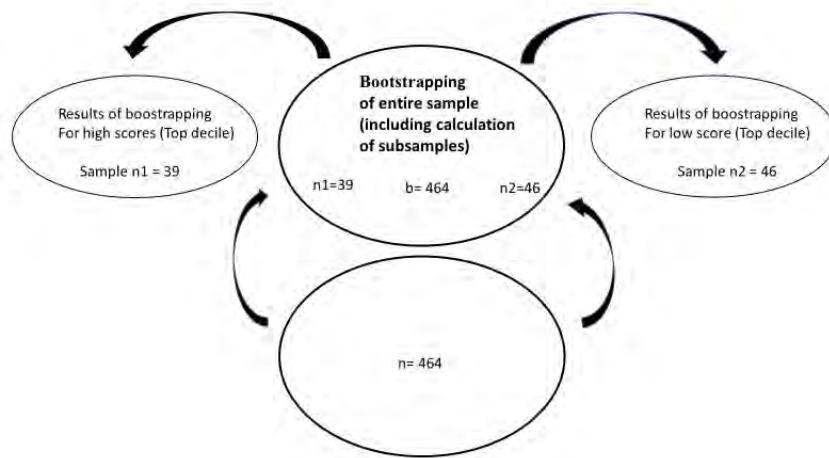
Figure 1: Bootstrapping approach 1 based on predefined extreme groups



Re 1: Bootstrapping approach 1 – using the extreme groups only:

Figure 1 and Figure 2 illustrate the applied approach to bootstrapping of the CI for statistic metrics based on the survey sample. Figure 1 shows the restricted version of the bootstrap procedure where the subsamples were predefined before the bootstrap procedure was applied on the subsamples $n1=39$ and $n2=46$ rather than the entire sample.

Figure 2: Bootstrapping approach 2 based on the entire sample



Re 2: Bootstrapping approach 2 using the entire sample

Figure 2 illustrates the applied approach to bootstrapping of the CI for descriptives statistics based on the entire survey sample ($n=464$). The bootstrap procedure was based on a stata program which calculates the top an bottom deciles of the sample. This program was bootstrapped with respect to the desired statistics metrics such as the mean costs, coefficient of variation, variation index and all antibiotic categories in such a way that all statistic metrics was calculated based on a bootstrap procedure with 1000 replications.

Besides the following examples of the approaches 1, 2 and 3, this study only reports results based on approach 2.

2.1.1 Example: Bootstrapping of CIs for descriptive statistics in extreme groups

In this example the yearly prescribing patient cost related to Macrolides J01F are applied to illustrate the alternative approaches illustrated in Figure 1 and Figure 2 to bootstrap CIs for statistic metrics in extreme group analysis. In Table 1, the mean and CV metrics was chosen to show examples including related bootstrapped CIs for each of the three approaches. The first approach uses predefined subgroups for bootstrapping, the second applies the entire sample and the third uses “serial” bootstrapping for each variable:

Table 1 Alternative uncertainty measures of cost of macrolides prescriptions

Bootstrapping approach:	Mean	High-empathy GPs † (n=39)		
		[95% CI]	CV	[95% CI]
Approach 1 predefined subgroup:				
Macrolides (B) J01F	3,911.6	[2,878.7; 4,944.4]	0.85	[0.67;1.02]
Approach 2 using the entire sample:				
Macrolides (B) J01F	3,935.6	[3,039.7; 4,831.6]	0.82	[0.68;0.94]
Approach 3 serial approach:				
Predefined subgroups:				
Macrolides (B) J01F:	3911,6	[2900.3;4922.9]	0.85	[0.54;0.79]
Using entire sample:				
Macrolides (B) J01F:	3935,6	[3047.6;4823.6]	0.81	[0.68;0.96]

These three alternative approaches may be applied in practice to estimate uncertainty of statistic metrics in top decile and bottom decile groups of GPs as part of an EGA. The applied illustrative example for Macrolides (J01F) for the different approaches is based on data from the high empathy group and includes the 95% CIs. According to approach 2, which uses the entire sample of GPs to estimate the CIs, the mean yearly Macrolides J01F patient prescription cost was DKK 3,935.6 within a 95% CI of [3,039.7;4,831.6] and the coefficient of variation was 0.82 within a 95% CI of [0.68;0.94]. In contrast, approach 1 only uses the reduced predefined subgroup (n=39) to estimate the mean and related CI. Compared to approach 2, this shows a similar mean and a broader CI [2,878.7;4,944.4] corresponding to larger uncertainty, when a smaller subgroup is used to bootstrap CIs. This was also the case for the CV and the related CI when approach 1 and approach 2 were compared. Both the CV [0.85] and the 95% CI [0.67;1.02] in approach 1 reflected larger variation an uncertainty than in approach 2 with CV [0.82] and 95% CI [0.68;0.94].

The GP profile characteristics in the subsamples with high and low empathy have been reported in a related study¹⁴. So far, the characteristics of the entire sample have been described in a previous study⁹. Therefore, this study only focuses on the cost of antibiotic prescribing in the reported subgroups.

3. Results

The results of the EGA of yearly prescribing patient cost in Table 2 and Table 3 show the estimated yearly mean antibiotic prescriptions patient costs and within subgroup variation in terms of the CV for the groups of high empathy GPs (n1= 39, Table 1) versus the same yearly mean cost and CV for the subgroup of low empathy GPs (n2=46, Table 3). In Table 2, the results including bootstrapped 95% CIs for the mean and CV are divided into subcategories of penicillin, non-penicillin antibiotics, antifungals and broad versus narrow spectrum antibiotics.

Table 2 yearly cost of antibiotic prescriptions for high empathy GPs

Antibiotic category:	High-empathy GPs † (N=39)			
	Mean	[95% CI]	CV	[95% CI]
Penicillins:				
Extended spectrum (B) J01CA	28,855.8	[23,457.6; 34,254.1]	0.68	[0.48;0.88]
Beta-lactamase sensitive (N) J01CE	6,652.0	[5,721.6; 7,582.3]	0.53	[0.42;0.63]
Beta-lactamase resistant (N) J01CF	6,129.0	[3,712.4; 8,545.6]	1.47	[0.94;1.99]
Combinations with beta lactamase inhibitors (B) J01CR	2,162.4	[1,459.2; 2,865.6]	1.24	[0.87;1.61]
All penicillins: J01C	48,546.5	[41,191.7; 55,901.4]	0.56	[0.42;0.69]
Non-penicillin antibiotics :				
Tetrayclines (B) J01A	8,887.6	[1,555.9; 16,219.4]	3.04	[1.04;5.06]
Cephalosporins J01D	32.6	[0.84; 64.3]	3.66	[1.37;5.94]
Sulfonamides J01E	10,175.7	[5,971.5; 14,379.8]	1.52	[1.25;1.78]
Macrolides (B) J01F	3,935.6	[3,039.7; 4,831.6]	0.82	[0.68;0.94]
Quinalones (B) J01M	594.8	[392.3; 797.3]	1.23	[0.85;1.62]
Other antibiotics J01X	3,815.5	[602.0; 7,029.0]	3.10	[1.97;4.23]
All other antibiotics	27,441.8	[17,760.4; 37,123.2]	1.35	[0.92;1.77]
Antifungals, J02	4,983.8	[2,705.1; 7,262.5]	1.72	[1.25;2.20]
Total all antibiotics:				
Narrow spectrum (N)	91,188.8	[74,855.0;107,522.6]	0.67	[0.53;0.80]
Broad spectrum (B)	34,202.4	[26,109.0; 42,295.9]	0.88	[0.65;1.11]
	42,977.2	[33,478.6; 52,475.7]	0.83	[0.50;1.15]

Narrow spectrum antibiotic was defined via the following ATC-codes: J01CE, J01CF, J01DB, J01DF, J01EA, J01EB, J01FA, J01FF, J01XA, J01XC, J01XD, J01XE, J01XX. Broad spectrum antibiotics included: J01AA, J01CA, J01CR, J01DC, J01DD, J01DH, J01EE, J01GB, J01MA, J01MXB.

Overall, the estimated patient costs in Table 2 and Table 3 indicate that there is a trend that high empathy GPs prescribe under circumstances which means their patients have lower cost of antibiotic prescribing compared to the low empathy (bottom decile) GPs. The within subgroup or intra group variation in terms of the CV indicates a trend towards a slightly higher variation in the patient cost among GPs with lower empathy for penicillins. In contrast, the measured CV seems to be lower for the non-penicillins for the subgroup with the low empathy.

Across both tables, the lowest within-category-variation was in the group of penicillins. The largest within-category-variation was for Cephalosporins in the categories of non-penicillin antibiotics.

Table 3 yearly cost of antibiotic prescriptions for low empathy GPs

Antibiotic category:	low-empathy GPs † (N=46)			
	Mean	[95% CI]	CV	[95% CI]
Penicillins:				
Extended spectrum (B) J01CA	40,489.8	[29,195.5; 51,784.2]	0.85	[0.64;1.07]
Beta-lactamase sensitive (N) J01CE	8,343.0	[6,469.0; 10,217.0]	0.71	[0.56;0.85]
Beta-lactamase resistant (N) J01CF	5,971.4	[4,003.6; 7,939.1]	1.05	[0.53;1.58]
Combinations with beta lactamase inhibitors (B) J01CR	2,746.9	[1,988.9; 3,504.8]	0.90	[0.71;1.09]
All penicillins: J01C	64,852.2	[49,822.6; 79,881.8]	0.71	[0.555;0.86]
Non-penicillin antibiotics :				
Tetracyclines (B) J01A	4,775.0	[3,304.2; 6,245.7]	0.97	[0.72;1.22]
Cephalosporins J01D	21.6	[0.4; 42.9]	3.16	[1.19;5.12]
Sulfonamides J01E	6,959.7	[3,391.0; 10,528.5]	1.61	[1.27;1.96]
Macrolides (B) J01F	4,093.5	[2,832.2; 5,354.8]	0.99	[0.78;1.21]
Quinalones (B) J01M	1,143.1	[572.6; 1,713.6]	1.63	[1.06;2.20]
Other antibiotics J01X	6,903.1	[389.8; 13,416.5]	2.93	[1.83;4.02]
All other antibiotics	23,896.0	[15,495.3; 32,296.7]	1.12	[0.86;1.38]
Antifungals, J02	4,990.0	[2,568.6; 7,711.5]	1.56	[1.13;1.98]
Total all antibiotics:	106,931.5	[80,334.4;133,528.7]	0.76	[0.63;0.89]
Narrow spectrum (N)	34,478.6	[24,836.2; 44,121.1]	0.90	[0.65;1.15]
Broad spectrum (B)	53,985.5	[40,138.3; 67,833.0]	0.78	[0.62;0.94]

3.1 High versus low empathy inter-group differences

Table 4 shows an analysis of the difference between the high and low empathy GPs. This means the group differences between the patient costs of each antibiotic category based on Table 2 and Table 3. The latter differences are measured in terms of a) a variation index (VI), b) the mean group difference and c) tests of the differences between the costs of antibiotic prescribing. The variation index (VI) indicates a trend that high empathy GPs have about 20-30% lower patient cost of antibiotic prescribing of most penicillins than low empathy GPs. Furthermore, Table 4 shows that the variation was more fluctuating for non-penicillin antibiotics where the high empathy GPs had higher yearly cost of prescribing for tetracyclines, cephalosporins and sulfonamides. High empathy GPs seems to prescribe less broad-spectrum antibiotics. However, the above mentioned trends were not confirmed by the Mann-Whitney (ranksum) test of group difference and t-test of the group mean difference and related confidence intervals. Only the group of all penicillins was boarder-line significant according to the Mann-Whitney/t-test. In the e-return list of the bootstrap procedure in stata there was a Z0-

test-score for the mean difference of each antibiotic category in Table 4. The last column in Table 4 includes this Z-score for the Mean difference.

Table 4 yearly cost of antibiotic prescriptions – high versus low empathy group differences

	V1	[95% CI]	Mean	[95% CI]	MannWhit-test	Z0
Penicillins:						
Extended spectrum (B) JOICA	0.71	[0.47-0.95]	-11.634.0	[-23,846.1; 578.1]	0.0763/0.0325	0.0258*
Beta-lactamase sensitive (N) JOICE	0.80	[0.58-1.01]	-1,691.0	[-3,762.1; 380.1]	0.2169/0.0760	0.0000***
Beta-lactamase resistant (N) JOICF	1.03	[0.51-1.54]	157.6	[2,863.6; 3,178.8]	0.4482/0.9743	0.0000***
Combinations with beta lactamase inhibitors (B) JOICR	0.79	[0.44-1.14]	-584.5	[-1,620.6; 451.7]	0.1065/0.3233	0.0517
All penicillins; JOIC	0.75	[0.55-0.95]	-16,305.6	[-32,523.4; -87.9]	0.0720/0.0301	0.0258*
Non-penicillin antibiotics:						
Tetracyclines (B) JOIA	1.86	[0.163-3.56]	4,112.6	[-3,385.5; 11,610.8]	0.4588/0.2594	0.1661
Cephalosporins JOID	1.51	[2.10-5.11]	10.9	[-26.8; 48.7]	0.6760/0.8307	0.0181*
Sulfonamides JOIE	1.46	[0.322-6.60]	3,215.9	[-2,390.9; 8,822.7]	0.4588/0.2594	0.0129*
Macrolides (B) JOIF	0.96	[0.581-1.34]	-157.8	[-1,671.7; 1,356.0]	0.9367/0.8202	0.0232*
Quinalones (B) JOIM2	0.52	[0.17-0.87]	-58.3	[-1,151.3; 54.7]	0.2552/0.1221	0.0439*
Other antibiotics (O) X	0.55	[0.80-1.91]	-3,087.6	[-10,355.1; 4,179.8]	0.5139/0.6925	0.0621
All other antibiotics	1.15	[0.54-1.76]	3,545.8	[-9,168.2; 6,259.7]	0.7177/0.3227	0.0414*
Antifungals, JO2	1.00	[0.24-1.76]	-6.2	[-3,327.9; 3,315.4]	0.2627/0.9478	0.0439*
Total all antibiotics:	0.85	[0.59-1.12]	-15,742.7	[-46,579.7; 15,094.2]	0.4588/0.4104	0.0336*
Narrow spectrum (N)	0.99	[0.32-1.36]	-276.2	[-12,583.1; 12,105.9]	0.8255/0.8720	0.0026**
Broad spectrum (B)	0.80	[0.32-1.07]	-11,008.5	[-27,656.6; 5,639.7]	0.2270/0.2486	0.0078**

The differences between the groups have been tested: The Mann-Whitney (ranksum) test was applied. *P<0.05, **P<0.01, ***P<0.001. Furthermore, 95% CIs for the mean group differences have been estimated. All reported confidence intervals are based on bootstrapping (1000 reps) of the observations in the high and low empathy groups respectively.

4. Discussion

This exploratory pilot study of profiles of antibiotic prescribing patient cost among GPs indicate that both high and low empathy GPs prescribe categories of antibiotics in a similar way. This may be interpreted as good news for GP patients. Nevertheless, based on the results one hypothesis may be that high empathy GPs prescribe some penicillin categories in a different way that results in lower patient costs than low empathy GPs. Another hypothesis may be that high empathy GPs prescribe antibiotics in a more homogenous way than the group of low empathy GPs. One explanation could be that high empathy GPs communicate more or better with their patients and therefore are able to explain to them that they do not need the more expensive and broad spectrum antibiotics. Thus, the results reveal that it may be meaningful to conduct further confirmatory research with respect to behavioral research agendas since empathy could have an important influence on patient and subsequently societal costs.

The applied extreme group analysis has strength and weaknesses⁵. However, it is a recognized method to undertake exploratory pilot research and considered well suited to make profile pictures of GP behavior.

Furthermore, it is a strength of this study that the estimated CIs for the profiles allow the readers to know the level of uncertainty of the patient cost in each ATC-category. To assess the sensitivity of the uncertainty to the applied bootstrapping approach this study has used at least three alternative bootstrapping approaches to estimate CIs in subgroup analysis. This experience and the examples included in Table 1 shows that CIs based on bootstrapping on the predefined or reduced subgroups yields broader CIs. Furthermore, serial bootstrapping does not appear to give systematically different results than the preferred approach 2 which is based on the entire sample. Finally, the results demonstrate the chosen approach may have an impact on the results in this type of research and should be investigated further.

Patient cost of antibiotic prescribing was measured in terms of retail prices. The prices did not include any administrative fees for fillings and any final drug preparation fees and can only be perceived as a proxy for patient costs. For instance, changes in patient subsidies and profit margins may have had influence on the patient retail prices in 2017. Besides the applied cost metric did not include other potential patient opportunity cost due to other patient cost such as lost income.

The aim of this study has not been to explain variation in antibiotic prescribing patient costs. It is clear, that fluctuations in the investigated antibiotic patient cost may be determined by a range of patient characteristics such as differences in patient morbidity in terms of co-morbidities/multimorbidity and socioeconomic patient characteristics. Nevertheless, these limitations do not preclude EGA or subgroup analysis of profiles of antibiotic prescribing patient cost across antibiotic categories among GPs with high and low empathy levels according to the Jefferson Scale of Empathy for health care professionals.

5. Conclusion

This study shows that extreme group analysis and related GP profiles can be used to explore antibiotic prescribing behavior among GPs. One generated hypothesis is that high empathy GPs prescribe some penicillin categories in a way that results in lower patient costs for antibiotics compared to prescribing of the same antibiotics categories among low empathy GPs.

From a methodological point of view, bootstrapping can be used to estimate the level of uncertainty in the estimated patient costs across the applied extreme subgroups of GP empathy. However, it should be noticed that the subgroup results are dependent on the way the extreme groups are defined as part of the bootstrap procedure.

Literature

1. Colgan R, Powers JH. Appropriate antimicrobial prescribing: approaches that limit antibiotic resistance. *Am Fam Physician*. 2001;64(6):999-1004.
2. Cole A. GPs feel pressurised to prescribe unnecessary antibiotics, survey finds. *BMJ*. 2014;349:g5238. doi:10.1136/bmj.g5238
3. Alber K, Kuehlein T, Schedlbauer A, Schaffer S. Medical overuse and quaternary prevention in primary care—A qualitative study with general practitioners. *BMC family practice*. 2017;18(1):1-13.
4. Feldt LS. The use of extreme groups to test for the presence of a relationship. *Psychometrika*. 1961;26(3):307-316. doi:10.1007/BF02289799
5. Preacher KJ, Rucker DD, MacCallum RC, Nicewander WA. Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychol Methods*. 2005;10(2):178-192. doi:10.1037/1082-989X.10.2.178
6. Yuguero O, Ramon Marsal J, Esquerda M, Vivanco L, Soler-González J. Association between low empathy and high burnout among primary care physicians and nurses in Lleida, Spain. *Eur J Gen Pract*. 2016;23(1):4-10. doi:10.1080/13814788.2016.1233173
7. Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. *BMJ*. 2015;351:h5651. doi:10.1136/bmj.h5651
8. Hodges S, Klein K. Regulating the costs of empathy: The price of being human. *The Journal of Socio-Economics*. 2001;30:437-452. doi:10.1016/S1053-5357(01)00112-3
9. Charles JA, Ahnfeldt-Møllerup P, Søndergaard J, Kristensen T. Empathy Variation in General Practice: A Survey among General Practitioners in Denmark. *Int J Environ Res Public Health*. 2018;15(3). doi:10.3390/ijerph15030433
10. Eakin BK, McMillen DP, Buono MJ. Constructing Confidence Intervals Using the Bootstrap: An Application to a Multi-Product Cost Function. *The Review of Economics and Statistics*. 1990;72(2):339-344. doi:10.2307/2109725
11. Campbell MK, Torgerson DJ. Bootstrapping: estimating confidence intervals for cost-effectiveness ratios. *QJM*. 1999;92(3):177-182. doi:10.1093/qjmed/92.3.177
12. Efron B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*. 1979;7(1):1-26. doi:10.1214/aos/1176344552
13. Breivik Ø, Aarnes OJ. Efficient bootstrap estimates for tail statistics. *Natural Hazards and Earth System Sciences*. 2017;17(3):357-366. doi:10.5194/nhess-17-357-2017
14. Profiles of GPs with High and Low Physician Empathy: Personal, Professional, and Antibiotic Prescribing Characteristics (preprint)
<https://www.researchsquare.com/article/rs-1624965/v1>



What is SAS® Viya® for Learners?

SAS® Viya® for Learners delivers free access to advanced analytics software for teaching and learning. It is a suite of cloud-based software that supports the entire analytics life cycle – from data, to discovery, to deployment – and lets you code in SAS, Python or R.

Sign Up Today!

As a student or an academic educator, you can easily get access by following this link:
www.sas.com/viya-learners

