Bayes, Confidence Distributions, and US Pre-War Economy

Nils Lid Hjort / Stability and Change 2022-2023, CAS

NSF Oslo møte, 9/v/2023

Pre-talk comment: Stability and Change at CAS

We've had six workshops, with two core groups, the Stat People (the Nils crowd) and the PRIO people (the Håvard crowd):

0 PreCas Workshop (May 11-12, 2022)

1 ProcMod, From Processes to Models (Oct 19-20, 2022)

2 PredUnc, Prediction with Uncertainty (Dec 6-7, 2022)

3 N000, Counting the Uncounted (22 Feb 2023)

4 ChangeTrend, Changes, Trends, Windows (Mar 28-29, 2023)

5 DemoIndex, Democracy Indexes (Apr 26-27, 2023)

* Follow-up Workshop (Oct 2023)

Open, friendly, conducive atmosphere, international participants from both statistics and political science – lots o' common ground, and more to come.

This talk: mini-summary & footnote

Christopher Sims won the 2011 Sveriges Riksbank Prize in Economic Sciences im Memory of Alfred Nobel. For his Stockholm acceptance speech (and paper): showcased Bayes and MCMC, for a particular dataset and macroeconomic vector time series model, to reach inference for a few crucial parameters.

I use Confidence Distributions instead (same data, same model, but no priors). Our analyses differ significantly (and I claim I'm right).

- ♠ I like to think that I mostly give new talks about new things.
- ♠ But 75% of today's theme is inside Confidence, Likelihood, Probability (Schweder and Hjort, 2016, Ch. 14); also, this was essentially Tore's idea, analysis, reporting.
- ♠ But I have redone both the Bayes and the CD analyses, have 25% more material, and understand issues better. Will land in Statistical Inference: 666 Exercises, 66 Stories (Hjort and Stoltenberg, 2023).

C.A. Sims won the ... Sveriges Riksbank Prize in 2011, with T.J. Sargent. He praises Haavelmo, and uses modern Bayes.

Plan

A The macroeconomic dataset: (C_t, I_t, G_t), for $t = 1929, \ldots, 1940$ (Consumption, Investment, Government spending).

B The Sims model: 6 regression parameters, 3 variance parameters. The Bayes with MCMC method, and results: $B(\theta_1 \,|\, \mathrm{data})$ (and for other parameters).

C Confidence Distributions, what they are, how to compute them, t-bootstrapping.

D When Bayes and CD tend to agree – and when they're not.

E Finding $C(\theta_1 \,|\, \mathrm{data})$ (and CDs for other parameters).

F Remarks.

A: Consumption, Investment, Government spending

The Sims prewar macroeconomic dataset, for (C_t, I_t, G_t):

year	C	I	G
1929	736.3	101.4	146.5
1930	696.8	67.6	161.4
1931	674.9	42.5	168.2
1932	614.4	12.8	162.6
1933	600.8	18.9	157.2
1934	643.7	34.1	177.3
1935	683.0	63.1	182.2
1936	752.5	80.9	212.6
1937	780.4	101.1	203.6
1938	767.8	66.8	219.3
1939	810.7	85.9	238.6
1940	752.7	119.7	245.3

The unit is 1 billion 1932 dollars.

B: the Sims model (with priors)

Vector autoregressive time series model (translated from his paper):

$$\begin{aligned} C_t &= \beta_0 + \beta_1(C_t + I_t + G_t) + \sigma_C\,\varepsilon_{1,t}, \\ I_t &= \theta_0 + \theta_1(C_t - C_{t-1}) + \sigma_I\,\varepsilon_{2,t}, \\ G_t &= \gamma_0 + \gamma_1 G_{t-1} + \sigma_G\,\varepsilon_{3,t}, \end{aligned}$$

with the $\varepsilon_{j,t}$ being i.i.d. standard normal.

Such models, with such data, also called multiple structural equation models, started with Haavelmo (1943, 1944).

So 6 regression parameters and 3 variance parameters.

Sims further says $\theta_1 \geq 0$, and also $\gamma_1 \leq 1.03$, $1 - \beta_1(1 + \theta_1) > 0$: model does not explode.

Sims uses Bayes, with

(i) flat priors on the line for $\beta_0, \beta_1, \theta_0, \theta_1, \gamma_0, \gamma_1$, subject to these constrains;

(ii) $\sigma_1, \sigma_2, \sigma_3$ having independent $1/\sigma_j$ priors.

Likelihood for the Sims model

With $y_t = (C_t, I_t, G_t)^{\mathrm{t}}$, rewriting the model in matrix form:

$$H_0 y_t = c + H_1 y_{t-1} + \varepsilon_t \quad \text{for } t = 1, \ldots, n,$$

with the $\varepsilon_t \sim f_0$. Translation is ok with

$$H_0 = \begin{pmatrix} 1-\beta_1, & -\beta_1, & -\beta_1 \\ -\theta_1, & 1, & 0 \\ 0 & 0 & 1 \end{pmatrix}, \; H_1 = \begin{pmatrix} 0, & 0, & 0 \\ -\theta_1, & 0, & 0 \\ 0, & 0, & \gamma_1 \end{pmatrix}, \; c = \begin{pmatrix} \beta_0 \\ \theta_0 \\ \gamma_0 \end{pmatrix}.$$

Then from $y_t = H_0^{-1} z_t$, with $z_t = c + H_1 y_{t-1} + \varepsilon_t$ having density $f_t(z_t \,|\, y_{t-1})$, likelihood is

$$\begin{aligned} L &= \prod_{t=1}^n g(y_t \,|\, y_{t-1}) \\ &= \prod_{t=1}^n f_t(H_0 y_t \,|\, y_{t-1}) |H_0| = \prod_{t=1}^n f_0(H_0 y_t - c - H_1 y_{t-1}) |H_0|. \end{aligned}$$

For Sims case, $\varepsilon_t \sim \mathrm{N}_p(0, D)$, with diagonal D.

Bayes and posterior distributions

The above leads to log-likelihood

$$\ell = \sum_{t=1}^{n}\Big[\log|H_0| + \sum_{j=1}^{p}\{-\log\sigma_j - \tfrac{1}{2}\widetilde{\varepsilon}_{t,j}^2/\sigma_j^2\}\Big]$$

where $\widetilde{\varepsilon}_{t,j} = H_0 y_t - c - H_1 y_{t-1}$.

With priors $\pi_a(\alpha)$ and $\pi_s(\sigma)$ for $\alpha = (\beta_0, \beta_1, \theta_0, \theta_1, \gamma_0, \gamma_1)^{\rm t}$ and $\sigma = (\sigma_1, \sigma_2, \sigma_3)^{\rm t}$, posterior becomes

$$\pi(\alpha, \sigma \,|\, {\rm data}) \propto \pi_a(\alpha)\pi_s(\sigma)\exp\{n\log|H_0(\alpha)|\} \prod_{j=1}^{3}(1/\sigma_j)^n \exp\{-\tfrac{1}{2}Q_j(\alpha)/\sigma_j^2\},$$

with $Q_j(\alpha) = \sum_{t=1}^{n}\widetilde{\varepsilon}_{j,t}(\alpha)^2$. With independent $1/\sigma_j$ priors:

$$\pi(\alpha \,|\, {\rm data}) \propto \pi_a(\alpha)\exp\{n\log|H_0(\alpha)|\}\prod_{j=1}^{3} 1/Q_j(\alpha)^{n/2}.$$

MCMC for posterior of $\beta_0, \beta_1, \theta_0, \theta_1, \gamma_0, \gamma_1$

So I have $\pi(\alpha \,|\, \mathrm{data})$ for $\alpha = (\beta_0, \beta_1, \theta_0, \theta_1, \gamma_0, \gamma_1)^{\mathrm{t}}$. I set up a good MCMC in dimension 6. Flat prior, but constraints

$$\theta_1 \geq 0, \quad \gamma_1 \in [0, 1.03], \quad 1 - \beta_1(1 + \theta_1) > 0.$$

From α^{old}, proposal α^{prop} a symmetric push away, accepted with probability

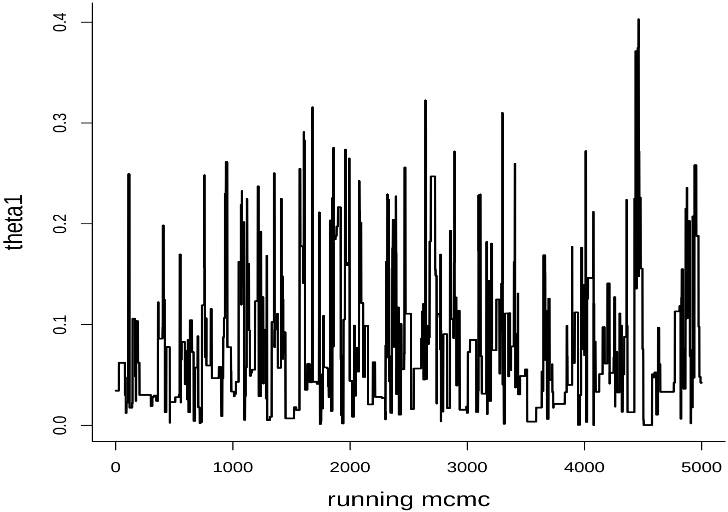
$$\mathrm{pr}_{\mathrm{accept}} = \min\Big\{1, \frac{\pi(\alpha^{\mathrm{prop}} \,|\, \mathrm{data})}{\pi(\alpha^{\mathrm{old}} \,|\, \mathrm{data})}\Big\}.$$

So $\alpha^{\mathrm{new}} = \mathrm{ok}\, \alpha^{\mathrm{prop}} + (1 - \mathrm{ok})\, \alpha^{\mathrm{old}}$, with $\mathrm{ok} = 1(\mathrm{accept})$.

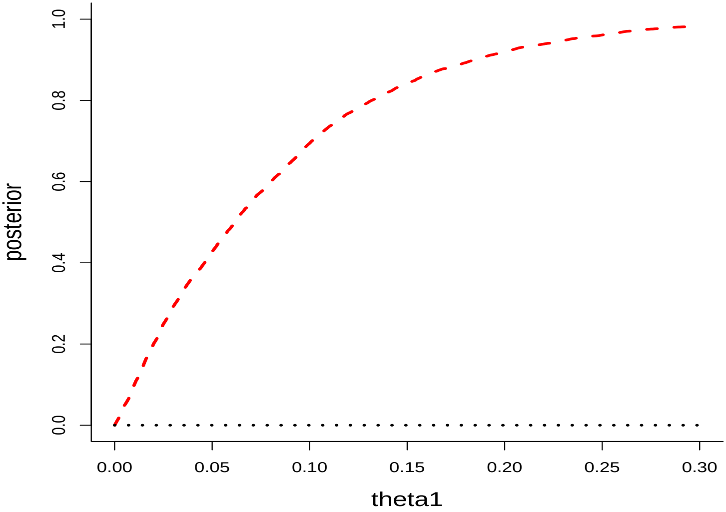
This works well (and is essentially what Sims did). Can in particular read off posteriors for the more crucial parameters, like θ_1 and γ_1.

Can also use this machinery for guided prediction (say 1941, 1942 based on 1929–1940 data; what happens to C_t, I_t, if government decides G_t?).

Running MCMC, in dimension 6, reading off for θ_1 (easy to let it run for 10^5 steps):



Reading off posterior $B(\theta_1 \mid \mathrm{data})$; median 0.06,
95% interval $[0, 0.24]$, etc. I've essentially re-done Sims (2012a).



C: Confidence Distributions

CDs are (almost) posteriors without priors.

Simpler versions go back to Fiducial Fisher 1930. Lots of modern developments, with connections in various directions; see Schweder and Hjort Confidence, Likelihood, Probability (2016).

With ϕ a focus parameter, and y the data, suppose $C(\phi, y)$ is a cdf in ϕ for any y, and that

$$U = C(\phi_{\rm true}, Y) \sim {\rm unif} \quad \text{at true position.}$$

Then

$$\Pr_{\rm true}(C^{-1}(0.05, Y) \le \phi_{\rm true} \le C^{-1}(0.95, Y)) = 0.90,$$

etc. So $C(\phi, {\rm data})$ acts like a frequentist posterior cdf for ϕ.

Also, confidence curve ${\rm cc}(\phi, {\rm data}) = |1 - 2\,C(\phi, {\rm data})|$ has the property

$${\rm CI} = \{\phi\colon {\rm cc}(\phi, {\rm data}) \le 0.90\} \quad \text{has} \quad \Pr_{\rm true}(\phi \in {\rm CI}) = 0.90$$

etc.

Computing the CD

With $\phi = g(\theta_1, \ldots, \theta_p)$ a focus parameter, various recipes, for different situations, in CLP.

Finding a pivot: Suppose $t = \{h(\widehat{\phi}) - h(\phi)\}/\widehat{\kappa}$ has distribution H, independent of parameters. Then

$$C(\phi, y) = 1 - H(\{h(\widehat{\pi}_{\rm obs}) - h(\phi)\}/\widehat{\kappa}_{\rm obs}).$$

Normal approximation: if $\widehat{\phi} \approx_d \mathrm{N}(\phi, \widehat{\kappa}^2)$, use

$$C(\phi, \mathrm{data}) = \Phi((\phi - \widehat{\phi})/\widehat{\kappa}).$$

There are various refinements.

Wilks theorems, with log-profile-likelihoods: via deviance $D(\phi) = 2\{\ell_{\rm prof}(\widehat{\phi}) - \ell_{\rm prof}(\phi)\}$, use

$$\mathrm{cc}(\phi, \mathrm{data}) = \Gamma_1(D(\phi)).$$

Again, refinements, via Bartletting etc.

CDs via t-bootstrapping

t-bootstrapping I: Suppose $t = (\widehat\phi - \phi)/\widehat\kappa$ has a distribution H approximately constant in the relevant parameter region. Then

$$C(\phi, y) = 1 - H((\widehat\phi_{\rm obs} - \phi)/\widehat\kappa_{\rm obs}),$$

where H may be computed by simulation of a high number of $t^* = (\widehat\phi^* - \widehat\phi)/\widehat\kappa^*$.

t-bootstrapping II: Write the full model parameter as (ϕ, γ). Write the cdf of $t = (\widehat\phi - \phi)/\widehat\kappa$ as $H(\cdot, \phi, \gamma)$. Then

$$C_0(\phi, y) = 1 - H((\widehat\phi_{\rm obs} - \phi)/\widehat\kappa_{\rm obs}, \phi, \gamma)$$

would've been perfect (but we can't use it in general). I use

$$\begin{aligned} C(\phi, y) &= 1 - H((\widehat\phi_{\rm obs} - \phi)/\widehat\kappa_{\rm obs}, \phi, \widehat\gamma_{\rm obs}(\phi)) \\ &= {\rm Pr}_*((\widehat\phi^* - \widehat\phi_{\rm obs})/\widehat\kappa^* \ge (\widehat\phi_{\rm obs} - \phi)/\widehat\kappa_{\rm obs}). \end{aligned}$$

For each ϕ, simulate a high number of $t^* = (\widehat\phi^* - \widehat\phi_{\rm obs})/\widehat\kappa^*$.

D: Bayes and CDs often agree (but not always)

In regular, smooth models, with reasonable

$$\text{ratio} = (\text{data volume}) \;/\; (\text{model complexity}),$$

Bayes and CDs will be in reasonable agreement.

This is the terrain of Bernshteĭn–von Mises theorems, and also prior matching, and more.

In particular: a Bayesian machine produces intervals with ok coverage: if $\mathrm{BI}(0.90)$ is a Bayesian posterior 0.90 interval for some ϕ, then

$$\mathrm{Pr}_{\mathrm{true}}(\phi_{\mathrm{true}} \in \mathrm{BI}(0.90)) \approx 0.90,$$

etc.

It's not forbidden to be purist Bayes, with your own special prior, and its consequent special posterior – but for lots of science, be more objective and careful.

Boundary parameters

When parameters are close to boundaries: things are more messy, for frequentists and Bayesians: the Big Industrial Complex of approximate normality, Maximum Likelihood, Wilks theorems, AIC, BIC, FIC, Bayes (driving 90 percent of applied statistics?), does not work without care and modifications.

Case in point, here, prototype situation: $y \,|\, \theta \sim \mathrm{N}(\theta, 1)$, a priori $\theta \geq 0$. Perfect CD:

$$C(\theta, \mathrm{data}) = \Phi(\theta - y) \quad \text{for } \theta \geq 0,$$

with pointmass $\Phi(-y)$ at zero, etc. Bayes with flat prior:

$$b(\theta \,|\, y) = \phi(\theta - y)/\{1 - \Phi(-y)\} \quad \text{for } \theta \geq 0.$$

But the coverage can be far off (if θ_{true} small):

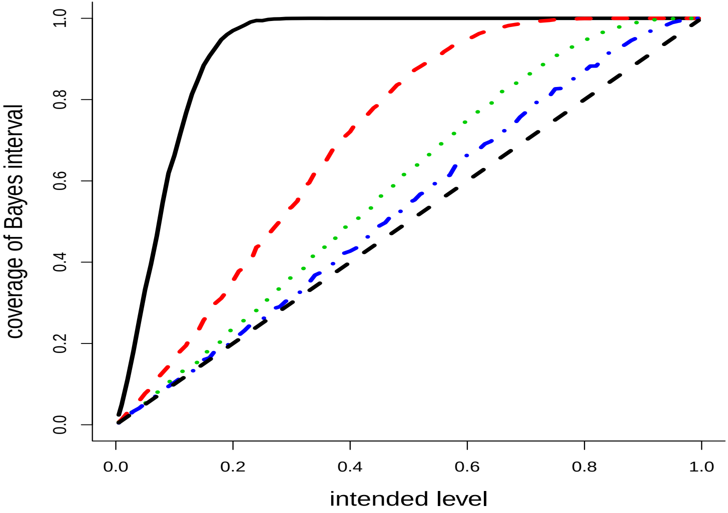
$$\mathrm{Pr}_{\mathrm{true}}(\theta_{\mathrm{true}} \leq B^{-1}(q \,|\, y)), \text{ with } B^{-1}(q \,|\, y) = \Phi^{-1}(1 - (1-q)\Phi(y))$$

is often seriously overshooting q.

We shall see, for θ_1: Sims' 0.75 interval has coverage 0.9999, etc.

Coverage $\Pr_{\rm true}(\theta_{\rm true} \in {\rm BI}(q))$, for levels $q = 0.01, \ldots, 0.99$, for cases $\theta_{\rm true} = 0.1, 0.5, 1.0, 1.5$.

The CD is perfect, regardless of $\theta_{\rm true}$.



E: Finding the CDs in Sims setup

First I estimate the $6 + 3$ parameters in the Sims model via ML.

With $\alpha = (\beta_0, \beta_1, \theta_0, \theta_1, \gamma_0, \gamma_1)^{\mathrm{t}}$, I profile out $\sigma_1, \sigma_2, \sigma_3$: starting from

$$\ell = n \log |H_0(\alpha)| + \sum_{j=1}^{3} \Big\{ -n \log \sigma_j - \tfrac{1}{2} \sum_{t=1}^{n} \widetilde{\varepsilon}_{t,j}^2 / \sigma_j^2 \Big\},$$

I reach

$$\ell_{\mathrm{prof}}(\alpha) = n \log |H_0(\alpha)| - n \sum_{j=1}^{3} \log \widehat{\sigma}_j(\alpha) - \tfrac{1}{2} 3n$$

with $\widehat{\sigma}_j(\alpha)^2 = Q_j(\alpha)/n$ and $Q_j(\alpha) = \sum_{t=1}^{n} \widetilde{\varepsilon}_{j,t}(\alpha)^2$.

This reduces numerical optimisation from $\dim = 9$ to $\dim = 6$.

With focus on θ_1 (similarly for other focus parameters): For each θ_1 on a grid: use

$$\widehat{H}_0 = \begin{pmatrix} 1-\widehat{\beta}_1, & -\widehat{\beta}_1, & -\widehat{\beta}_1 \\ -\theta_1, & 1, & 0 \\ 0 & 0 & 1 \end{pmatrix}, \ \widehat{H}_1 = \begin{pmatrix} 0, & 0, & 0 \\ -\theta_1, & 0, & 0 \\ 0, & 0, & \widehat{\gamma}_1 \end{pmatrix}, \ \widehat{c} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\theta}_0 \\ \widehat{\gamma}_0 \end{pmatrix}$$

to simulate 10^3 datasets

$$y_t^* = \widehat{H}_0^{-1}(\widehat{c} + \widehat{H}_1 y_{t-1} + \varepsilon_t) \quad \text{for } t = 1, \ldots, n$$

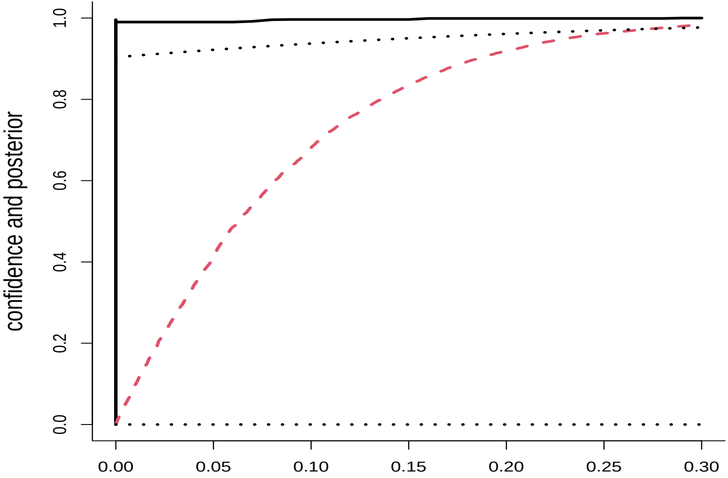
with y_0 fixed to values for 1929. For each simulated dataset, estimate 6 + 3 parameters, and compute $t^* = (\widehat{\theta}_1^* - \widehat{\theta}_{1,\mathrm{obs}})/\widehat{\kappa}^*$. In the end I read off

$$C(\theta_1, \mathrm{data}) = \mathrm{Pr}_*(t^* \geq (\widehat{\theta}_{1,\mathrm{obs}} - \theta_1)/\widehat{\kappa}_{1,\mathrm{obs}}).$$

Quite heavy computing, as I need nontrivial ML estimation job for 10^3 new datasets, for each θ_1.

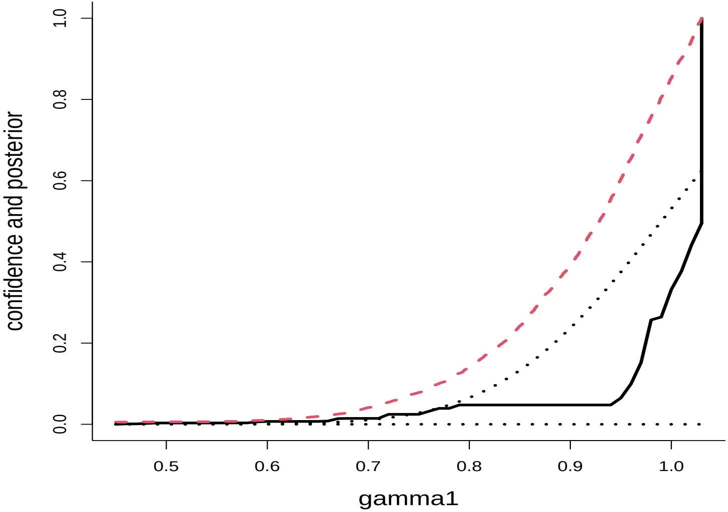
Bayes and CD for θ_1

Result: CD for θ_1 has pointmass $C(0, \mathrm{data}) = 0.989$ at zero. Figure shows Bayes posterior; good CD (very close to 1); simple normal approximation CD. Sims does not detect that θ_1 is very likely zero; I am 99% confident about this.



Bayes and CD for γ_1

Prior constraint: $\gamma_1 \leq 1.03$ (decided by Sims). Bayes (with the Sims priors) does not detect that γ_1 is very likely in $[0.95, 1.03]$. The CD finds this.



Bayes and CD for other parameters

The Sims model takes

$$\begin{aligned} C_t &= \beta_0 + \beta_1(C_t + I_t + G_t) + \sigma_C\,\varepsilon_{1,t}, \\ I_t &= \theta_0 + \theta_1(C_t - C_{t-1}) + \sigma_I\,\varepsilon_{2,t}, \\ G_t &= \gamma_0 + \gamma_1 G_{t-1} + \sigma_G\,\varepsilon_{3,t}. \end{aligned}$$

Sims says θ_1 is crucial, and gives parameter constraints

$$\theta_1 \ge 0, \quad \gamma_1 \in [0, 1.03], \quad 1 - \beta_1(1+\theta_1) > 0.$$

For the other 4 parameters, Bayes (with MCMC) and CD (with t-bootstrapping) give about the same results; the drama is for θ_1 and γ_1, due to the boundary parameter effect.

If θ_1 and γ_1 had been a bit away from boundaries (to the right of 0, to the left of 1.03), there would have been no drama (and Tore + Nils would have not have detected any trouble).

F: Remarks

- So unlike Sims, I am 99% confident that increase in consumption does not affect investment.
- In the $6+3$ parameter model, Sims imposes the $\theta_1 \geq 0$ restriction. Constrained ML is $\widehat{\theta}_1 = 0$:

	β_0	β_1	θ_0	θ_1	γ_0	γ_1
Sims	166	0.566	63.00	0.00	10.70	0.991
Nils	201	0.525	63.88	0.00	10.79	0.990

  Unconstrained ML is $\widehat{\theta} = -0.566$ (with sd approx 0.434).
- Same phenomenon in regression models: If $y_i = a + b_1 x_{i,1} + b_2 x_{i,2} + \varepsilon_i$, and postulate $b_2 \geq 0$, then the perfect CD is

$$C(b_2, \mathrm{data}) = G_{\mathrm{df}}((b_2 - \widehat{b}_2)/\widehat{\kappa}_2) \quad \text{for } b_2 \geq 0,$$

  with G_{df} the cdf for t_{df}, here $\mathrm{df} = n - 3$. A Bayes setup needs to factor in that perhaps $b_2 = 0$. Same in logistic regressions, Poisson regression, etc.

- There is a priori nothing wrong with Bayes – but using flat priors on $[0, \infty)$ for nonnegative parameters may go wrong: ok if some distance away from zero, but very wrong if truth is close to zero. If $\theta = 0$, you do not detect it! CDs do the job.

- The Bayes analysis becomes more clever (and agrees more with the CDs) with more clever priors. May try $\theta_1 \sim \frac{1}{2}\delta_0 + \frac{1}{2}I_{(0,\infty)}$ (will do so, later, for Sims Story for Hjort-Stoltenberg 2023 book; it takes a more complicated MCMC setup).

- For prototype setup, $y \,|\, \theta \sim \mathrm{N}(\theta, 1)$, our CD is optimal. With a prior $\mathrm{d}\pi(\theta)$, posterior is

$$\mathrm{d}\pi(\theta \,|\, y) = \mathrm{d}\pi(\theta)\phi(\theta - y)/c(y) \quad \text{for } \theta \geq 0.$$

  But there is no prior succeeding in giving a posterior identical to the CD.

- Would be useful to develop a FIC for vector autoregressive time series models.

(Some) references

G Claeskens, NL Hjort (2008). *Model Selection and Model Averaging.* CUP.

C Cunen, NL Hjort (2022). Combining information across diverse sources: the II-CC-FF paradigm. *Scandinavian Journal of Statistics.*

NL Hjort, EAa Stoltenberg (2023). *Statistical Inference: 666 Exercises, 66 Stories (and Solutions to All).* CUP.

T Haavelmo (1943). The statistical implications of a system of simultaneous equations. *Econometrics.*

T Haavelmo (1944). The Probability Approach in Econometrics. *Econometrica* (115 pages + supplement!).

T Schweder, NL Hjort (2016). *Confidence, Likelihood, Probability.* CUP.

T Schweder (2018). Bayesian Analysis: Always and Everywhere? *FocuStat Blog Post.*

C Sims (2012a). Statistical modeling of monetary policy and its effects [Sveriges Riksbank Prize in Memory of Alfred Nobel lecture]. American Economic Review.

C Sims (2012b). Appendix: inference for the Haavelmo model. Technical report, Puplic Policy & Finance, Princeton University, Princeton, NJ.