2002

2002-12 Minimal Asymptotic Distributions for Estimators of Panel Data Models

Tiemen Woutersen

ISSN:0318-725X
ISBN:0-7714-2416-7

RESEARCH REPORT 2002-12

Minimal Asympototic Distributions for Estimators of Panel Data Models

by

Tiemen Woutersen

December 2002

Department of Economics
Social Science Centre
University of Western Ontario
London, Ontario, Canada
N6A 5C2
econref@uwo.ca

http://www.ssc.uwo.ca/economics/econref/html/RRPaper.html

Minimal Asymptotic Distributions for Estimators of Panel Data Models*

BY TIEMEN WOUTERSEN†

December 12, 2002

Abstract

This paper develops the minimal asymptotic distributions for estimators of panel data models with incidental parameters.

KEYWORDS: Incidental parameters, panel data

JEL CLASSIFICATION: C33, C31, C14

1 Introduction

THE AVAILABILITY OF PANEL DATA allows the econometrician to control for the heterogeneity by allowing for time-invariant, individual specific parameters. This fixed effect approach introduces many parameters into the model which causes the 'incidental parameter problem' of Neyman and Scott (1948): the maximum likelihood estimator is in general inconsistent. For some models with exogenous regressors, such as the linear, logit and Weibull model, one can find a sufficient statistic for the fixed effect so that one can use a conditional likelihood that does not depend on the fixed effect. For these models, one can estimate the parameters of interest consistently using datasets with a large number of individuals and just two observations per individuals, see Chamberlain (1984, 1985) and Baltagi (1995) for overviews.

*I thank Jinyong Hahn for his encouragement. The 'Social Science and Humanities Council' in Canada provided financial support. All errors are mine.

†Mailing Address: University of Western Ontario, Department of Economics, Social Science Center, London, Ontario, N6A 5C2, Canada. Email: twouters@uwo.ca.

However, for dynamic panel data models, no sufficient statistic has been found for the fixed effect for any model. For the dynamic linear model, several authors have derived estimators that are based on taking differences of the dependent variables and instrumental variables techniques, see the overview of Arellano and Honoré (2001). However, for nonlinear models, one cannot difference away the fixed effect. Alvarez and Arellano (1998) recognize that in many applications, the number of observations per individual is larger than two and develop an asymptotic approximation where both the number of individuals, n, and the number of observations per individuals, T increases. This asymptotic approximation is used by Phillips and Moon (1999), Hahn and Kuersteiner (2002), Hahn, Hausman and Kuersteiner (2002) and Woutersen and Voia (2002) to evaluate estimators for the dynamic linear model with fixed effects. Hahn and Newey (2002) use the same asymptotic approximation to evaluate their bias corrected likelihood estimator and Woutersen (2002) uses the approximation to evaluate his integrated moment estimator for dynamic panel data models.

This paper derives the efficiency bound for panel data models with fixed effects. In particular, it shows that the variance of the minimal asymptotic distribution equals the Cramér-Rao bound. This equality is already established for cross section models. The key insight, due to LeCam (1953), is that estimators can only improve on 'regular' estimators on a subset of the parameter space that has Lebesgue measure zero, see Van der Vaart and Wellner (1996) or Van der Vaart (1998) for overviews. The paper also shows that the efficiency bound of an asymptotic approximation with n and T increasing does not depend on the relative rate at which T increases. We conclude by considering regressor parameter of the exponential model and compare the minimal asymptotic distribution of this parameter to the semiparametric efficiency bound as derived by Hahn (1994).

This paper is organized as follows. Section 2 derives the efficiency bound and provides examples. Section 3 concludes and all the proofs are in the appendix.

2 Efficiency Bounds

Many estimators can be written as weighted averages of random variables. For example, the least squares estimator is the weighted average of the error terms of the linear model. More generally, the influence function of the moment estimators is a weighted average of moment functions, see Newey and McFadden (1994). Suppose we have a cross-section dataset with n individuals and assume the data generating process in independent across individuals. Then such estimators converge at $n^{-1/2}$ subject to identification and regularity conditions. The asymptotic variance is a good measure to compare different method of moment estimators. However, one can always find an estimator that converges at an arbitrarily fast rate for particular parameter values. In particular, consider the following example, given by Hodges in 1951.

Example 1 (Hodges' estimator): Let M_n be the mean of a sample of size n from the $N(\theta, 1)$ distribution. Define a second estimator S_n through

$$S_n = \begin{cases} M_n \text{ if } |M_n| \geq n^{-1/4} \\ 0 \text{ if } |M_n| < n^{-1/4}. \end{cases}$$

Note that $\sqrt{n}(S_n - \theta) \rightarrow_d N(0,1)$ for $\mu \neq 0$ and $q_n(S_n - \theta) \rightarrow_p 0$ for $\theta = 0$ for every q_n.

LeCam (1953) showed that such improvements over an estimator M_n can only be made for a subset of the parameter space that has Lebesgue measure zero. Because a set with Lebesgue measure zero is small, LeCam introduces the idea of estimating a *sequence of parameter values* in order to rule out such small sets. In particular, consider estimating the sequence $\theta + h/r_n$ where h is fixed and norming operator r_n is increasing with n. For the method of moment estimators discussed above, we have $r_n = \sqrt{n}$. Requiring that the asymptotic distribution of $\theta + h/r_n$ does not depend on h/r_n rules out estimators such as the Hodges estimator. In particular, let H denote a Banach space and let an estimator M_n be *regular* if, for every $h \in H$ and every θ,

$$(1) \qquad r_n\{M_n - (\theta + h/r_n)\} \rightarrow_d \mathfrak{L}(\theta)$$

for some limiting distribution $\mathfrak{L}(\theta)$. A crucial feature is that $\mathfrak{L}$ does not depend[1] on h. Loosely speaking, the definition says that an estimator is regular if, for every θ, the rate of convergence at θ is the same as the rate of convergence in a small neighborhood of θ.

The next step is to compare properties of $\mathfrak{L}$ for different estimators. In particular, $\mathfrak{L}$ has a *minimal asymptotic distribution* if $\mathfrak{L}$ is centered at zero has the smallest possible variance. The minimal asymptotic distribution can be derived from properties of the densities $p(y_1, ..., y_n|\theta)$ and $p(y_1, ..., y_n|\theta + h/r_n)$ where $y_i, i = 1, ..., n$, denotes the dependent variable and the $\theta, \theta + h/r_n$ denote the parameters of the data generating process. The following example shows how this can be done for the exponential model.

Example 2 Suppose $y_i, i = 1, ..., n$, is distributed exponentially with mean $\frac{1}{\theta}$. This implies the following densities for individual i,

$$\begin{aligned} p(y_i|\theta) &= \theta e^{-y_i} \text{ and} \\ p(y_i|\theta + h) &= (\theta + h)e^{-(\theta+h)y_i}. \end{aligned}$$

Consider

$$\begin{aligned} \ln \frac{p(y_i|\theta + h)}{p(y_i|\theta)} &= \ln \frac{(\theta + h)e^{-(\theta+h)y_i}}{\theta e^{-y_i}} \\ &= \ln(1 + \frac{h}{\theta}) - hy_i \\ &= \frac{h}{\theta} - \frac{1}{2}(\frac{h}{\theta})^2 + o(h^2) - hy_i \\ &= h\{\frac{1}{\theta} - y_i\} - \frac{1}{2}h^2\frac{1}{\theta^2} + o(h^2). \end{aligned}$$

The expression $\ln \frac{p(y_i|\theta+h)}{p(y_i|\theta)}$ can also be written as a function of the score of the log likelihood[2] and the Jacobian,

$$\ln \frac{p(y_i|\theta + h)}{p(y_i|\theta)} = hL_\theta^i + \frac{1}{2}h^2 L_{\theta\theta}^i + o(h^2).$$

Similarly,

$$\begin{aligned} \ln \prod_i \frac{p(y_i|\theta + h/\sqrt{n})}{p(y_i|\theta)} &= \ln \frac{(\theta + h)^n e^{-\sum_i(\theta+h)y_i}}{\theta^n e^{-\sum_i \theta y_i}} \\ &= h\frac{L_\theta}{\sqrt{n}} + \frac{1}{2}h^2\frac{L_{\theta\theta}}{n} + o(h^2). \end{aligned}$$

It follows from Van der Vaart (1998, Theorem 8.9) that the 'minimal' asymptotic distribution of an estimator for θ is $N(0, [\frac{EL_{\theta\theta}}{n}]^{-1})$.

We now derive the minimal asymptotic distributions for panel data model with incidental parameters. Suppose we observe n individuals for T periods and we want to control for heterogeneity by allowing for individual (or incidental) parameters. Let the data generating process be conditional independent and suppose that we want to use an asymptotic approximations in which both n and T increase. In this case, the individual parameters of individual i only appear in the likelihood contribution of individual i. As a consequence, we can only estimate the individual parameters at rate $T^{-1/2}$ while we potentially estimate the common parameters at rate $(nT)^{-1/2}$. For estimation, the rate at which T increases relative to n may be important (see Hahn and Kuersteiner (2002) and Woutersen (2002)) but we show that the minimal asymptotic distribution does not depend at the rate at which T increases relative to n. An intuition for this result is that the Cramér-Rao bound does not depend on the relative rates of T and n and the following example illustrates this.

Example 3 Suppose that we observe n individuals for T periods and that want to estimate an exponential hazard model. Since we have more than one observation for each individual, we can allow for a fixed effect. Let x_{is} denote the vector of characteristics of individual i for spell s. If the spells are independent across individuals as well as across spells, then the hazards of individual i can be written as

$$\theta_{is}(y_{is}) = f_i e^{x_{is}\beta},\ s = 1, ..., T \text{ and } i = 1, ..., n.$$

This hazard implies the following log likelihood,

$$L(f_1, ..., f_n, \beta) = \sum_i T \ln f_i + \sum_i \sum_s x_{is}\beta - \sum_i \sum_s f_i e^{x_{is}\beta} y_{is}.$$

Differentiating $L(f_1, ..., f_n, \beta)$ gives

$$\begin{aligned} L_\beta &= \sum_i \sum_s x_{is} - \sum_i \sum_s x_{is} f_i e^{x_{is}\beta} y_{is} \\ L_{f_i} &= \frac{1}{f_i} - \sum_s e^{x_{is}\beta} y_{is}. \end{aligned}$$

Projecting L_β on $L_{f_1}, ..., L_{f_n}$ gives the efficient score S_β,

$$\begin{aligned} S_\beta &= L_\beta - \sum_i L_{f_i} \frac{EL_{\beta f_i}}{EL_{f_i f_i}} \\ &= L_\beta - \sum_i \frac{L_{f_i} \sum_s x_{is}}{f_i} \\ &= -\sum_i \sum_s \tilde{x}_{is} f_i e^{x_{is}\beta} y_{is} \end{aligned}$$

where $\tilde{x}_{is} = x_{is} - \frac{\sum_s x_{is}}{T}$. The Cramér-Rao bound equals $[ES_\beta S'_\beta]^{-1}$ and we will show that the variance of the minimal asymptotic distribution is also equal to $[ES_\beta S'_\beta]^{-1}$.

Van der Vaart and Wellner (1996, theorem 3.11.2) give the minimal asymptotic distribution of regular estimators. We use a version of this theorem that is given in Hahn and Kuersteiner (2002, Theorem 6).

Let $P_{n,h}$ denote a probability measure on a measurable space $(\mathfrak{X}_n, \mathfrak{A}_n)$ and assume the following.

Assumption 1

$P_{n,h} : h \in H$ is asymptotically normal.

A good discussion of local asymptotic normality is given in Van der Vaart (1998, section 7.6). A data generating process is asymptotically normal if the log likelihood can be approximated by a quadratic function around the true value of the parameters. The data generating process of example 3 satisfies local asymptotic normality. Also assume that

Assumption 2

Let $T \propto n^\alpha$ where $\alpha \geq 0$.

This assumption formalizes the notion that we use an asymptotic approximation in which n increases or both n and T increase.

Assumption 3

Let the data generating process be independent across individuals conditional on the regressors x_{it} and individual effect f_i.

Assumption 4

Let $r_n = \sqrt{nT} = n^{(1+\alpha)/2}$

Moment estimators that are based on taking averages have a norming operator proportional to $\sqrt{nT}$ and we are interested in the set of estimators with this norming operator. However, some of the nonstationary cases considered by Phillips and Moon (1999) do not have this norming operator.

Assumption 5

Let assumption (i)-(v) of theorem 6, Hahn and Kuersteiner (2002), hold.

Let $\Delta_1 = \frac{L_\beta}{\sqrt{nT}}$ and that $\Delta_2 = \frac{L_f}{\sqrt{nT}}$. Let $\tilde{\Delta}_1$ be the residual in the projection of Δ_1 on Δ_2. That is

$$\begin{aligned}\tilde{\Delta}_1 &= \frac{L_\beta}{\sqrt{nT}} - \frac{E(L_\beta L_f')}{nT}[\frac{E(L_f L_f')}{nT}]^{-1}\frac{L_f}{\sqrt{nT}} \\ &= \frac{L_\beta}{\sqrt{nT}} - \frac{EL_{\beta f}}{nT}[\frac{EL_{ff}}{nT}]^{-1}\frac{L_f}{\sqrt{nT}}.\end{aligned}$$

Note that $\frac{EL_{\beta f}}{nT}[\frac{EL_{ff}}{nT}]^{-1}\frac{L_f}{\sqrt{nT}}$ is $O(1)$. The minimal asymptotic distribution is determined by $\tilde{\Delta}_1$. We conclude that

Theorem

Let assumption 1-5 hold. Then the minimal asymptotic distribution is given by $N(0, \{E(\tilde{\Delta}_1\tilde{\Delta}_1')\}^{-1})$. Moreover, the variance of the minimal asymptotic distribution, $\{E(\tilde{\Delta}_1\tilde{\Delta}_1')\}^{-1}$, equals the Cramér-Rao bound.

Note that the variance of the minimal asymptotic distribution does not depend on the relative rate at which increase T increases. We now compare the variance of the minimal asymptotic distribution to the semiparametric efficiency bound of a particular model. Semiparametric efficiency bounds are not available for all models but Hahn (1994) and Hahn

(1997) derived the semiparametric efficiency bounds for a set of models for which there exists a sufficient statistics for the fixed effects.

Example 3 (Continued) Hahn (1994) derives the semiparametric efficiency bound for the exponential hazard model under the assumption that f_i is a random effect. This bound has the following expression,

$$[E\tilde{S}_\beta \tilde{S}'_\beta]^{-1} \text{ where}$$
$$\tilde{S}_\beta = -\sum_i \frac{\sum_s \tilde{x}_{is} e^{x_{is}\beta} y_{is}}{\sum_s e^{x_{is}\beta} y_{is}}.$$

Note that $[E\tilde{S}_\beta \tilde{S}'_\beta]^{-1} > [ES_\beta S'_\beta]^{-1}$ for every finite T where $[ES_\beta S'_\beta]^{-1}$ denotes the Cramér-Rao bound of example 3. Suppose the 'true model' is a random effects model so that no estimator can have a variance smaller than $[E\tilde{S}_\beta \tilde{S}'_\beta]^{-1}$. In that case, no fixed effects estimator estimator can have a variance smaller than $[E\tilde{S}_\beta \tilde{S}'_\beta]^{-1} > [ES_\beta S'_\beta]^{-1}$ for every finite T. Thus, no estimator reaches the Cramér-Rao bound $[ES_\beta S'_\beta]^{-1}$ for any finite T. Therefore, the Cramér-Rao bound is especially relevant in an asymptotic with T increasing since some estimators may reach the Cramér-Rao bound in this asymptotic. For this reason, we used an asymptotic with T increasing in this paper.

3 Conclusion

This paper develops the minimal asymptotic distributions for estimators of models with incidental parameters. We emphasize an asymptotic approximation with both T and n increasing. The reason for this is that recent papers use this asymptotic approximation to evaluate estimators, see Hahn and Newey (2002) and Woutersen (2002). This paper gives the minimal asymptotic distribution for estimators under the assumptions of these papers. We also compared the minimal asymptotic distribution of a mixed exponential model to its semiparametric efficiency bound and showed that the minimal asymptotic distribution has a smaller variance for finite T than the variance implied by the semiparametric efficiency bound. This implies that the variance of the minimal asymptotic distribution only gives a 'sharp bound' for the variance of regular estimators in an asymptotic with T increasing.

4 Appendices

Short review of the Cramér-Rao bound:

Stuart et al. (1991, section 17.13-17.17, 17.24-17.28, and 18.15-18.16) give a clear exposition of the Cramér-Rao bound. Let $L(\theta)$ denote the log-likelihood and let the derivative of the log-likelihood with respect to θ be denoted by $L_\theta(\theta)$. Let $L_\theta = L_\theta(\theta_0)$. Then the Cramér-Rao bound is given by $\{E(L_\theta L_\theta')\}^{-1}$. Let $L_\theta' = \{L_\beta', L_f'\}$ where L_f is a vector with length N. Then the information bound for the incidental parameter models is given by

$$\begin{pmatrix} EL_{\beta\beta} & EL_{\beta f} \\ EL_{f\beta} & EL_{ff} \end{pmatrix}^{-1} = \begin{pmatrix} \{EL_{\beta\beta} - EL_{\beta f}(EL_{ff})^{-1}EL_{f\beta}\}^{-1} & -(EL_{\beta\beta})^{-1}(EL_{\beta f})F \\ -F(EL_{f\beta})(EL_{\beta\beta})^{-1} & F \end{pmatrix}^{-1},$$

where $F = \{EL_{ff} - EL_{f\beta}(EL_{\beta\beta})^{-1}EL_{\beta f}\}^{-1}$. Using the fact that EL_{ff} is a diagonal matrix we can also write the upper left element of this matrix as follows,

$$\{EL_{\beta\beta} - EL_{\beta f}(EL_{ff})^{-1}EL_{f\beta}\}^{-1} = [E\sum_i \{L^i_{\beta\beta} - \frac{EL^i_{\beta f}EL^i_{f\beta}}{EL^i_{ff}}\}]^{-1}.$$

Following LeCam (1953), Van der Vaart (1998, chapter 8) discusses how the Cramér-Rao bound gives a lower bound on the variance for regular estimators of cross section models.

Proof of Theorem 1

Theorem 6 of Hahn and Kuersteiner (2002) was originally derived to deal with weak instruments where the number of instruments increases. However, the theorem is general enough to cover an asymptotic approximation for panel data models with N and T increasing. Hahn and Kuersteiner (2002, page 1653) note that "Theorem 6 implies that the 'minimal' asymptotic distribution is $N(0, \{E(\tilde{\Delta}_1\tilde{\Delta}_1')\}^{-1})$.". Thus, we only need to prove that $\{E(\tilde{\Delta}_1\tilde{\Delta}_1')\}^{-1}$ equals the Cramér-Rao bound. Note that the Cramér-Rao bound can be written as the inverse of the variance of the efficient score S_β. In particular, the efficient score S_β is the projection of the score L_β on $\{L_{f_1}, ..., L_{f_N}\}$, $S_\beta = L_\beta - \sum_i L_{f_i}\frac{EL_{\beta f_i}}{EL_{f_i f_i}}$. Note that $\tilde{\Delta}_1$ has the same form. Thus, the Cramér-Rao bound equals the variance of the minimal asymptotic distribution.

5 References

Alvarez, J. and M. Arellano (1998): "The Time Series and Cross-Section Asymptotics of Dynamic Panel Data Estimators," Working Paper 9808, CEMFI, Madrid.

Arellano, M., and B. E. Honoré, (2001): "Panel Data Models: Some Recent Developments," in *Handbook of Econometrics,* Vol. 5, ed. by J. Heckman and E. Leamer, Amsterdam: North-Holland.

Baltagi, B. H. (1995): *Econometric Analysis of Panel Data,* New York: John Wiley and Sons, New York.

Chamberlain, G. (1984): "Panel Data," in *Handbook of Econometrics,* Vol. 2, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland.

Greene, W. H. (2002): *Econometric Analysis,* fifth edition, Prentice Hall, Upper Saddle River New Jersey.

Hahn, J. (1994): "The Efficiency Bound of the Mixed Proportional Hazard Model," *Review of Economic Studies,* 61,607-629.

Hahn, J. (1997): "A Note on the Efficient Semiparametric Estimation of Some Panel Models", *Econometric Theory, 13, 583-588.*

Hahn, J., J. Hausman, and G. Kuersteiner (2001): "Bias Corrected Instrumental Variables Estimation for Dynamic Panel Models with Fixed Effects", MIT working paper.

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, 1639-1658.*

Hahn J. and W. Newey (2002): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models", MIT manuscript.

Hsiao, C. (1986): *"Analysis of Panel Data".* New York: Cambridge University Press.

LeCam, L. (1953): "On some Asymptotic Properties of Maximum Likelihood Estimates and related Bayes' estimates," *Annals of Mathematical Statistics,* 41, 802.

Mundlak, Y. (1961): "Empirical Production Function Free of Management Bias," *Journal of Farm Economics,* 43, 44-56.

Neyman, J., and E. L. Scott (1948): "Consistent Estimation from Partially Consistent Observations," *Econometrica*, 16, 1-32.

Newey, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99-135.

Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics,* Vol. 4, ed. by R. F. Engle and D. MacFadden. Amsterdam: North-Holland.

Phillips and Moon (1999): "Linear Regression Limit Theory for Nonstationary Panel Data", *Econometrica*, 67, 1057-1111.

Van der Vaart, A. W. (1998): *Asymptotic Statistics,* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, UK.

Van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes,* Springer-Verlag, New York.

Woutersen, T. M. (2002): "Robustness against Incidental Parameters," Working paper, Department of Economics, University of Western Ontario.

Woutersen, T. M. and M. Voia (2002): "Adaptive Estimation of the Dynamic Linear Model with Fixed Effects," Working paper, Department of Economics, University of Western Ontario.

Notes

[1]See Van der Vaart and Wellner (1996 page 413) for a more general definition of a regular estimator.

[2]Note that $L(\theta) = \sum_i L^i(\theta) = N \ln \theta - \sum_i \theta y_i$.