Aðgreining á áhrifum aldurs og kynslóða: Mjúkir fordómar (smoothness priors)

helgito@hi.is

Helgi Tómasson

19-03-2021

Skipulag fyrirlestrar

- 35 ára doktorsafmæli. Varði doktorsritgerð 19.mars 1986 (*Estimation and Prediction ARMA models*
- Sögulegt yfirlit, þróun fagsins, val mitt á framhaldsnámi.
- Hugmyndafræði matsaðferða í 200 ár. Undarlegheit í ályktunarfræði(*inference*), Lindley mótsögn ofl.
- Mesti sennileiki (maximum-likelihood), Bayes, pre-test, James-Stein. Fordómar borga sig.
- Mjúkir fordómar (smoothness-priors)
- Notkun á mjúkum fordómum við sundurliðun aldurs- og kynslóðaáhrifa. Kennslubókardæmi og praktíst dæmi um þróun á tíðni sjálfsvíga í aldri og tíma. Í samvinnu við Högna Óskarsson geðlækni.
- Lokaorð

Sögulegt yfirlit

- Fyrir 200 árum gátu menn (t.d. Laplace) reiknað líkur á að fá 3svar sömu útkomu er peningi var kastað upp 10 sinnum.
- Ef peningi hefur verið kastað 10 sinnum, við fengið 3svar sömu útkomu, hvað segir þetta um peninginn?
- Þetta leiddi til umræðu um „*inverse probability*". Þ.e. hendingin (rand-variable) Y, er mæld 10 sinnum og útkoman er $y_1, \ldots, y_{10}$.
- Gert er ráð fyrir að köstin séu óháð og að $P(Y = 1) = \theta$. Hvað segir þessi útkoma (summa kastanna sé 3) um θ?
- θ er óþekkt og ég vil nota gögnin $y_1, \ldots, y_{10}$ til að álykta um θ.

- Hin bayesíska nálgun gerir ráð fyrir að lýsa megi vissu (óvissu) um θ með líkindadreifingu. Þ.e. θ er hending (random-variable).
- Gangurinn er þannig að fyrirframvissan er sett fram með $\pi(\theta)$ og eftir-á-vissan með:

$$\pi(\theta|y_1, \ldots y_{10}) \propto P(y_1, \ldots y_{10}|\theta)\pi(\theta).$$

- Við köllum fallið L, $L(\theta|y_1, \ldots y_{10}) = P(y_1, \ldots y_{10}|\theta)$, sennileikafall (likelihood-function) og fallið P, fall af $y_1, \ldots, y_{10}$, líkindafall (probability-function). Atriði að orðin hljómi ólíkt.
- Ef fyrirframvissan (fordómarnir) í þessu tilfelli er beta-dreifing, $\pi(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$, þá eftir-á-vissan líka beta-dreifing. Fordómarnir eru að meðaltal peningsins sé $a/(a+b)$. Áhrif fordómanna hverfa með vaxandi fjölda mælinga.
- Eftir-á-dreifinginn inniheldur allar upplýsingar um θ. Þ.e. $I_{total} = I_{data} + I_{prior}$.

- Á 19. öld gerðist ýmislegt, t.d. gat Laplace giskað á massa Júpíters og Satrúnusar. Til varð kúltúr sem t.d. Spanos kallar *biometric tradition*.
- Undor lok 19. aldar kom fram maður Pearson-eldri sem setti fram ýmsar hugmyndir (Pearsonian kerfi af dreifingum o.s.frv).
- Upp úr 1900 kom fram guðfaðir nútímatölfræði Fischer, sem umbylti ályktunarfræðinni. Pearson-eldri hafði talið að hlutverk tölfræði væri að lýsa gögnum. Fischer taldi hlutverk tölfræðinnar vera að setja fram vísindalegar alhæfingar. Hann rökstuddi að aðferð mesta-sennileika, *maximum-likelihood*, væri góð aðferð (í einhverjum skilningi optimal).
- Fischer átti í deilum við ýmsa menn, t.d Neyman og Pearson yngri um túlkanir á kenningaprófunum.
- Kenningaprófanir gengu út á að gögnum væri safnað og athugað hvort fengin útkomma væri ósennileg gefið H_0 rétt. Fischer og Neyman-Pearson deildu um hvernig túlka skyldi þetta.
- Í grófum dráttum fer þetta þannig fram að reiknaðar eru líkur á fenginni útkomu eða ótrúlegri, p-gildið, og ef p-gildið er lítið er ályktað að um marktækan (*significant*) mun sé að ræða.

Lindley paradox

- Maður að nafni Jeffey blandaði sér í deilurnar. Hann vildi leysa málið og notaði bayesíska nálgun. Nálgunin er gróflega:
- Set fram H_0, t.d. $\theta = 1/2$. Set fordóma a-priori líkur π_0 á að H_0 sé rétt. Þar með er a-priori odds π_0/π_1, $(\pi_1 = 1 - \pi_0)$. Þegar gögnum hefur verið safnað má reikna eftir-á-odds,
  Bayes-factor=BF= $\frac{\text{líkur á } H_0 \text{ rétt}}{\text{líkur á } H_1 \text{ rétt}}$.
- Jeffey lagði til, ca. hafna H_0 ef BF er lítið, samþykkja ef BF er stórt.
- Afleiðingin, meiri deilur.

Use of p-values

An example from Young and Smith (2005, p. 77).

- Bayesian testing is based on the Bayes-factor:

$$BF = \frac{P(H_0 \text{ true}|\text{data})}{P(H_1 \text{ true}|\text{data})}$$

- Large values of BF support H_0, small values support H_1.
- Suppose X is binomial $B(n, \theta)$, $H_0 : \theta = \theta_0$, $H_1 : \theta \neq \theta_0$. The prior under H_1 is uniform, $U(0, 1)$.
- The observed data is $X = x$. Then

$$BF = \frac{(n+1)!}{x!(n-x)!}\theta_0^x(1-\theta_0)^{n-x}.$$

- Lets say $n = 100$, $\theta_0 = \frac{1}{2}$, and $x = 60$. I.e. $\hat{\theta} = 0.6$.
- What should we say about H_0?
- The classical z-value is 2, so we should reject H_0 if $\alpha = 0.05$.
- However, BF=1.09, that is the Bayesian says we have increased our support for H_0.
- This impact is much more dramatic if n is large.
- How should this be interpreted in practice?

My interpretation:

- If someone tells me a about a study that shows a small but "statistically significant"(e.g., $p < 0.05$) in a very large dataset. This usually means: **Increased support** for H_0 (which is typically no impact).
- Bayarri and Berger (1999) give a review article, "Quantifying Surprise in the Data and Model Verification". Several prominent statisticians, J.M. Robbins, D.K. Pauler, B.P. Carlin, M. Evans, D.V. Lindley, X-L. Ming, A. Gelman, M. Mouchart, T. O'Hagan, L.R. Pericchi, D.A. van Dyk complement with advanced discussion.

- Hypothesis-testing and p-values are complicated issues. Try to avoid them. A lot of smart people have been trying to figure them out for decades. "Could Fisher, Jeffreys and Neyman Have Agreed on Testing?", Berger (2003) I think this is still an open problem.
- Spanos (1999, page 725) on the debate Fishers/Neyman: *Fisher and Neyman would turn in their graves if they were to find out how the modern textbook hybrid on hypothesis testing managed to reconcile what they considered irreconcilable differences!* .
- Lindley (1999) segir: *My personal view is tha p-values should be relegated to the scrap heap and not considered by those who wish to think and act coherently*.

- Hubbard and Lindsay (2008, page 82) on Fisher's *p*-values: *his widely misunderstood and defective p-values blanket the empirical literature*.
- Gigerenzer[The empire of chance, p. 106] says: Although the debate[Fisher Neyman] continues among statisicians, it was silently resolved in the "cookbooks"written in the 1940s to the 1960s, largely by non-statisticians, to teach students in the social sciences statisticians, "the rules of statistics". We call this compromise the "hybrid theory"and it goes without saying tha neither Fisher nor Neyman and Pearson would have looked with favor on this offspring of their forced marriage
- Hagfræðingar geta kynnt sér: The Cult of Statistical Significance, eftir Ziliak og McCloskey. Heilbrigðisstéttir: Why Most Published Research Findings Are False, eftir Ionnadis.

A medical example:

- Recently I attended a lecture on the connection of consumptions fish-liver-oil/fatty-oil (lýsi) and the risk for prostata cancer (Brasky et al., 2013).
- The study design was:
  1. A lot of data that needed some massage, pooling and coordination.
  2. Some runs of (a misspecified) Cox-proportional hazard model.
  3. Result: An odds-ration just above 1.
- My interpretation: Fish-liver-oil is quite safe.
- Other examples: Vaccinations of influenza and measels, genetically manipulated vegetables, etc.
- If you need 10.000 or more observations to get a "significant"impact, you may have increased support for your null hypothesis being true (almost true).

Beware of "metaanalysis"/big-data

- Some may want to increase "power"and merge many studies and pretend that their n just got larger.
- They may say: We are well aware of the Simpson paradox and therefore we have performed som I^2 tests. I find the I^2 test somewhat intransparent.
- Boos and Stefanski (2013) give an overview of the statistics of meta-analysis. My impression is that the Bayesian approach is more conservative.
- The model uncertainty has to be take account for. Do no take Cox-proportional hazard for granted.

Matsaðferðir, mesti-sennileiki ofl.

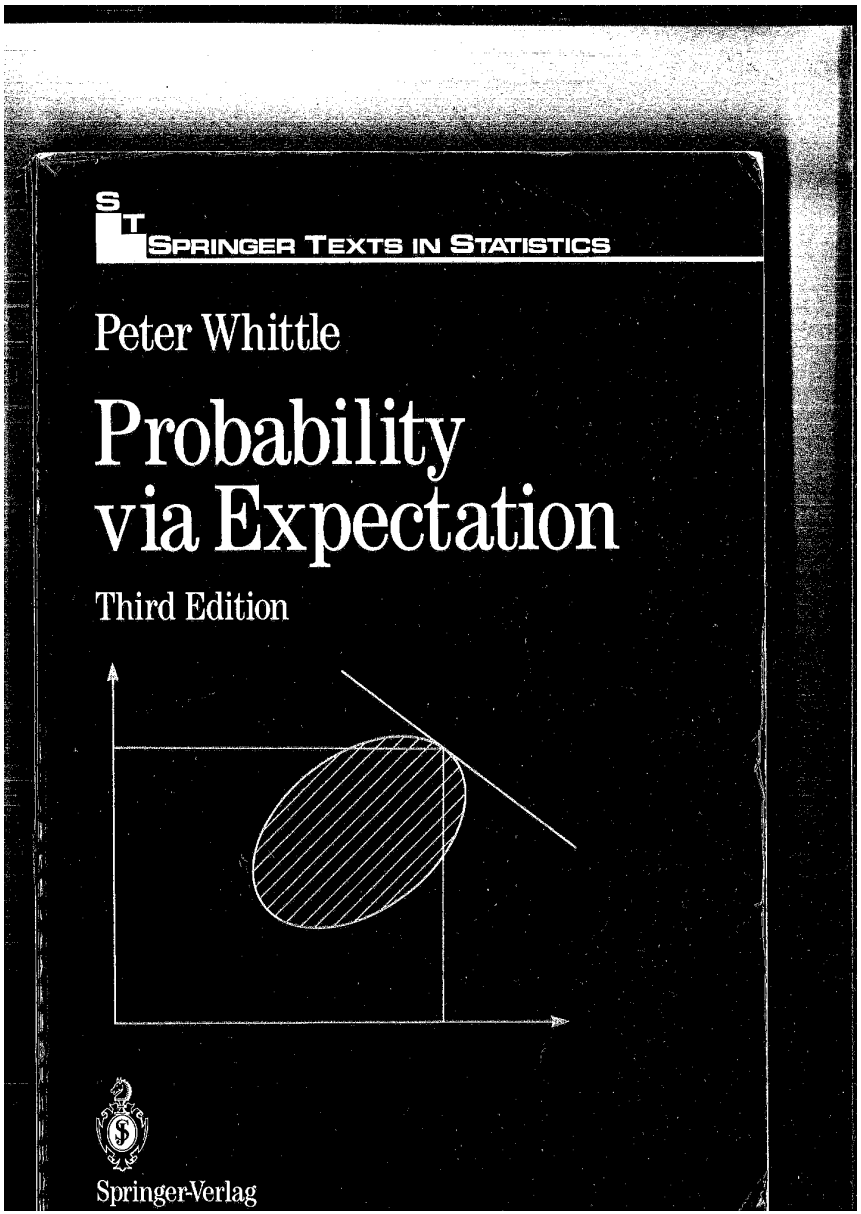
- Stærðfræðileg sönnun lá fyrir um að ML aðferð hefði optimal-eiginleika þegar fjöldi mælinga stefnir á óendanlegt.
- Ef meta á 1 eða 2 tvö meðaltöl í normal-líkani þá er ML einnig best mælt með MSE=mean-square-error.
- Það var því áfall þegar Stein sannaði upp úr 1950 að þegar vídd líkans væri 3 eða meira þá mætti finna betri metil.
- Í fyrstu héldu menn að þetta væri stærðfræðilegt *peculiarity* og að menn gætu aldrei skrifað niður formúluna.
- Upp úr 1960 fundu menn slíkar formúlur. Þær reyndust náskyldar Bayes-aðferðum, þ.e. eins konar *empirical-bayes*.

Framhaldsnám

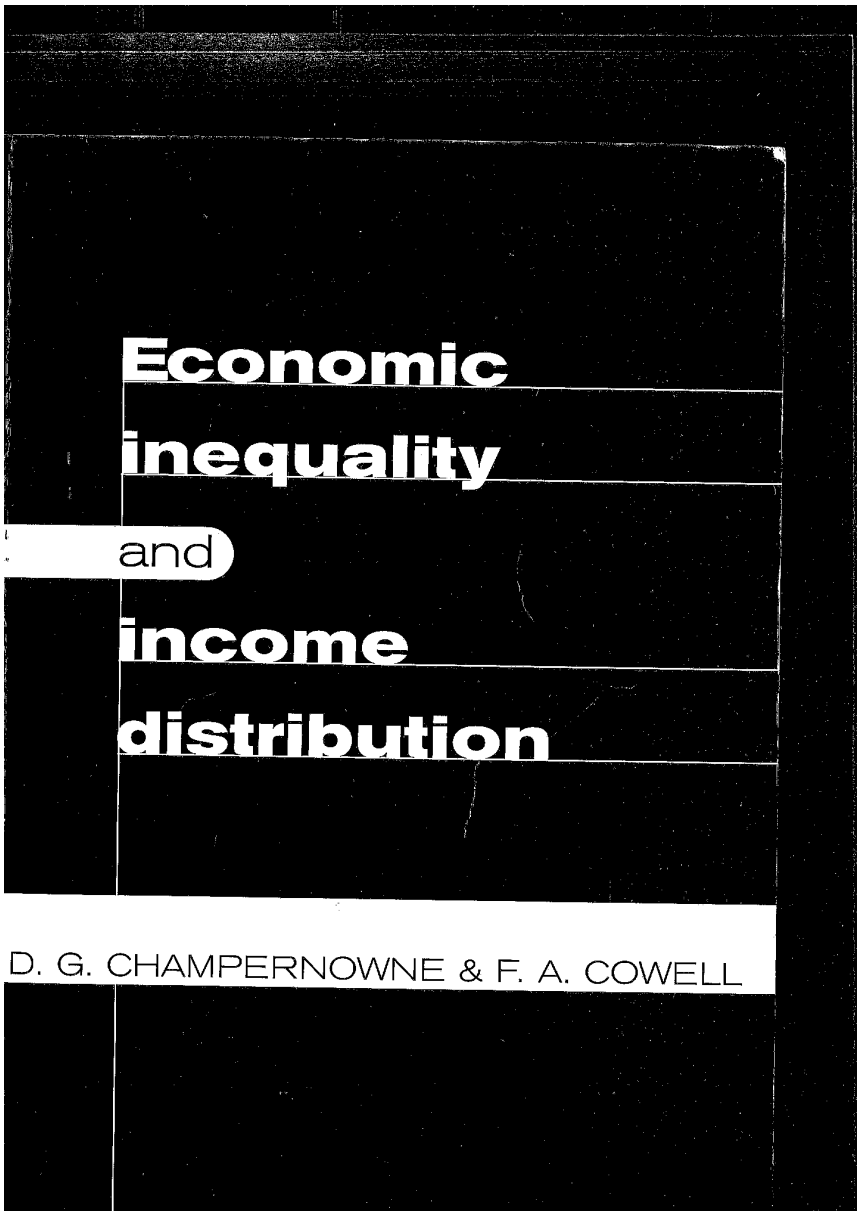
- Ég hafði lært stærðfræði, lært um tölfræði og tímaraðir, forritað tímaraðir o.s.frv. Tveir Guðmundar höfðu áhrif Guðmundur Guðmundsson og Guðmundur Magnússon.
- Fór í framhaldsnám í ekónometríu (hagrannsóknir) til Gautaborgar þar sem hinn frægi H. Wold hafði verið. Konunglegur ekónómetríuprófessor, Anders Klevmarken bauð mig sérstaklega velkominn.
- Herman Wold, hafði þróað ARMA líkön og haft tvo fræga doktorsnemendur, Jöreskog, og Whittle. Herman Wold hafði litla trú á þeirri ekónómetríu eftirstríðsáranna sem nefnd voru „*simulaneaous models*". Hann sagði að simultan líkön væru skemmtileg stærðfræði, hagfræðin væri bara ekki simultan. Hann ýtti vandamálinu til Jöreskog og útkoman varð LISREL:

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \Gamma\boldsymbol{\xi} + \boldsymbol{\zeta}.$$

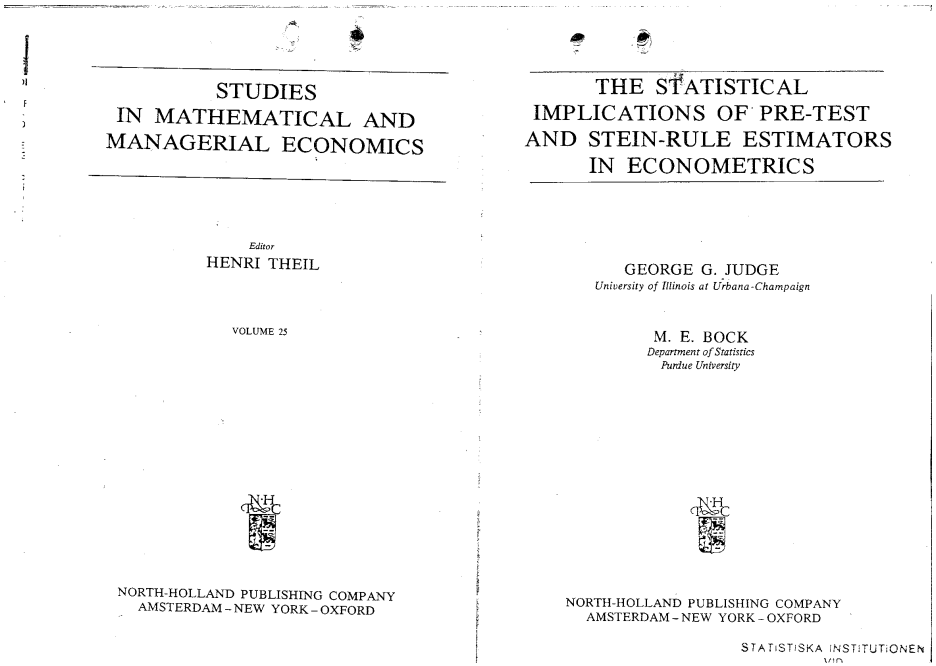
- Klevmarken var áhugasamur um míkróekónometríu, hvernig fólk eyðir tíma inni á heimilum o.s.frv.



Ég var látinn lesa fræga bók í líkindafræði eftir Whittle. Líkindafræðin er leidd út með því að byrja á að skilgreina væntanlega gildið í stað þess að byrja á að tala um hlutfallslega tíðni atburða í endurteknum tilraunum. Ætti að henta hagfræðingum og öðrum sem hugsa í væntingum vel. Inniheldur Markov líkan fyrir tekjudreifingar þar sem rökstutt er að tekjudreifingar hafi Pareto tail. Því var fyrsta hugmynd að doktorsverkefni fyrir mig um tekjudreifingar.



Champernowne var frægur Englendingur. Vinur Alan Turing. Þeir hönnuðu skáktölvu, TuroChamp ári 1948. Notaði hans nálgun í verkefni til að setja upp Markov-líkan fyrir tekjuþróun. Leiða út jafnvægisdreifingu o.s.frv. Empirískt verkefni með mikilli forritun.



STUDIES
IN MATHEMATICAL AND
MANAGERIAL ECONOMICS

Editor
HENRI THEIL

VOLUME 25

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM – NEW YORK – OXFORD

THE STATISTICAL
IMPLICATIONS OF PRE-TEST
AND STEIN-RULE ESTIMATORS
IN ECONOMETRICS

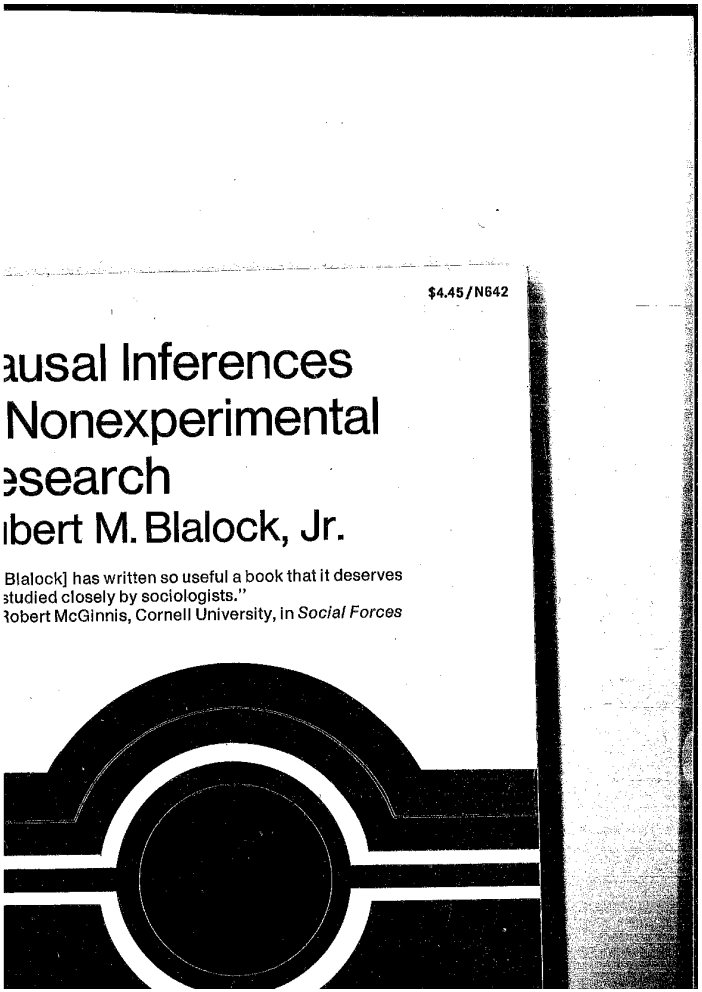
GEORGE G. JUDGE
University of Illinois at Urbana-Champaign

M. E. BOCK
Department of Statistics
Purdue University

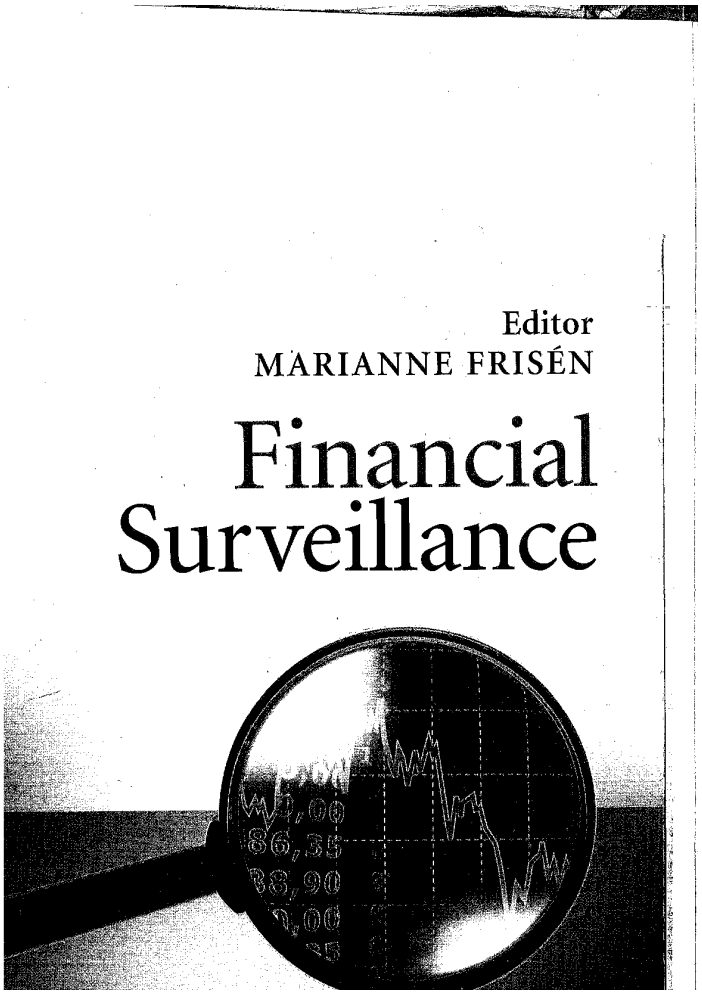
NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM – NEW YORK – OXFORD

STATISTISKA INSTITUTIONEN

Ég hafði forritað Box-Jenkins aðferðirnar og því taldi leiðbeinandi minn að þessar hugmyndir hentuðu mér. Markmiðið væri betri spár og hlutdrægni í parametermati því ekki svo heilagt. Fór í skiptiprógram til Purdue og kynntist sanntrúuðum Bayesistum.



Undir lok doktorsnámsins var ég látinn taka heimspekinámskeið í aðferðfræði félagsvísinda. Las m.a. þessa bók eftir Blalock. Judea Pearl vitnar í hana í book of why. Hermann Wold hafði birt greinar í Econometrica um causality. Einhverjir Ameríkanar köllluðu hina ekónómetrísku nálgun Wold, *Swedish school of causality*. Sbr. Nóbelsverðlaunahafa, Granger, Sims, ofl.



Leiðbeinandi minn var Marianne Frisén. Hún eftir m.a. unnið í vöktun „*surveillance*. Þróað m.a. kerfi til að vakta inflúensu ca. 2000. Á kafla í þessari bók um vöktun í fjármálum.

Pre-test (algengur praxís) og James-Stein

- Hugsanlegt að flétta saman test, þ.e. H_0:parameter=0, og metil
- Það eru til fræði um „pre-test" metla þ.e. að búinn er til metill sem er samsettur úr prófi og t.d. LS-metli, $\boldsymbol{b}$. Hugsum okkur einfalt línulegt líkan

$$\boldsymbol{y} = \boldsymbol{X}\beta + \varepsilon \quad E(\varepsilon) = 0 \quad E(\varepsilon\varepsilon') = \sigma^2\boldsymbol{I}$$

Ef kenningunni $H_0 : \boldsymbol{R}\beta = \boldsymbol{r}$ er ekki hafnað er notaður metillinn

$$\boldsymbol{b}^* = \boldsymbol{b} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}(\boldsymbol{R}\boldsymbol{b} - \boldsymbol{r})$$

Þetta má skrifa sem einn metil á þann hátt að:

$$\tilde{\beta} = I_{[0,c)}(u)\boldsymbol{b}^* + I_{[c,\infty)}(u)\boldsymbol{b}$$

Hér er u prófstærð sem segir til um hvort H_0 er hafnað og $I_{(s,r)}(u)$ er fall sem tekur gildið 1 ef $s < u < r$ og 0 annars.

- $MSE = mean - square - error = bias^2 + variance$, er mælikvarði á gæði metla.
- LS=least-squares er best-unbiased, þannig að framfarirnar eru í að minnka varíans mikið með smá bias. Pre-test metill hefur lítinn varíans ef H_0 er næstum rétt.

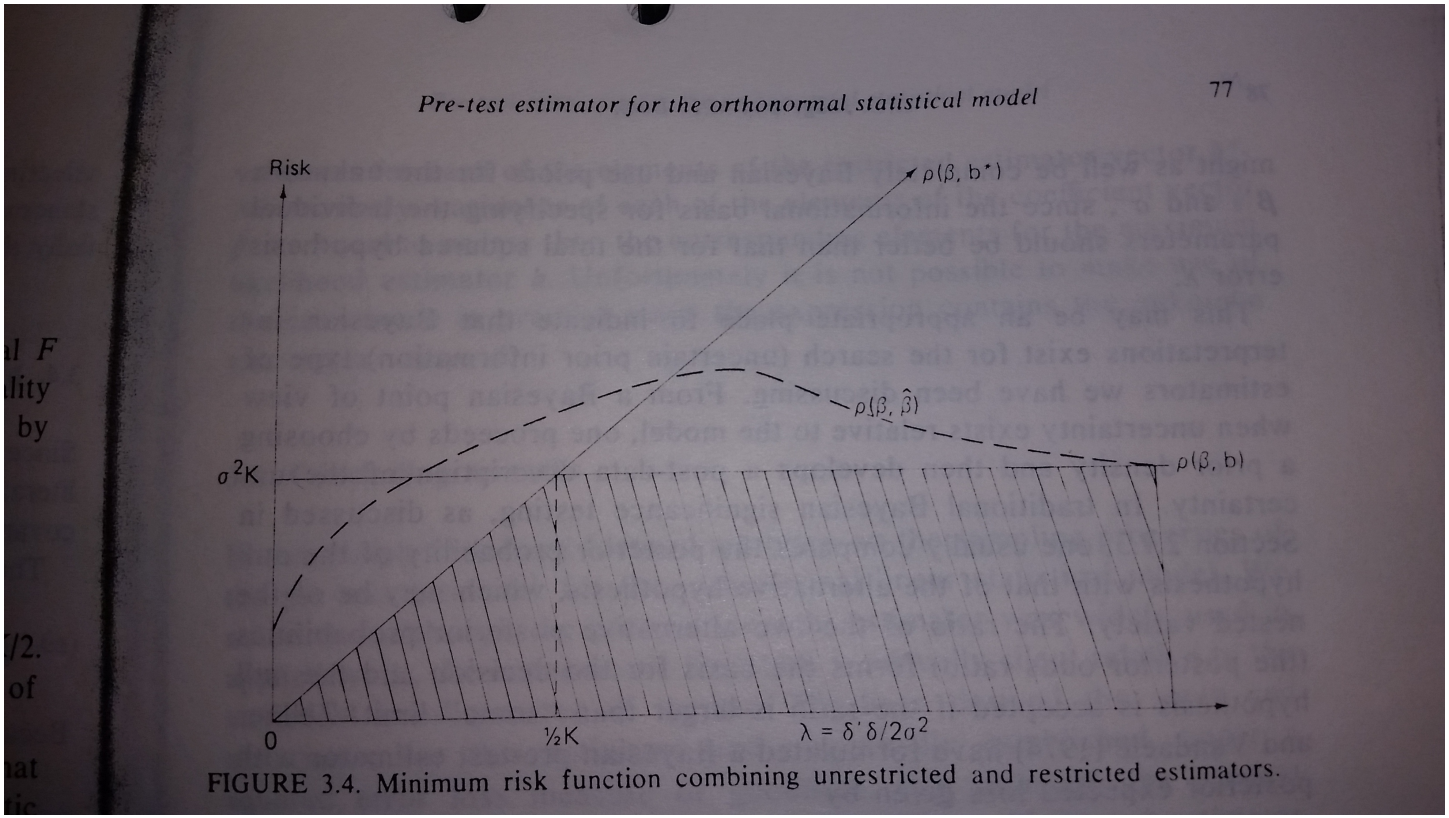


FIGURE 3.4. Minimum risk function combining unrestricted and restricted estimators.

Empirical-Bayes

- Pre-test atferlið er hlutdrægni í átt að H_0.
- Menn héldu lengi að fyrir línulegt líkan væri ekki hægt að finna metil sem hefði lægri *MSE* en least-squares metill
- Það kom því á óvart þegar Stein(1956) sýndi fram á tilvist slíks metils.
- Skömmu síðar fannst slíkur metill (James og Stein(1961)).

$$\hat{\beta}_{JS} = (1 - a/\boldsymbol{b}'\boldsymbol{b})\boldsymbol{b} \tag{1}$$

Þ.e. *OLS* metillinn er minnkaður aðeins. Ef $\sigma = 1$ má t.d. nota $a = K - 2$.

- E.t.v. er erfitt að sjá hvernig mönnum datt þetta í hug.
- Hægt er að rökstyðja formúlu (1) með reglu Bayes.

$$\boldsymbol{\theta} \sim N(0, \tau^2 I_K) \quad \text{a priori vissa um } \boldsymbol{\theta} \tag{2}$$

$$\boldsymbol{Z}|\boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \sigma^2 I_K) \tag{3}$$

líkan fyrir mælingu, Ef gert er ráð fyrir að $\sigma = 1$

þá er vissa að lokinni mælingu samkvæmt reglu Bayes

$$\theta|\boldsymbol{Z} \sim N(\boldsymbol{Z}(1 - 1/(\tau^2 + 1)), I_K/(1 + \tau^{-2})) \tag{4}$$

$$E(1/\chi^2(K)) = 1/(K-2)$$

þ.e.

$$E((K-2)/\boldsymbol{Z}'\boldsymbol{Z}) = 1/(\tau^2 + 1) \tag{5}$$

Væntanlegt gildi bayes-metilsins í jöfnunni er:

$$\boldsymbol{Z}(1 - 1/(\tau^2 + 1))$$

Ef τ er óþekkt og metið með jöfnu (5) fæst:

$$\hat{\boldsymbol{\theta}}_{EB} = (1 - (K-2)/\boldsymbol{Z}'\boldsymbol{Z})\boldsymbol{Z}$$

sem er hliðstætt James-Stein metlinum (1).

Umræða

- Merkilegt fyrir margra hluta sakir. Hér eru gögnin notuð tvisvar, a **big no-no in statistics**, bæði til að álykta um θ og til að giska á fyrirfram vissuna, varíans í prior.
- Á þessum árum voru sterkir flokkadrættir, Bayesista versus frequentistar. Stein-aðferðirnar voru sannanlega góðar í frequentist skilningi.
- Á seinni árum er miklu meira umburðarlyndi ríkjandi og flokkunin Bayesistar aðrir ekki eins afgerandi.
- Menn eru meira afslappaðir gagnvart Bayesismanum, bæði af heimspekilegum ástæðum og tæknilegum.
- Fordómar borga sig, betra að þeir séu nálægt réttu, bara ekki halda of fast í þá.

Smoothness-priors

- Mæli breytu Y_{at}, á aldri a á tíma t, $a = 1, \ldots, A$ og $t = 1, \ldots T$.
- Vil álykta um $E(Y_{at})$, þ.e. AT meðaltöl, mælingar eru AT.

$$E(Y_{at}) = \alpha_a + \beta_t + \gamma_{at}, \text{ til einföldunar óháðar og, } V(Y_{at}) = \sigma^2.$$

- Hugsum okkur að væntanlegu gildin séu puntar á samfelldur ferli $y(a, t)$. Einnig að ferillinn sé mjúkur:

$$\frac{\partial^2 y(a,t)}{(\partial t)^2} = 0, \quad \frac{\partial^2 y(a,t)}{(\partial a)^2} = 0, \quad \frac{\partial^2 y(a,t)}{\partial t \partial a} = 0.$$

- Í discrete aldri og tíma:

$$\Delta_t^2 y(a,t) = y(a,t) - 2y(a,t-1) + y(a,t-2) = 0,$$
$$\Delta_t^2 y(a,t) = y(a,t) - 2y(a-1,t) + y(a-2,t) = 0,$$
$$\Delta_{at} y(a,t) = (y_{at} - y_{a,t-1}) - (y_{a-1,t} - y_{a-1,t-1}).$$

- Þessa diffur-operatora máskrifa á fylkjaformi.

$$D_a = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \ddots & \cdots & 0 \\ \vdots & \vdots & \cdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & -2 & 1 \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_A \end{bmatrix},$$

- D_β er svipað, nema þvingað $\beta_1 = 0$,og fyrsta línan byrjar $(-2\ 1)$.
- D_γ er aðeins flóknar í venjulegri röð eru mest 0, tveir -1 og tveir 1.

- Hægt er að meta líkan með hliðarskilyrðum:

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} (\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}})$$
$$D_\alpha \boldsymbol{\alpha} = 0,$$
$$D_\beta \boldsymbol{\beta} = 0,$$

- Hlutafleiðuskilyrðin D_γ þarf ekki að nota.
- Þetta er jafngilt því að meta línulega trend í aldri og tíma.

$$Y_{at} = \mu + \theta_1 a + \theta_2 t + u_{at}, V(u_{at}) = \sigma^2.$$

Sýnidæmi úr R

```
n1=10
n2=8
aa=rep(1:n1,n2)
tt=vector(mode="numeric",length=0)
for (ii in (1:n2)) tt=c(tt,rep(ii,n1))
y=rnorm(n1*n2)+aa+2*tt
Da=Dasmooth(n1)
Db=Dbsmooth(n2)

xxmat=sparse.model.matrix(~-1+factor(aa)+factor(tt))

xxinv=solve(t(xxmat)%*%xxmat)
bhat=xxinv%*%t(xxmat)%*%y
R=bdiag(Da,Db)
betar=bhat-xxinv%*%t(R)%*%solve(R%*%xxinv%*%t(R))%*%R%*%bhat

lm(formula = y ~ aa + tt)
```

```
lm(formula = y ~ aa + tt)

Coefficients:
(Intercept)           aa           tt
     0.4185       0.9642       1.9631


diff(betar)[1:3]
[1] 0.9642157 0.9642157 0.9642157

diff(betar)[10:13]
[1] -10.060688   1.963123   1.963123   1.963123
```

Smoothness priors

- Hugmynd er að í stað $D_\alpha \boldsymbol{\alpha} = 0$ séu skilyrðin $D_\alpha \boldsymbol{\alpha} = \boldsymbol{\varepsilon}$, þar sem $\boldsymbol{\varepsilon}$ eru lítil.
- Svipað fyrir D_β og D_γ.
- Nota þetta til að setja fram prior um $E(Y_{at})$. Priorinn gengur út á að meðaltölin liggi á surfaci sem séu nálægt lausn á hlutafleiðujöfnukerfi(í aldri og tíma).
- Eins og í öllum hlutafleiðujöfnudæmum þarf randskilyrði, hér set ég $\beta_1 = 0$, $\gamma_{1t} = 0, t = 1, \ldots, T$ og $\gamma_{a1} = 0, a = 2, \ldots A$.
- Parameter sem met skal er:

$$\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_A, \beta_2, \ldots, \beta_T, \gamma_{22}, \ldots, \gamma_{AT})$$

- Til að lausn sé á hlutafleiðunni þarf einnig skilyrði á α_1, α_2 og β_2.
- Þessir 3 parametrar fá stóran a-priori varíance (eru metnir út frá mælingum). Hinir eru þvingaðir af smoothness skilyrðum.

Baysísk framsetning er því:

$$\boldsymbol{Y} = X\theta + \boldsymbol{u}, \quad \boldsymbol{u} \sim N(0, \sigma^2 I),$$

$$R_*\boldsymbol{\theta} \sim N(0, \Lambda),$$

$$R_* = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & \cdots \\ D_\alpha & & & 0 & 0 \\ 0 & 0 & & 1 & 0 \\ 0 & 0 & 0 & D_\beta & 0 \\ 0 & 0 & 0 & 0 & D_\gamma \end{bmatrix},$$

$$\Lambda = \begin{bmatrix} \lambda_1^2 & 0 & , \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2^2 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \lambda_\alpha^2 I & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \lambda_3^2 & 0 & 0 \\ 0 & \cdots & \cdots & 0 & \lambda_\beta^2 I & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & \lambda_\gamma^2 I \end{bmatrix},$$

i.e.

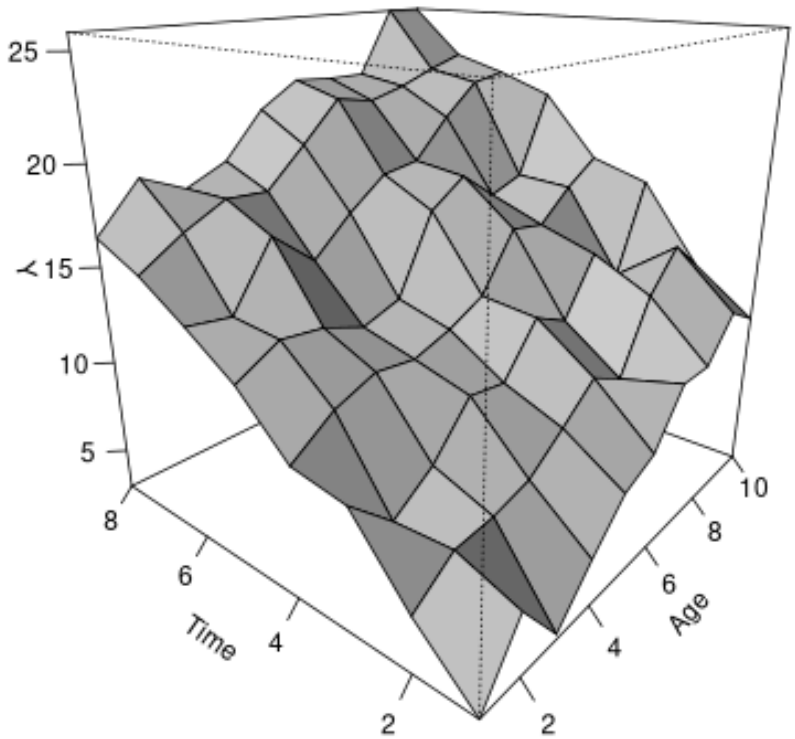
$$\boldsymbol{\theta} \sim N(0, R_*^{-1}\Lambda R_*'^{-1}).$$

- Hér er X grunnfylki (full-rank) fyrir myndrúmið, design-matrix, vektorar sem erum að mestu 0 og 1. Í örðum hagnýtum getur X t.d. verið byggt á margliðum,Fourier-föllum eða Wavelets.
- Væntanlegt gildi Bayes posterior dreifingarinnar fyrir $\boldsymbol{\theta}$ er:

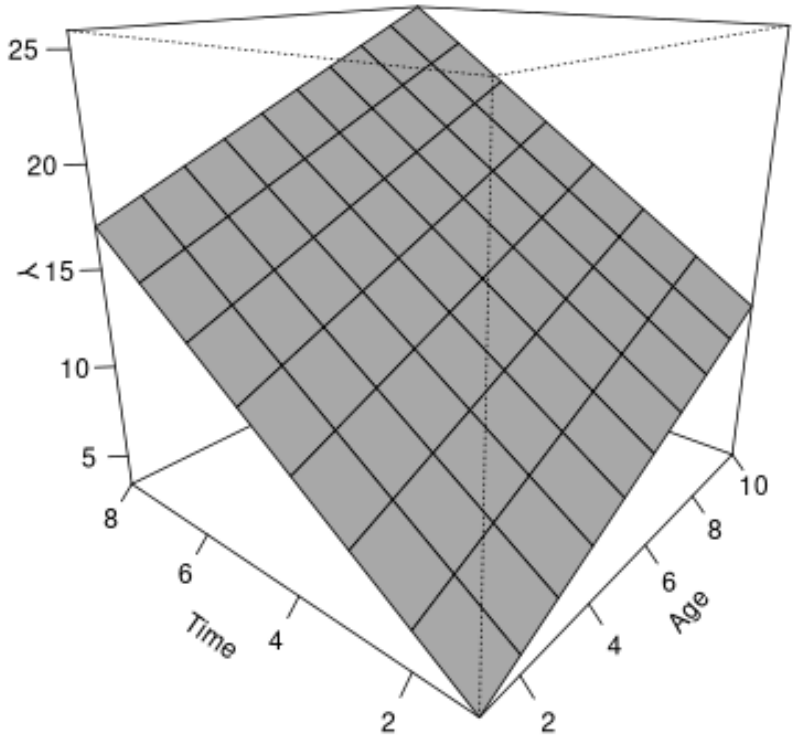
$$E_{Bayes}(\boldsymbol{\theta}) = (X'X + (R_*^{-1}\Lambda R_*'^{-1})^{-1})^{-1}X'\boldsymbol{y}$$

- Hér er gott að nota fylkin eru sparse til að spara minni og reiknitíma.

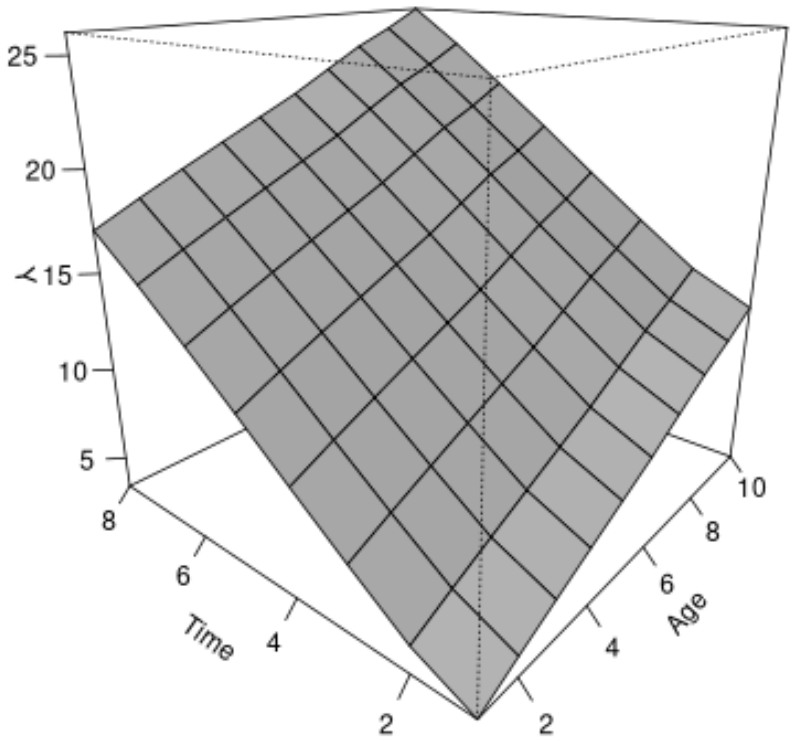
Nokkrar myndir



Mynd: Sýnidæmi 10x8



Mynd: Restricted 10x8



Mynd: Smoothness-prior

Hagnýtt dæmi/Högni óskarsson

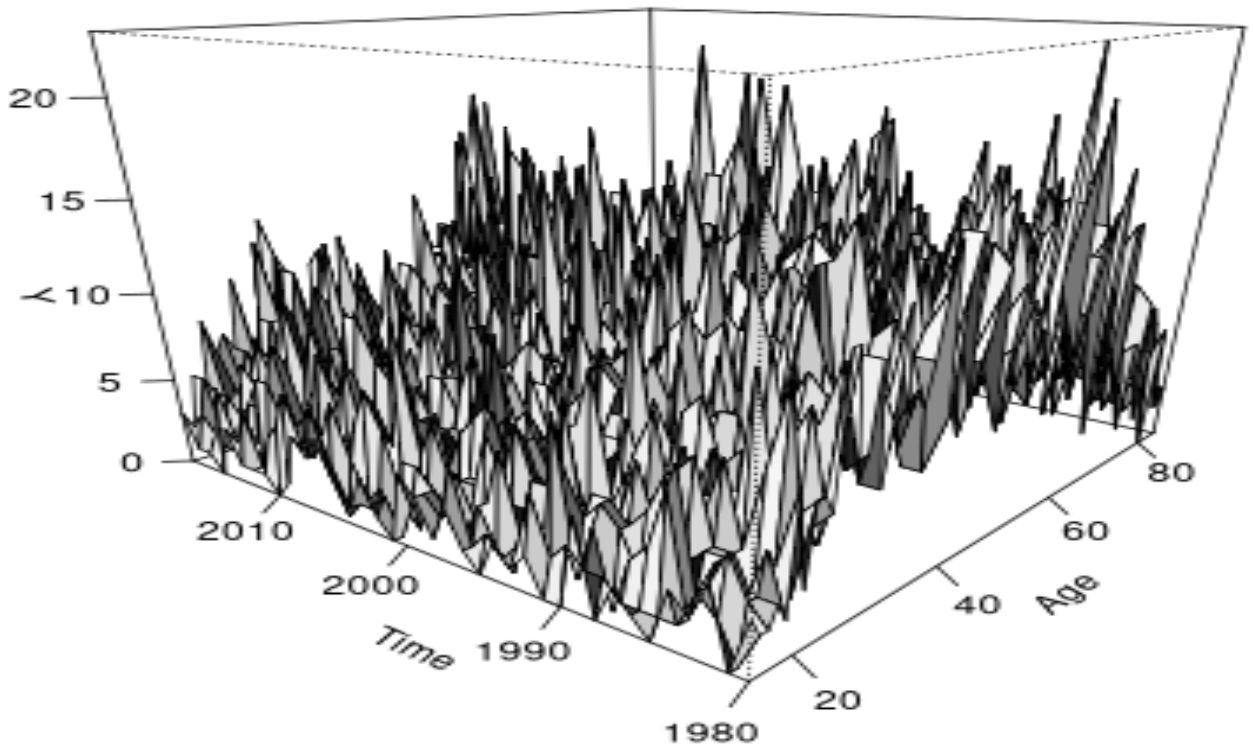
- Fyrir liggja gögn um tíðni sjálfsvíga fyrir 70 aldurshópa í 39 ár. Eðlilegt likan er:

$$Y_{at} = n_{at}\theta_{at} + \varepsilon_{at}.$$

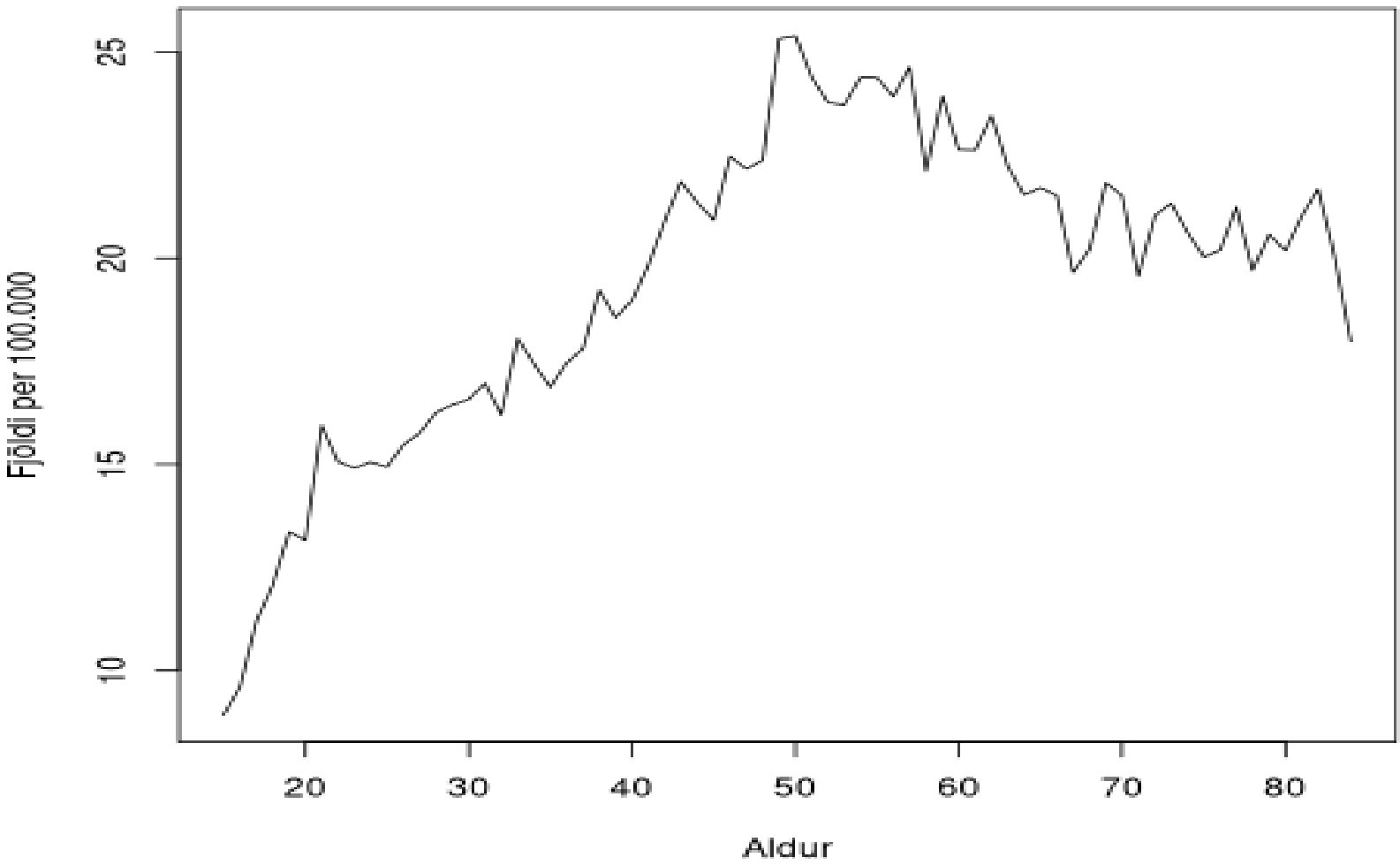
Þar sem Y_{at} er fjöldi sjálfsvíga á aldri a á tíma t. θ_{at} eru líkur á að einstaklingur á fremji sjálfsvíg á aldri a á tíma t. n_{at} er fjöldi einstaklinga á aldri a á tíma t. Meta þarf 2730 meðaltöl.

- X grunnfylkið inniheldur 1 eða 0 og röð at er margfölduð með $n_a t$.

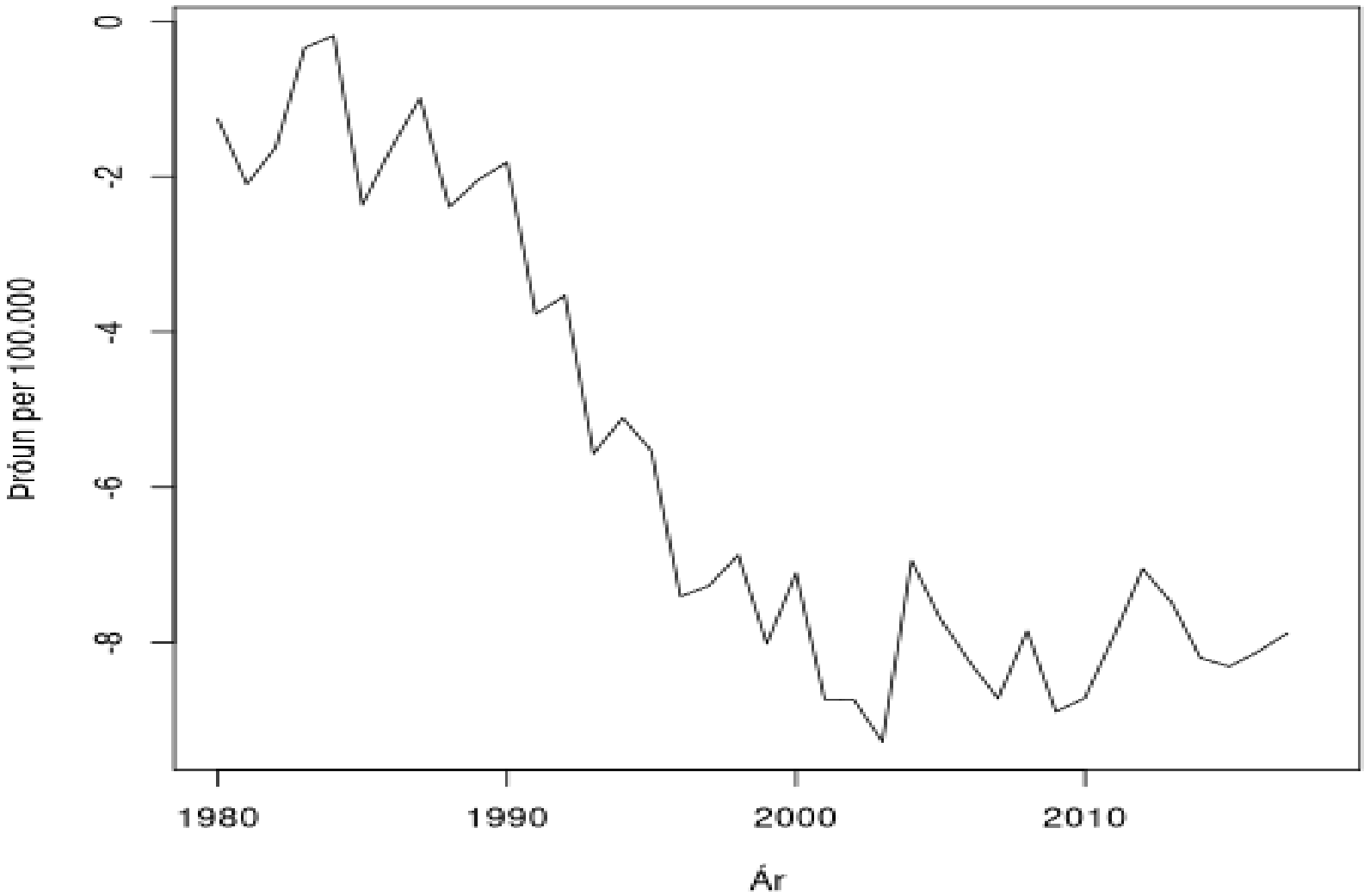
Hrágögn



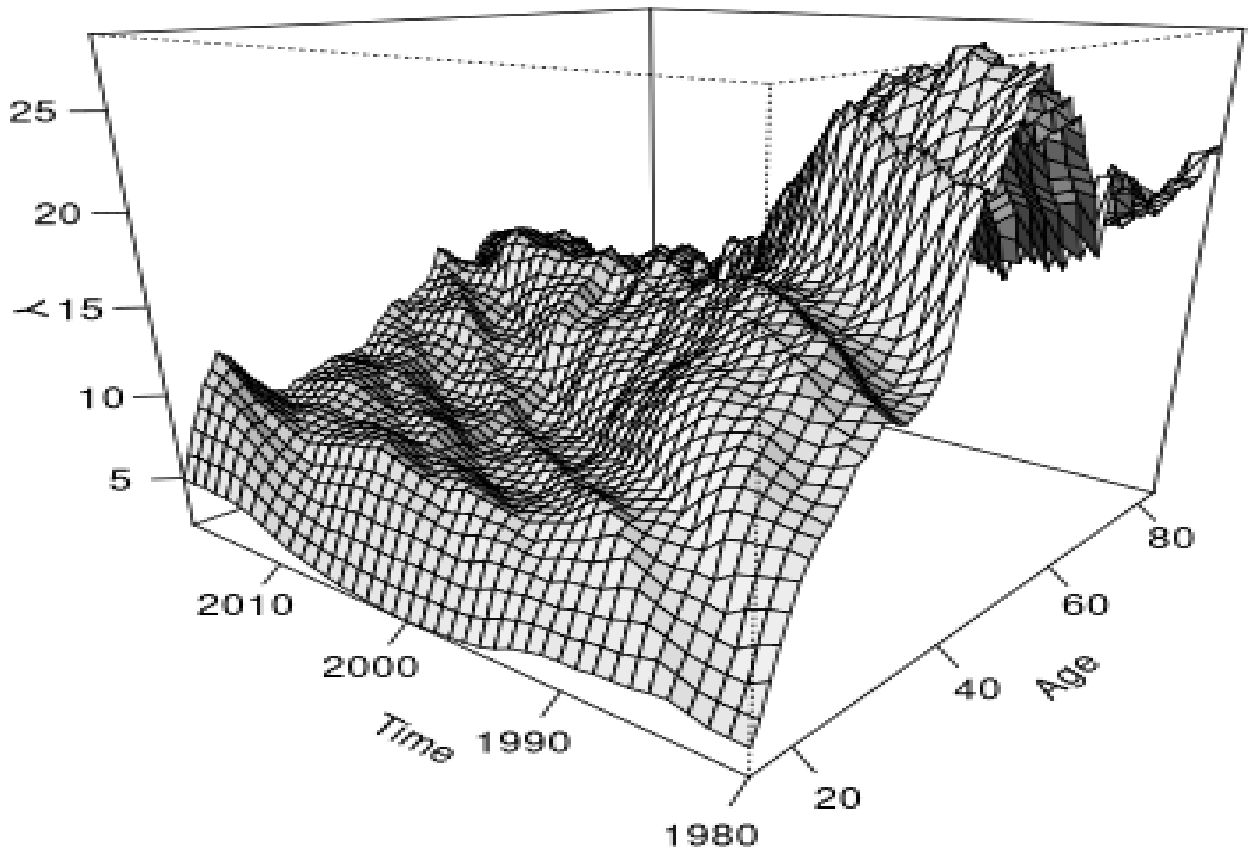
Mynd: Hrágögn sænskir karlar, mældur fjöldi



Mynd: Meðaltal per 100.000 eftir aldri



Mynd: Þróun í tíma per 100.000



Mynd: Smoothuð tíðni per 100.000, sænskir karlar, $\lambda_\alpha = 0.002$, $\lambda_\beta = 0.001$, $\lambda_\gamma = 0.0005$.

Lokaorð

- Smoothing stuðlar voru ekki optimeraðir.
- Hægt að reikna varíans fyrir gefna smoothing stuðla. Held að það sé erfiðara fyrir optimeraða smoothing stuðla.
- Overdispersion var hverfandi.
- Hér var notast við normal nálgun á binomial mælingum. Ef mæld tíðni er mjög lág gæti verið skynsamlegt að vinna með Poisson-módel beint.
- Kannski er betra að módellera cumulative-hazard.
- Afstaða mín til p-gilda er svipuð og til áfengis. Vil ekki banna þau, eins og amk. eitt sálfræðitímarit hefur gert. Nægir að vara við þeim.
- Bayesískar aðferðir hafa orðið reiknivænni með árunum. Trúarbragðagjáin hefur minnkað.
- Forrita of sjálfur til að tékka af stærðfræðilega niðurstöður (Fortran, C++/Rcpp, Gretl, R, Octave, Julia, Python/Pylab, ...)

Bayarri, M. and Berger, J. (1999). Quantifying surprise in the data and model verification. In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 6*. Clarendon Press Oxford.

Berger, J. O. (2003). Could fisher, jeffreys and neyman have agreed on testing? *Statistical Science*, 18(1):1–32.

Boos, D. and Stefanski, L. (2013). *Essential Statistical Inference: Theory and Methods*. Springer New York Heidelberg.

Brasky, T., Darke, A., Song, X., Tangen, C., Goodman, P., Thompson, I., Meyskens Jr., F., Goodman, G., Minasian, L., Parnes, H., Klein, E., and Kristal, A. (2013). Plasma phospholipid fatty acids and prostate cancer risk in the select trial. *Journal of the National Cancer Institute*, 105(15):1132–1141.

Hubbard, R. and Lindsay, R. M. (2008). Why p values are not a useful measure of evidence in statistical significance testing. *Theory & Psychology*, 18(1):69–88.

Lindley, D. V. (1999). Comment Bayarri and Berger: Quantifying surprise in data. In Dawid, A. and Smith, A., editors, *Bayesian Statistics 6*, volume 6, page 75. Oxford: Clarendon.

Spanos, A. (1999). *Probability Theory and Statistical Inference*. Cambridge University Press.

Young, G. and Smith, R. (2005). *Essential of Statistical Inference*. Cambridge University Press.