Likelihood Inference for Dynamic Panel Models

by

Marc Nerlove

WP 98-16

Department of Agricultural and Resource Economics
The University of Maryland, College Park

files/papers97/Paris/LikelihoodPanel.doc **FINAL REVISION 03.31.98**
To be published in *L'Annales d'Économie et de Statistique* de l'INSEE in late 1998.

Comments welcome.

Likelihood Inference for Dynamic Panel Models

Marc Nerlove
Department of Agricultural and Resource Economics
University of Maryland
Tel: (301) 405-1388 Fax: (301) 314-9032
e-mail: mnerlove@arec.umd.edu
homepage: http://www.arec.umd.edu/mnerlove/mnerlove.htm

ABSTRACT

The *likelihood principle* is applied to the problem of inference in dynamic panel models. The principle states that the likelihood function contains "...*all* the information which the data provide concerning the relative merits of..." alternative parametric hypotheses. The usual asymptotic theory of maximum likelihood is based on a quadratic approximation to the likelihood function in the nearby neighborhood of a local maximum of the function. One needs to look at the entire function more broadly in order to ascertain the true significance of the data for the hypotheses under consideration, not only because of the possibilities of multiple local maxima and boundary solutions, but also because the data are typically differentially informative with respect to different regions of the parameter space. In order to handle cases in which the likelihood function depends on more than two parameters, The devices of "concentrating" and of "slicing" or sectioning the function in the direction of a hyperplane or surface reflecting the variation of all but two of the parameters are introduced. The likelihood functions for two basic dynamic panel models: (1) a model involving individual-specific effects which reflect the influence of latent time-persistent variables; (2) a model involving individual-specific time trends which reflect the nonstationarity introduced by trending latent variables, are derived. The methods are applied to the analysis of cross-country economic growth. The findings demonstrate the power and feasibility of general methods of likelihood inference, especially to reveal problems of inference and areas of ignorance.

*This paper was prepared for the Seventh Conference on Panel Data Econometrics, 19-20 June 1997, Paris. The research on which it is based was supported by the Maryland Agricultural Experiment Station, Project A-53. The paper will appear as MAES Paper No. ****.
I am indebted to G. S. Maddala for his comments, to two anonymous referees for theirs, and to Anke Meyer for numerous helpful suggestions and for her wise expository counsel. Jinkyo Suh provided able computational assistance.

What has now appeared is that the mathematical concept of probability is ... inadequate to express our mental confidence or diffidence in making ... inferences. and that the mathematical quantity which usually appears to be appropriate for measuring our order of preference among different possible populations does not in fact obey the laws of probability. To distinguish it from probability, I have used the term **"Likelihood"** to designate this quantity; since both the words "likelihood" and "probability" are loosely used in common speech to cover both kinds of relationship.

R. A. Fisher. *Statistical Methods for Research Workers*, 1925.

Within the framework of a statistical model, *all* the information which the data provide concerning the relative merits of two hypotheses is contained in the likelihood ratio of those hypotheses on the data. ...For a continuum of hypotheses, this principle asserts that the likelihood function contains all the necessary information.

A. W. F. Edwards, *Likelihood*, 1972.

You are living on a Plane. What you style Flatland is the vast level surface of what I may call a fluid, or in, the top of which you and your countrymen move about, without rising above or falling below it.

I am not a plane Figure, but a Solid. You call me a Circle; but in reality I am not a Circle, but an infinite number of Circles, of size varying from a Point to a Circle of thirteen inches in diameter, one placed on the top of the other. When I cut through your plane as I am now doing, I make in your plane a section which you, very rightly, call a Circle. For even a Sphere--which is my proper name in my own country--if he manifest himself at all to an inhabitant of Flatland--must needs manifest himself as a Circle.

E. A. Abbott, *Flatland*, 1884.

It was six men of Indostan
To learning much inclined,
Who went to see the Elephant
(Though all of them were blind),
That each by observation
Might satisfy his mind.
The First approached the Elephant,
And happening to fall
Against his broad and sturdy side,
At once began to bawl:
"God bless me! but the Elephant
Is very like a wall!"
The Second, feeling of the tusk,
Cried, "Ho! what have we here
So very round and smooth and sharp?
To me 'tis mighty clear
This wonder of an Elephant
Is very like a spear!"

The Third approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
"I see," quoth he, "the Elephant
Is very like a snake!"
The Fourth reached out an eager hand,
And felt about the knee.
"What most this wondrous beast is like
Is mighty plain," quoth he;
" 'Tis clear enough the Elephant
Is very like a tree!"
The Fifth, who chanced to touch the ear,
Said: "E'en the blindest man
Can tell what this resembles most;
Deny the fact who can
This marvel of an Elephant
Is very like a fan!"

The Sixth no sooner had begun
About the beast to grope,
Than, seizing on the swinging tail
That fell within his scope,
"I see," quoth he, "the Elephant
Is very like a rope!"
And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!
Moral:So oft in theologic wars,
The disputants, I ween,
Rail on in utter ignorance
Of what each other mean,
And prate about an Elephant
Not one of them has seen!

John Godfrey Saxe , "The Blind Men and the Elephant: A Hindoo Tale," 1880

Preface

This paper applies the *likelihood principle* of Fisher (1921, 1922, 1925 and 1932), Barnard (1949, 1951, 1966, 1967 and Barnard, Jenkins and Winsten, 1962) and Birnbaum (1962) to the problem of inference in dynamic panel models.[1] Beginning with Chamberlain (1984) an extensive literature on non-likelihood methods for estimation an inference about dynamic panel models has emerged, much of it surveyed in Sevestre and Trognon (1996) and by Baltagi (1996, Chapter 8, pp. 125 - 148). I do not propose to survey this literature here or to compare the alternative estimates suggested with the results of likelihood inference, although such comparison would no doubt be highly useful.

In section 1, I develop the principle that the likelihood function contains "...*all* the information which the data provide concerning the relative merits of..." alternative parametric hypotheses. The usual asymptotic theory of maximum likelihood is shown to be based on a quadratic approximation to the likelihood function in the nearby neighborhood of a local maximum of the function. I argue that one needs to look at the entire function more broadly in order to ascertain the true significance of the data for the hypotheses under consideration, not only because of the possibilities of multiple local maxima and boundary solutions, but also because the data are typically differentially informative with respect to different regions of the parameter space. In order to handle cases in which the likelihood function depends on more than two parameters, I introduce the devices of "concentrating" and of "slicing" or sectioning the function in the direction of a hyperplane or surface reflecting the variation of all but two of the parameters.

In section 2, I derive the likelihood functions for two basic dynamic panel models: (1) a model involving individual-specific effects which reflect the influence of latent time-persistent variables; (2) a model involving individual-specific time trends which reflect the nonstationarity introduced by trending latent variables. In developing the likelihood functions for these two leading cases, I argue for reduction of models of type (2) to stationary models of type (1) by differencing. In this case, however it is necessary to modify the likelihood functions to reflect the effects of differencing on the unobserved residual variation. The differenced model now has a different interpretation from the original models of type (1) in that the individual-specific effects now represent individual-specific trend slopes. I further argue that in stationary cases, which include both levels models and differenced models, the initial observations of the dependent variables contain useful information on the process which must have generated those observations in the past, before the panel was observed, and that this information depends positively on their variance and on the number individuals in the panel, and is thus of particular importance for "shallow" panels.[2]

---

[1] Although *likelihood* and inference from likelihood resembles Laplace's method of *inverse probability* (Laplace, 1774 - 1814), which provides the principal basis for the Bayesian approach to inference, Fisher (1932) was a great pains to distinguish the two, and, indeed, was sharply critical of the use of prior distributions, especially of the use of "non-informative" priors to represent ignorance.

[2] Maddala (1971) discusses a similar problem, pooling cross-section and time-series data, from a Bayesian point of view. The analysis with diffuse priors is similar a number of respects to that presented here based on the likelihood principle. Some of Maddala's results are discussed below. As is the case with much of the literature in this area, however, those about likelihood or maximum likelihood are based on a likelihood function which conditions on the initial observations. Breusch's (1987) remarkable result, for example, about the convergence of iterated Generalized Least Squares to the ML estimates holds only for the case in which the likelihood function is conditional on the initial observations in the dynamic case.

Finally, in section 3, to assess the feasibility and power of likelihood methods for inference about dynamic panel models, I use data on 94 countries for the period 1960 - 1985, and a subsample of 22 OECD countries, from the Penn World Tables 5.6, publicly available from the NBER web site. The 22-country sample consists of primarily European countries, all highly developed and tied together by a network of trading relations; the 94-country sample is much more heterogeneous, consisting of the aforementioned 22 plus 72 additional countries ranging from Mozambique and Haiti to the "Asian Tigers." This is the same data set which has been used in dozens of previous studies.

In an previous paper (Nerlove, 1996), I compared some commonly used methods of estimation in dynamic panel models with one another and contrasted the results obtained from likelihood methods which take account of the information contained in the initial observations about the process which must have generated those observations in the pre-sample period. I showed that many of the earlier findings are probably statistical artifacts arising from biases in the econometric methods employed. Here I focus especially on the need to take advantage of the relatively large amount of information contained in the initial observations and to take account of differing country-specific trends. Using a simple variant of the Solow-Swan growth model widely used in recent studies of the convergence process, I demonstrate here that likelihood methods which take account of individual-specific trends and of the information present in the initial observations leads to acceptance of the convergence hypothesis, with the best-supported value of conditional convergence in the order of about 90% within 13 or 14 years for a broad sample of 94 countries. The analysis demonstrates the power and feasibility of general methods of likelihood inference, especially to reveal problems of inference and areas of ignorance.

1. Introduction: The Likelihood Principle

Although clearly implied in what Fisher wrote in the 1920's (1922,1925), the likelihood principle, which essentially holds that the likelihood function is the sole basis for inference, did not come into prominence until the 1950's and 1960's, principally through the work of Barnard, Birnbaum, and Edwards (see the references cited below, Barndorff-Nielsen, 1988, and Lindsey, 1996) written largely in reaction to both the classical Neyman-Pearson (frequentist) and the Bayesian approaches to inference (Jeffereys, 1934, 1961; see also Press, 1989).

A statistical model consists of a random vector $x \varepsilon X$ of observations having a joint distribution function $F(x;\theta)$, with corresponding density $f(x;\theta)$, depending on the unknown parameters $\theta \varepsilon \Theta$. It is assumed that F is known. The *likelihood function* determined by any given outcome x is defined as the function on Θ equal to $cf(x;\theta)$ where c is an arbitrary positive constant which may depend on x but does not depend on θ. Two likelihood functions defined on the same parameter space Θ, whether arising from the same "experiment" or from different "experiments," E_1 and E_2, are *equivalent* if their ratio is positive and independent of Θ for all $\theta \varepsilon \Theta$ except possibly at points at which both functions are zero (so that the ratio is undefined).

The *likelihood principle* asserts that for a given experiment E, the evidential meaning of any outcome x, for inference regarding θ is contained entirely in the likelihood function determined by x. All other aspects of how the data may have been generated are irrelevant, e.g., the sample space, provided, of course, that the sample space itself doesn't depend on θ. It follows that if two "experiments," E_1 and E_2, have pdf's $f(x,\theta)$ and $g(y,\theta)$, respectively, and if for some particular outcomes, x^* of E_1 and y^* of E_2,

$$f(x^*,\theta) = h(x^*, y^*)g(y^*,\theta),\ h(x^*,y^*) > 0, \text{ for all } \theta \varepsilon \Theta,$$

then these outcomes must result in the same inference about θ.

Birnbaum (1962) derives the likelihood principle from the sufficiency principle and a still more basic assumption, the so-called *conditionality principle.* This principle states that if an "experiment" involving θ is chosen from a collection of possible experiments, *independently of θ*, then any experiment not chosen is irrelevant to the statistical analysis. The conditionality principle makes clear the implication of the likelihood principle that any inference should depend only on the outcome observed and not on any other outcome we might have observed and thus sharply contrasts the method of likelihood inference from the Neyman-Pearson, or frequentist, approach, in which inference does depend crucially on a hypothetical sequence of experiments, the outcome of but one of which is observed. In particular, questions of unbiasedness, minimum variance, consistency and the like and the whole apparatus of confidence intervals, significance levels, and power of tests, are ruled out of bounds. While maximum-likelihood estimation does satisfy the likelihood principle (and thus sufficiency and conditionality), the frequentist assessment in terms of asymptotic properties is irrelevant. In this paper, I apply the likelihood principle to the problem of inference about the parameters of dynamic panel models and try to make clear the role of the maximum of the likelihood function and its Hessian evaluated at the maximum in approximating the whole of the likelihood function for purposes of inference.

The likelihood principle is clearly incomplete from the standpoint of inference since it nowhere states how the evidential meaning of the likelihood function is to be determined. To the principle, therefore, "likelihoodists" generally append the *method of support* (a term coined by Jefferevs, 1934). The *support function* is defined as the natural logarithm of the likelihood function. Since the likelihood function incorporates an arbitrary constant, the support function is defined only up to the addition of an arbitrary constant. Conventionally, this constant is often taken to be the value which makes support at the maximum equal zero. In multiplicative terms, this is equivalent to normalizing the likelihood function by dividing it by its value at the maximum. Only relative support for a particular parameter value over another can be interpreted in any case, so the constant disappears when looking at the difference between support values of different parameter values. The *method of maximum support* is the *method of maximum likelihood*. But the interpretation of the parameter value which yields this maximum and of the inverse of the negative of the Hessian at the point of maximum is different than in the frequentist interpretation in terms of asymptotic properties. The likelihoodist interpretation of these magnitudes is in terms of a quadratic approximation to the support function in the neighborhood of its maximum..

It is clear that the difference in the value of the support function at two different values of a parameter has the significance that the value for which support is greater is more consistent with the observed data than the value of lesser support. What we have is essentially a likelihood ratio test without the frequentist apparatus of asymptotic chi-square. It is also clear that the values of parameters for which maximum support is obtained (that is, the maximum-likelihood estimates), especially if the maximum is unique, have a special significance in relation to other possible values. Moreover, how sharply defined such a maximum of the likelihood function, if a unique maximum exists, is also clearly relevant to any inference we may wish to draw. On the negative side, a poorly behaved likelihood function, for example, one having ridges of equal likelihood, many local maxima, or a maximum on the boundary of an a priori admissible region of the parameter space, is generally indicative of an incompletely or ill-formulated underlying statistical model.

From a frequentist point of view what matters about the likelihood function is only its maximum and curvature in the neighborhood of the maximum, and all the desirable properties and the assessment of the reliability of the maximum-likelihood estimates are only asymptotic. Greene (1993, pp.111-116) gives a very brief discussion of these matters; Davidson and MacKinnon (1993, Chapter 8, pp.243-287) give a more complete and rigorous discussion; a more intuitive discussion with many econometric examples is given by Cramer (1986). That only the maximum and the Hessian at the maximum are all the matters from a frequentist point of view is perhaps not surprising in view of the fact that for the mean of a normal distribution the quadratic approximation is exact (see the discussion below) and because of the central limit theorem in its many forms many estimators, including ML estimators in regular cases, tend to normality in distribution.