Normalization in Econometrics

James Hamilton, Dan Waggoner and Tao Zha

Econometric Reviews

What do we mean by Normalization?

Restricting the parameter space to achieve identification
Example:

- toy model:

$$y_t = \sigma \varepsilon_t \tag{1}$$

  with $\varepsilon_t \sim N(0, 1)$
  Log-likelihood is given by:

$$\log f(y; \sigma) = -T/2 \log(2\pi) - T/2 \log(\sigma^2) - \sum_{t=1}^{T} y_t^2/(2\sigma^2) \tag{2}$$

  Here the 'sensible' normalization would be $\sigma > 0$

- Note: this paper only deals with normalization issues in likelihood-based inference

What is the Problem?

Picking a certain normalization not only amounts to ’picking’ a point estimate, but also what area of the parameter space is used to compute confidence intervals.

> *A poor normalization can lead to multimodal distributions, disjoint confidence intervals, and very misleading characterizations of the true statistical uncertainty*

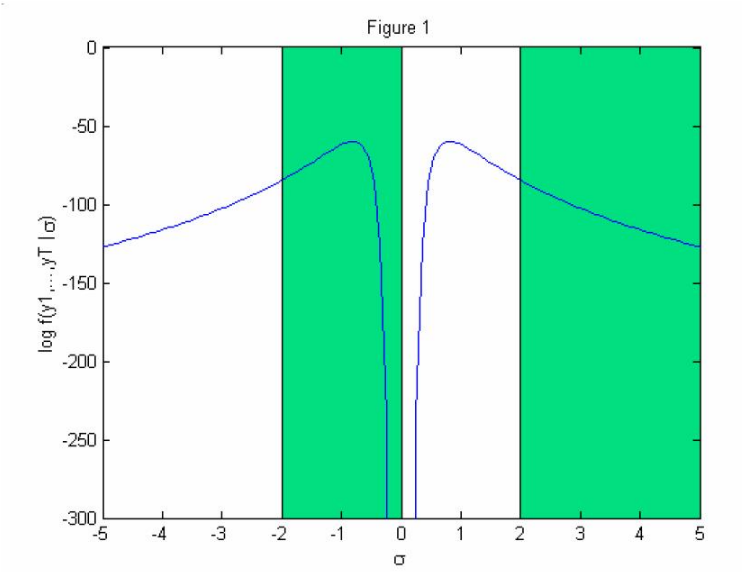


Figure 1
0
-50
-100
-150
-200
-250
-300
-5
-4
-3
-2
-1
0
1
2
3
4
5
σ
log f(y1,...,yT | σ)

Some Definitions

Let $\theta \in \mathcal{R}^k$ be the parameter vector of interest and $f(y;\theta)$ the likelihood function

- two parameter vectors θ_1 and θ_2 are observationally equivalent if

$$f(y;\theta_1) = f(y;\theta_2)\forall y \tag{3}$$

- the model is said to be locally identified at θ_0 if there exists an open neighborhood around θ_0 containing no other value of θ that is observationally equivalent to θ_0

The Identification Principle

There will usually be loci along which the model is locally unidentified because the information matrix is singular or the likelihood is 0 ($\sigma = 0$ in the previous example). The following *identification principle* rules out 'bad' normalizations:
The set of admissible parameter values should be chosen in such a way that the loci along which the model is locally unidentified or the likelihood is 0 form the boundary of this set.

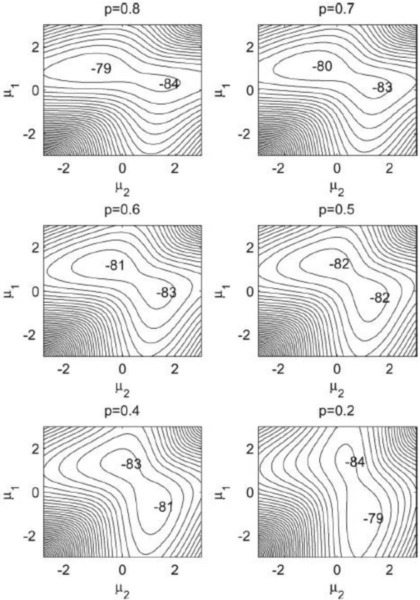
Example 1: Mixture of Normals

$$f(y_t;\mu_1,\mu_2,p) = \frac{p}{\sqrt{2\pi}} exp\left(\frac{-(y_t-\mu_1)^2}{2}\right) + \frac{1-p}{\sqrt{2\pi}} exp\left(\frac{-(y_t-\mu_2)^2}{2}\right) \tag{4}$$

Two possible normalization schemes have appeared in the literature:

- $p > 0.5$
- $\mu_1 > \mu_2$

Example: 50 observations, $\mu_1 = 1$,$\mu_2 = -1$ and $p = 0.8$


p=0.8
-79
-84
p=0.7
-80
-83
p=0.6
-81
-83
p=0.5
-82
-82
p=0.4
-83
-81
p=0.2
-84
-79
μ_1
μ_2

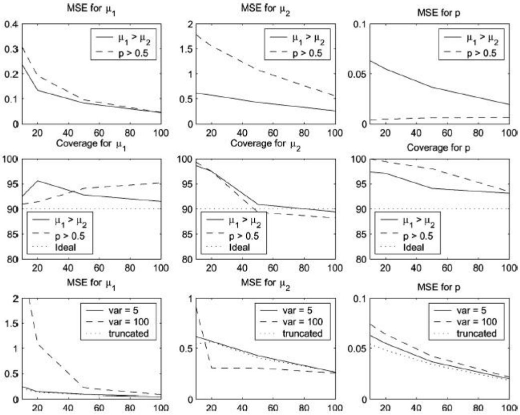


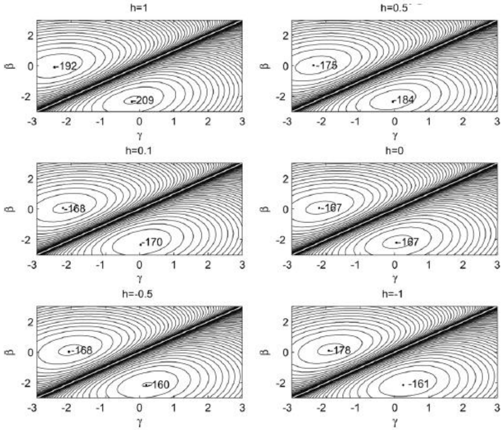
FIGURE 2 Performance of estimators with sample size on the horizontal axis. Top row: average squared difference between posterior mean and true value for indicated parameter when normalized by $\mu_1 > \mu_2$ (solid line) or by $p > .5$ (dashed line). First column: performance for estimate of μ_1; second column: performance for estimate of μ_2; third column: performance for estimate of p. Second row: ninety-percent coverage probabilities for indicated parameter when normalized by $\mu_1 > \mu_2$ (solid line) or by $p > .5$ (dashed line); dotted line gives nominal 90% goal. Third row: average squared difference between posterior mean and true value for indicated parameter when normalized by $\mu_1 > \mu_2$ for three different priors. Solid line: prior variance = 5; dashed line: prior vairance = 100; dotted line: truncated Normal prior.

Example 2: A Structural VAR model

$$B_0 y_t = k + \sum_{j=1}^{p} B_j y_{t-j} + u_t \tag{5}$$

with $u_t \sim N(0, D^2)$, D diagonal

$$B_0 = \begin{pmatrix} 1 & -\beta & 0 \\ 1 & -\gamma & -h \\ 0 & 0 & 1 \end{pmatrix} \tag{6}$$


h=1
h=0.5
h=0.1
h=0
h=-0.5
h=-1
β
γ
-192
-209
-175
-184
-168
-170
-167
-167
-168
-160
-178
-161

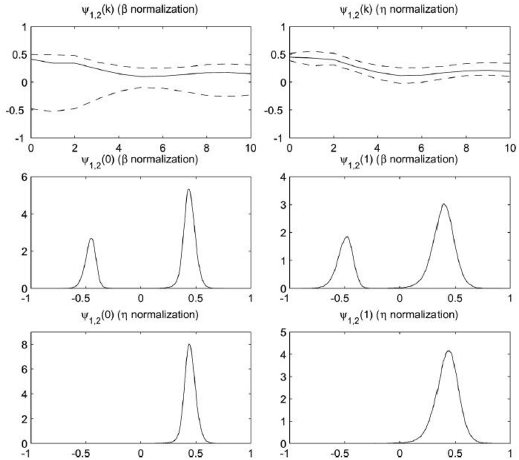


FIGURE 4 Top row: impulse-response function and 90% confidence interval for the effect of a one standard deviation increase in quantity demanded on the price k periods later under the β-normalization (left panel) and η-normalization (second panel). Second row: posterior density for the $k = 0$ (left panel) and $k = 1$ (right panel) values of the impulse-response function plotted in the upper left panel. Third row: posterior density for the $k = 0$ and $k = 1$ values of the impulse-response function plotted in the upper right panel.

Conclusion

- careless normalization can lead to misleading statistical inference
- check different normalizations
- algorithm for structural VARs is available (Waggoner &Zha 2003)