Does Jeffrey's Prior Alleviate the Incidental Parameter Problem?

Jinyong Hahn

February 12, 2003

1 Motivation

Panel data allow the possibility of controlling for unobserved individual heterogeneity. Models and methods of controlling for unobserved heterogeneity in linear models are well established. Controlling for unobserved heterogeneity is more difficult in nonlinear models, although some progress has been made by Hahn and Kuersteiner (2002), Hahn and Newey (2002), Lancaster (2002), and Woutersen (2002). The only Bayesian result in this short list is due to Lancaster (2002), who asked if there exists a prior such that the resultant Bayes estimator is robust to the incidental parameters discussed by Neyman and Scott (1948). Lancaster showed that a (diffuse) prior on orthogonalized incidental parameters has the desired robustness property. Unfortunately, parameter orthogonalization requires a solution to certain differential equations, which does not exist in general. Therefore, his result cannot be applicable to many panel models.

It may be speculated that Jeffrey's prior may provide such a solution. Jeffrey's prior is well-known to have the invariance property.[1] It should therefore be invariant to the parameter orthogonalization. As such, it may be hoped that a Bayes estimator based on Jeffrey's prior may be robust to incidental parameters.

In this note, it is shown that Jeffrey's prior does *not* have such property. I consider a simplified version of the model discussed in Chamberlain (1980), and show that a Bayes estimator based on Jeffrey's prior is subject to the incidental parameter problem. We do not yet have any prior distribution that provides a *general* solution to the incidental parameter problem.

[1] See, e.g., Zellner (1971, p. 47).

2 A Simple Model and Bayes Estimator with Jeffrey's Prior

We consider a simplified version of the model discussed in Chamberlain (1980). Assume that

$$x_{it} \sim N\left(\alpha_i, \theta\right), \qquad i = 1, \ldots, N : t = 1, \ldots, T \tag{1}$$

Here, θ is the common parameter, and α_i is a fixed effect. This is an interesting model despite its simplicity because the MLE $\widehat{\theta}$ for θ is inconsistent as $N \to \infty$ while T is fixed, whereas the "usual" Bayes estimator for θ based on the diffuse priors on α_i and θ is consistent. Consistency of such Bayes estimator can be explained by the fact that the parameter orthogonality is built-in for this model. See information matrix (2) presented in Appendix.

For this model, it can be shown[2] that the Jeffrey's prior is proportional to $\theta^{-\frac{N+2}{2}}$. It follows that the posterior $p\left(\theta, \alpha_1, \ldots, \alpha_N\right)$ is proportional to

$$\theta^{-\frac{NT}{2}} \exp\left[-\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\alpha_i\right)^2}{2\theta}\right] \cdot \theta^{-\frac{N+2}{2}}$$
$$= \theta^{-\frac{N(T+1)+2}{2}} \exp\left[-\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T}\left(x_{it}-\overline{x}\right)^2 + T\left(\alpha_i-\overline{x}\right)^2\right)}{2\theta}\right]$$

Writing the RHS as

$$\theta^{-\frac{N(T+1)+2}{2}} \exp\left[-\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\overline{x}\right)^2}{2\theta}\right] \cdot \prod_{i=1}^{N} \exp\left[-\frac{\left(\alpha_i-\overline{x}\right)^2}{2\theta/T}\right]$$
$$\propto \theta^{-\frac{NT+2}{2}} \exp\left[-\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\overline{x}\right)^2}{2\theta}\right] \cdot \prod_{i=1}^{N} \frac{1}{\left(\theta/T\right)^{1/2}} \exp\left[-\frac{\left(\alpha_i-\overline{x}\right)^2}{2\theta/T}\right]$$

we obtain

$$\int p\left(\theta, \alpha_1, \ldots, \alpha_N\right)\left(\prod_{i=1}^{N} d\alpha_i\right) \propto \theta^{-\frac{NT+2}{2}} \exp\left[-\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\overline{x}\right)^2}{2\theta}\right]$$

It follows that the posterior mode for θ is equal to

$$\widetilde{\theta} \equiv \frac{1}{NT+2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\overline{x}\right)^2$$

[2]Proof in Appendix.

Fixing T, and letting $N \to \infty$, we can see that

$$\operatorname*{plim}_{N\to\infty}\left(\widetilde{\theta}-\theta\right)=\frac{\theta}{T}=O\left(\frac{1}{T}\right)$$

This is the same order of (asymptotic) bias as the MLE $\widehat{\theta}$.

3 Summary

Diffuse prior is subject to the incidental parameter problem in general.[3] Lancaster's (2002) parameter orthogonalization, which is robust to such problem, is not feasible in general. In this note, it was seen that even Jeffrey's prior is subject to the same problem. The search for a robust prior continues.

A Derivation of Jeffrey's prior

We first start with calculation of information. Ignoring the irrelevant constant, we can see that the log of the likelihood L is equal to

$$\log L=-\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\log\theta+\frac{(x_{it}-\alpha_i)^2}{\theta}\right)$$

We therefore have

$$\begin{aligned}
\frac{\partial^2\log L}{\partial\theta^2} &= \sum_{i=1}^{N}\sum_{t=1}^{T}\left(\frac{1}{2\theta^2}-\frac{(x_{it}-\alpha_i)^2}{\theta^3}\right)\\
\frac{\partial^2\log L}{\partial\theta\partial\alpha_i} &= -\sum_{t=1}^{T}\frac{(x_{it}-\alpha_i)}{\theta^2}\\
\frac{\partial^2\log L}{\partial\alpha_i\partial\alpha_j} &= 0\\
\frac{\partial^2\log L}{\partial\alpha_i^2} &= -\sum_{t=1}^{T}\frac{1}{\theta}
\end{aligned}$$

[3]For example, Bayes estimator based on the "usual" diffuse prior is numerically equivalent to OLS in the case of dynamic panel model with fixed effects, and the latter is known to suffer from the incidental parameter problem. See Nickell (1981).

and the information matrix $\mathcal{I}(\theta, \alpha_1, \ldots, \alpha_N)$ equal to

$$\mathcal{I}(\theta, \alpha_1, \ldots, \alpha_N) = \begin{bmatrix} \frac{NT}{2\theta^2} & 0 & & 0 \\ 0 & \frac{T}{\theta} & & 0 \\ & & \ddots & \\ 0 & 0 & & \frac{T}{\theta} \end{bmatrix} \tag{2}$$

from which we can conclude that

$$|\mathcal{I}(\theta, \alpha_1, \ldots, \alpha_N)| = \frac{NT}{2\theta^2}\left(\frac{T}{\theta}\right)^N = \frac{NT^{N+1}}{2}\frac{1}{\theta^{N+2}}$$

Jeffrey's prior is therefore proportional to $\theta^{-\frac{N+2}{2}}$.

References

[1] Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data," *Review of Economic Studies* 47, 225 – 238.

[2] Hahn, J., and G. Kuersteiner (2002), "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T are Large," *Econometrica* 70, 1639–1657.

[3] Hahn, J., and W. Newey (2002), "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models", *unpublished working paper.*

[4] Lancaster, T. (2002), "Orthogonal Parameters and Panel Data", *Review of Economic Studies* 69, 647 – 666.

[5] Neyman, J., and E.L. Scott (1948), "Consistent Estimation from Partially Consistent Observations", *Econometrica* 16, 1–32.

[6] Nickell, S. (1981), "Biases in Dynamic Models with Fixed Effects", *Econometrica* 49, pp. 1417 - 1426.

[7] Woutersen, T. (2002), "Robustness against Incidental Parameters", *unpublished working paper.*

[8] Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley: New York.