Properties of Alternative Estimators of Dynamic Panel Models

An Empirical Analysis of Cross-Country Data for the Study of Economic Growth

by
Marc Nerlove

WP 98-17

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

files/papers97/ParisMaddala/MaddalaVol.doc **FINAL REVISION 06.25.97**

Properties of Alternative Estimators of Dynamic Panel Models:

An Empirical Analysis of Cross-Country Data for the Study of Economic Growth

Marc Nerlove
Department of Agricultural and Resource Economics
University of Maryland
Tel: (301) 405-1388 Fax: (301) 314-9032
email: mnerlove@arec.umd.edu

ABSTRACT

The sensitivity of the estimates of both the "state" variable and the other explanatory variables to the econometric method employed in dynamic panel models is investigated in the context of recent empirical studies of growth rate convergence using panel data from the Penn World Tables. Models with country-specific intercept effects and models with country specific trends are estimated. Although the investigation has as its primary purpose the assessment of alternative estimators, all of the results reported support the conventional interpretation of the coefficient of the lagged dependent variable in terms of growth convergence conditional on savings and population growth rates; however, the rather different estimates of this coefficient obtained when different estimation techniques are used is illustrated. In particular, I show that the use of fixed-effects panel models biases the results towards finding relatively rapid convergence and that, when more appropriate maximum-likelihood estimates, unconditional on the initial observations, are employed, very slow convergence is implied. Biases in the estimates of the coefficient of the "state" variable for all of the usual methods of panel data analysis imply biased estimates of the coefficients of any other variables included if these are correlated with the "state" variable, which is typically the case. Thus, the significance and possibly the sign of any other explanatory variables may be seriously affected. Consequently, the conclusions of many recent studies of the determinants of growth employing dynamic panel models may largely reflect the econometric methods employed.

This paper draws extensively from "Growth Rate Convergence, Fact or Artifact? An Essay in Panel Data Econometrics," prepared for the Sixth Conference on Panel Data Econometrics, Amsterdam, 28-29 June 1996. 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. ****.

My interest in the question of how the econometric approach might have influenced recent findings with respect to the convergence hypothesis was stimulated by reading Islam (1995a). Although I wound up going directly to the Penn World Tables for the data underlying the analyses presented here, I began by working with the series Islam was kind enough to supply and for which I am greatly indebted to him.

I thank Hashem Pesaran for helpful discussions, and Robert Barro, Michael Binder, William Greene, Kajal Lahiri, G. S. Maddala, and C. Spohr for useful comments. Pietro Balestra suggested the method used to obtain the likelihood function for the differenced model. Special thanks are due to Anke Meyer, with whom I discussed every aspect of this work

I am also indebted to Jinkyo Suh and Timothy Thomas for computational counsel and their assistance in straightening out the GAUSS programs which I wrote to obtain the results reported here. Suh also checked and double checked all derivations and verified that my programs accurately reflected the formulae derived and carried out further analyses for this revision in connection with the differenced model.

This paper is dedicated to my long-time friend and one-time colleague G. S. Maddala.

1. Introduction

One of the most important implications of the classic papers of Solow (1956) and Swan (1956) is that the lower the starting level of real per capita GDP, relative to the long run or steady state position, the faster is the growth rate. The Solow-Swan model assumes a constant-returns-to-scale production function with two inputs, capital and labor, and substitution between inputs, a constant savings rate, and constant rate of growth of population and neutral technical change, all exogenously given. Convergence of economies starting out at different levels of per capita income to the same steady-state rate of growth reflects the diminishing returns to capital implied by the production function assumed: economies starting out with lower levels of real per capita GDP relative to the long run or steady state position have less capital per worker and therefore higher rates of return to capital. I will refer to this as the standard Barro-Baumol (BB) sense of the meaning of convergence. There is a good deal of current discussion regarding the appropriate definition of "convergence."[1] My purpose here is not to question this notion of convergence but rather to show that estimates of the coefficient of the lagged dependent variable in a dynamic panel model which has been used to study this phenomenon is extremely sensitive to the method of estimation employed. Because the steady states of the Solow-Swan model depend on the savings rate, the rate of growth of population, and the rate of technical progress, many have argued that these factors need to be held constant in attempting to test the hypothesis of growth rate convergence. Convergence is, in this sense, conditional. Convergence may also be conditional on other factors such as the accumulation of human capital or investment in infrastructure or, indeed other unobserved factors which are trending at different rates in different countries or regions.[2]

The problem of BB-convergence in the standard neoclassical model is treated both theoretically and empirically in the recent text by Barro and Sala-i-Martin (1995) and empirically in a recent paper by

---

[1] Barnard and Durlauf (1995) give a nice discussion emphasizing the restrictiveness of the approach, going back to the earlier work of Barro (1991) and Baumol (1986), on which my analysis of econometric methods is based (see also Quah, 1996), and which is the most prevalent. Other recent contributions include Islam (1995b), Casselli, Esquivel, and Lefort (1996), and the fine survey by de la Fuente (1997).

Islam (1995a). Bernard and Durlauf (1996) provide a useful framework for understanding the time-series and cross-sectional tests of the BB-convergence hypothesis and its relation to alternative definitions. Quah (1996) discusses the problem of convergence in more general form and distinguishes several different varieties. He argues that "Simply because panel data techniques happen to apply to data with extensive cross-section and time-series variation does not mean they are at once similarly appropriate for analyzing convergence." While I do not fault Quah's conclusion, current discussions do emphasize panel data and methods and derive strong conclusions regarding BB-convergence and the significance of other determinants of growth from such data. It is therefore appropriate to consider how these conclusions, *within the context of BB-convergence*, are affected by the econometric methods employed.

Perhaps even more important than the problem of convergence is the question of the determinants of growth. The World Bank Project on Economic Growth lists 15 published papers and 15 working papers almost all of which involve dynamic panel data analysis or cross-section analysis with a state variable or initial condition.[3] Although the focus of these papers is not convergence but the effects of the other variables included, if the coefficient of the state variable in the statistical analysis is inconsistently estimated, in this sense "biased," then the coefficient of any variable correlated with the state variable will also be biased. Hence, quite misleading conclusions may be drawn concerning the significance, sign and relative magnitude of other factors included in the analysis, conclusions which may significantly affect the policy implications of the analysis.

Section 2 examines recent empirical investigations of BB-convergence and the rate of convergence and argues that most are flawed by failure to allow for the inconsistencies of single cross-section or panel studies in a dynamic context.[4] In a dynamic context a single cross-section is best viewed

---

[2] When population growth is endogenously determined, this implication of convergence, conditional or otherwise, of the neoclassical model of economic growth does not necessarily follow; see Nerlove and Raut (1996).

[3] See http//www.worldbank.org/html/prdmg/grthweb.

[4] This point is also made by Caselli, et al. (1996) in a study which came to my attention after Nerlove (1996b) was written. A recent study by Lee, et al (1996) arrives at similar conclusions but proposes a number of alternatives different from those investigated here. In particular, I believe that their formulation of what they call the unconditional likelihood function is quite different from that proposed here because they include deterministic trends in their model and cannot therefore directly formulate the likelihood *unconditional on the initial observations* (see footnote 11).

as a panel with time dimension 1. I do not attempt here a general review of the effects of the methods used in more general studies of the determinants of growth, but elsewhere I have examined the effects on the estimated coefficients of the Barro-Lee (1993) estimates of the stock of human capital.[5]

In section 3, I discuss four common methods of estimating the coefficient of the "state" variable interpreted in terms of the rate of convergence (and thus testing for convergence), show that these four methods yield estimates which satisfy an inequality derived by Trognon and Sevestre (1996). One broad class of estimates with which I do not deal here are those based on the Generalized Method of Moments and due to Chamberlain (1984), the so-called Π-matrix method, or derived from his work (Holtz-Eakin, Newey and Rosen, 1988; Arellano and Bond, 1991).[6] These methods are not only somewhat difficult to implement but the resulting estimates are, by construction, insensitive to the way in which initial values of the "state" variable have presumably been generated. Below I argue that, if the process generating the data in the sample period is stationary, or can be made so by suitable transformation of the data, the initial values of the state variable convey a great deal of information about that process since they reflect how it has operated in the past. Thus conditioning on those initial conditions is clearly an undesirable feature especially when the time-dimension of the panel is short.

In section 3(f), I present a new method of maximum-likelihood estimation based on the density of the observations *unconditional on the initial or starting values of the dependent variable.* I argue more generally for methods of inference which look at more than just the maximum of the likelihood function, on the basis of the *likelihood principle* of Fisher (1922; 1925). This approach fully takes into account what information the initial conditions contain about how the process has operated in the past and is thus of special relevance to short time-dimension ("shallow") panels. I extend this method to the case of

[5] See the appendix to Nerlove(1996b).

[6] For an extensive exposition of these methods and a very general formulation see Crépon and Mairesse, 1996. The method is based on a series of transformations of the basic equation (1); the resulting equation is then estimated as a cross-section regression on the original explanatory variables in all periods. Caselli, et al., use a variant which is an application of the GMM method applied to the first differences of the series. They then use the stock or state variables as instruments. This implies that they are predetermined and therefore conditions on their values. If indeed the process, originally or in first-difference form is stationary, their procedure does therefore discard the information in the initial observations about it, which is just what unconditional ML seeks to avoid. Since conditional ML should give estimates with the

country-specific trends. These make the underlying processes being investigated nonstationary, but simple forms of nonstationarity can be removed by differencing the data.

Finally, in section 4, I apply all six methods to two panel data sets drawn from the Penn World Tables, for both a model with country-specific intercept effects and one with country-specific trends. The results show clearly how misleading the standard estimates can be in assessing growth rate convergence and in the estimation of the significance and magnitude of other variables included. The contrast between the conditional and the unconditional ML estimates for a small cross-section dimension and their similarity for a large cross-section dimension is illustrated, as is the importance of looking at the likelihood function itself more broadly. I also show that the usual procedures for doing feasible GLS or for obtaining starting values for ML are seriously flawed and likely to yield negative estimates of the random time persistent cross-sectional effects. The results also show that biases in the estimate of the coefficient of the lagged value of the dependent variable are transmitted to the estimates of other coefficients in the model, making inferences about the determinants of growth problematic unless appropriate econometric methods are used.

Section 5 concludes.

2. Recent Empirical Investigations of Convergence and the Rate of Convergence

Let y_t = per capita output, k_t = the capital-labor ratio, s = the savings rate, δ = the depreciation rate of capital, and n = the exogenous rate of population growth and labor force. All of these variables may differ over time as indicated by their subscript t, but also, in a cross-country context, they are certain to differ from one country to another in a fashion which persists over time. An additional subscript is introduced in the sections which follow this one to indicate that fact. If the production function is Cobb-Douglas, $y_t = A_t k_t^{\alpha}$, where A_t reflects other than conventional factors of production affecting growth and where α, the elasticity of per capita output with respect to the capital-labor ratio, is often interpreted in terms of capital's share as implied by payment of capital at its marginal product. Under these circumstances it can easily be shown using a simple partial adjustment model that

same desirable properties as GMM and is very easy to compute under these circumstances, it is not clear to me why Caselli, et al., and others have avoided its use.

$$\log y_t = \frac{\alpha(1-\gamma)}{1-\alpha}[\log s - \log(n+\delta)] + \frac{1-\gamma}{1-\alpha}\log A_t + \gamma \log y_{t-1}. \tag{1}$$

The speed of convergence to equilibrium is inversely proportional to γ. With growth convergence $0 < \gamma < 1$. In equilibrium, per capita GDP depends only on the parameters n, s, and the time path of A. In an empirical context, these differ from time to time and country to country. Clearly the extent of convergence is conditional on s, n, δ and the time path of A_t. In empirical investigations, changing n and s and sometimes a measure of changing A have been introduced. Below I examine models in which A is assumed to be constant although differing from one country to another and an alternative formulation in which A_t can be represented by a simple linear trend which plausibly also differs from country to country.

Equation (1) has been widely used to examine the hypothesis of growth convergence (Mankiw, et al . 1992, p.410; Barro and Sala-i-Martin, 1995, Chapter 12; Islam, 1995, p. 1133; Lee, et al. 1996, Casseli, et al. 1996). In empirical work, y_t is replaced by real per capita GDP; when varying s and n are taken into account, s is replaced by an average savings rate over the period t-1 to t, and n is replaced by the growth rate of population over the period t-1 to t. It is usual to use rates averaged over several years; following Islam (1995), I have used quinquennial averages. The restriction on the coefficients of ln(s) and ln(n+δ), which arises from the constant-returns-to-scale assumption implies that ln(s) and ln(n+δ) can be collapsed into a single variable. Testing the growth convergence hypothesis, in this context, revolves largely around the coefficient γ of the lagged level of per capita real GDP. If this is positive but much less than one, the implication is that on average countries with low initial values are growing faster than those with high initial values and is therefore evidence of convergence. Whereas if this coefficient is close to one, perhaps even slightly larger than one, the implication is that initial values have little or no effect or even a perverse one on subsequent growth; such a finding is therefore evidence against the neoclassical theory which implies convergence. For example, if $\gamma = 0.9$, convergence to within 90% of final equilibrium occurs only in 22 periods, which, given quinquennial data, implies 110 years! Similarly, 0.8 requires 53 years, 0.7 32 years, while 0.2 requires only 7 years and 0.1 is within 90% in 5 years.[7]

[7] Derivation of the model and the calculations behind this statement are given in Nerlove(1996, pp. 5 –8 and Appendix Table 2).

The estimates of γ for the levels model presented below using cross-country quinquennial data are generally in excess of 0.7 no matter what econometric procedure is employed. but vary over a wide range depending on the method, 0.7 to 0.98.But for the differenced model. many estimates of γ are much smaller, in the vicinity of 0.5.[8] It is apparent that, for all practical purposes, coefficients in excess of 0.7 represent negligible convergence, since, with unchanging s, n, and A, it would take more than a generation to achieve 90% of equilibrium real per capita GDP. Most recent work attempts to test whether $\gamma = 1$; however, this is a test for unit root in log y_{it}. Even under the best of circumstances testing for a unit root is problematic (see Diebold and Nerlove, 1990). Here the problems are compounded by the short time dimension of the typical panel. Basing a test on the size of γ rather than equality with 1 finesses a host of problems of the sort discussed extensively in Diebold and Nerlove.[9]

Tests based on a single cross-section (which can be viewed as a panel of time dimension 1) or on pooled cross-section time series (panel) data generally have yielded contradictory results: Pooled panel data studies tend to reject the hypothesis of BB-convergence (relatively high γ's), even after controlling for population growth rates, savings rates and other variables. Dynamic fixed-effects models are of course not possible for a single cross-section, but recent work (Islam, 1995a) using a dynamic fixed-effects panel model yields results supporting convergence. There are serious problems with tests such as these which rely on the estimated coefficients of the lagged, or initial value, of the dependent variable in dynamic panel models, or in the special case of a single cross-section, which arise from two sources of bias. In this paper, I show that some of these findings are probably statistical artifacts arising from biases in the econometric methods employed. This demonstrates the sensitivity of the conclusions drawn about γ to the econometric method employed, irrespective of the validity of the relationship estimated.

The first source of bias are omitted variables, especially infrastructure and investments over time in infrastructure, and the natural resource base available to each country in cross-sectional or panel

---

[8] Using a GMM estimator Caselli, et al., obtain an estimate of about 0.51 - 0.53, i.e., much more rapid convergence and close to the estimates obtained for the 94-country sample using either conditional or unconditional ML. My estimates for the 22-country sample are much higher, however.

[9] Barnard and Durlauf (1995) use cointegration techniques on rather longer time series for 15 OECD countries to test alternative time-series definitions of convergence and contrast the results with the standard BB-formulation.

studies. Systematic differences in these across countries or regions will systematically bias the conclusions. Because such variables are likely to be correlated with savings or investment rates in conventional or in human capital and with population growth rates it is not altogether clear what the net effect of omitting them on the coefficient of the initial value will be in a single cross-section. But in a pooled model it is clear that, to the extent such differences are persistent, they will be highly correlated with the initial value and therefore omitting them will bias the coefficient of that variable upwards towards one and thus towards rejecting convergence. This source of bias has been well-known since the early paper by Balestra and Nerlove (1966) and is well-supported by the Monte Carlo studies reported in Nerlove (1971). In this light, it is not surprising that pooled panel data, or single cross-sections, which are a special case of panels with T = 1, even with inclusion of additional variables, often reject convergence.

Second, since there are likely to be many sources of cross country or cross region differences, many of which cannot be observed or directly accounted for, it is natural to try to represent these by fixed effects in a panel context. But, as is well-known from the Monte Carlo investigations reported in Nerlove (1971) and demonstrated analytically by Nickell (1981), inclusion of fixed effects in a dynamic model biases the coefficient of the initial value of the dependent variable included as an explanatory variable downwards, towards zero and therefore towards support for the convergence hypothesis. This may account for Islam's (1995a) recent findings.

Alternative estimates based on more appropriate random-effects models, such as two-stage feasible Generalized Least Squares or maximum likelihood conditional on the initial observations are also biased in small samples and inconsistent in large, or in the case of Instrumental Variable estimates have poor sampling properties or are difficult to implement. For example, the papers by Knight, Loayza and Villanueva(1993), Loayza (1994), and Islam (1995a) employ a method, among others, proposed by Chamberlain (1984), generally referred to as the Π-matrix approach.[10] The alternative of unconditional

---

[10] See also Crépon and Mairesse (1996).

maximum likelihood suggested in Nerlove and Balestra (1996) is implemented for the first time in this paper.[11]

Even if one has little interest in the question of convergence, or its rate, per se, the question of whether the coefficient of the state variable, lagged dependent or initial value, is biased in the sense of being inconsistent is an important one since biases in this coefficient will affect the estimates of the coefficients of other variables correlated with it and their levels of significance. To the extent such estimates are important in the formulation of policies to promote growth, the matter is indeed a serious one.[12]

In the remainder of this paper, I investigate the sensitivity of the coefficient of the lagged dependent or state variable to the econometric method employed as well as the sensitivity of the estimates of the coefficients of other variables included. All of the results reported, except those for pooled panel data, support the growth convergence hypothesis conditional on savings and population growth rates but illustrate the rather different estimates of the rates of convergence, and of the coefficients of other explanatory variables, obtained when different estimation techniques are used. In addition, a technique for examining the shape of sections of a high-dimensional likelihood function is developed which reveals interesting and somewhat unexpected relationships among the various estimates.

3. Alternative Methods for Estimation[13]

A good summary of the current state of knowledge about the properties of various estimators in dynamic panel models is contained in Sevestre and Trognon (1992, 2nd. ed. 1996). Trognon (1978) was the first

---

[11] Lee, et al. (1996) also estimate from what they maintain is an unconditional likelihood function, but inasmuch as they do not transform to stationarity (their relationship includes both a constant and a linear trend), I do not think their formulation of the likelihood function is based on the unconditional density of the dependent variable as proposed here. In fact, they estimate from a likelihood based on the *conditional* density of the dependent variable given the initial value. The relation between conditional and unconditional likelihood is discussed at length in Nerlove(1997).

[12] For example in (1) the parameter α could be derived from the coefficient of the variable $\log s - \log(n+\delta)$ as coefficient/(coefficient $+1-\gamma$), so there is a double source of bias. Indeed, a number of authors accept or reject statistical formulations based on the estimated value of α which should approximate capital's share.

to show the possible inconsistency of maximum likelihood conditional on the initial individual observations. Nickell (1981) shows the inconsistency of the estimates of the fixed-effects in a dynamic panel model. Kiviet (1995) derives exact results for the bias of leading estimators. In this section, following Sevestre and Trognon, I review the leading estimators and their properties for dynamic panel models. I will assume a random effects model for the disturbance for the reasons set forth in Nerlove and Ballestra (1996) and because fixed effects can be viewed as a special case from the standpoint of estimation.

For simplicity, in this section I restrict attention to the simple model containing one exogenous variable x_{it} and one lagged value of the dependent variable y_{it-1} as explanatory. Extension to the case in which more than one exogenous explanatory variable is included presents no serious difficulty.

(2) $$y_{it} = \alpha + \beta x_{it} + \gamma y_{it-1} + \mu_i + \varepsilon_{it} \text{ , i=1,...N, t=1,...T.}$$

Taking deviations from overall means eliminates the constant α. The usual assumptions are made about the properties of the μ_i and the ε_{it}:

(i) $E(\mu_i) = E(\varepsilon_{it}) = 0$, all i and t,

(ii) $E(\mu_i \varepsilon_{jt}) = 0$, all i, j and t,

(iii) $$E(\mu_i \mu_j) = \begin{cases} \sigma_\mu^2 & i = j \\ 0 & i \neq j, \end{cases}$$

(iv) $$E(\varepsilon_{it}\varepsilon_{js}) = \begin{cases} \sigma_\varepsilon^2 & t = s,\ i = j \\ 0 & otherwise \end{cases}.$$

Both μ_i and ε_{it} are assumed to be uncorrelated with x_{it} for all i and t. While this assumption is far from innocuous, for example, if savings rates or population growth is not independent of per capita income or unobserved factors which affect it, I adopt it here, not only because it is conventional but because one has to cut off somewhere. Clearly, however, y_{it-1} cannot be assumed to be uncorrelated with μ_i. It is clear, therefore, that

---

[13] I rely extensively in this section on the excellent discussion of Sevestre and Trognon, Chapter 7 in Mátyás and Sevestre (1996, pp.120-144). Additional alternatives, more appropriate when longer time series are available, are treated by Lee, et al. (1996) and are not discussed or implemented here.

OLS applied to (2) ignoring the component nature of the disturbances $u_{it} = \mu_i + \varepsilon_{it}$, which I call the *pooled regression*, will yield inconsistent estimates. In particular, if $\gamma > 0$, γ_{pooled} is "biased" upwards. So, just as in the case of ordinary serial correlation, β_{pooled} is also "biased" and the OLS residuals understate the amount of serial correlation, which in this case is measured by the intraclass correlation coefficient ρ. This parameter measures the extent of unobserved or latent time-invariant, individual-specific, variation relative to the total unobserved variation in the sample, $\sigma_\mu^2 / (\sigma_\mu^2 + \sigma_\varepsilon^2)$. It is extremely important in understanding the nature of the variation, both observed and unobserved, in the panel.

(a) Inconsistency of the pooled-sample OLS estimates of the dynamic error-components model.

Since the panel has two dimensions, it is possible to consider asymptotic behavior as $N \to \infty$, $T \to \infty$, or both. Generally speaking it is easier to increase the cross-section dimension of a panel, so the most relevant asymptotics are as $N \to \infty$. This is called *semi-asymptotics* in the panel data literature. It is not necessary to assume $|\gamma| < 1$ as long as T is fixed, but the way in which the initial values of the dependent variable, y_{i0}, are assumed to be generated is crucial. To see why, write (2) as

$$(3) \qquad y_{it} = \gamma^t y_{i0} + \sum_{j=0}^{t-1} \gamma^j \beta x_{it-j} + \frac{1-\gamma^t}{1-\gamma}\mu_i + \nu_{it}, \text{ where } \nu_{it} = \sum_{j=0}^{t-1} \gamma^j \varepsilon_{it-j}.$$

Equation (3) expresses y_{it} as the sum of four terms: the first, $\gamma^t y_{i0}$, depends on the initial values; the second on lagged values of the exogenous variable; the third on the individual, time-invariant, component of residual variance; and the fourth on lagged values of the remaining component. This last term is an autoregressive process with initial values $\nu_{i0} = 0$ and $\nu_{it} = \gamma\nu_{it-1} + \varepsilon_{it}$. It need not be assumed to be stationary as long as T is fixed. It does not make sense in this context to assume that the y_{i0} are uncorrelated with either the μ_i or the lagged values of the x_{it}'s. On the other hand, ε_{i0} is a random variable with mean 0 and variance σ_ε^2 independently and identically distributed for all i. Thus, the initial observation can be written as a function of lagged x's, the μ_i and ε_{i0}:

$$(4) \qquad y_{i0} = f(x_{i0}, x_{i-1}, \ldots, \mu_i, \varepsilon_{i0}).$$

Clearly, if the individual effects μ_i are assumed to be fixed and the lagged x's to be given, the y_{i0} are also fixed and uncorrelated with the disturbances in (3), ν_{it}, t = 1,...,T. But, if the individual effects are considered to be random, as Nerlove and Balestra (1996) have argued they should be, the initial observations are not exogenous since they are correlated with them, as they are part of the disturbance term, namely the third and fourth terms of (3).

It is common in the literature on panel data to assume that the y_{i0} are i.i.d. random variables which are characterized by their second moments and correlations with the individual effects and not necessarily generated by the same process which generates the rest of the y_{it}'s. The properties of various estimators depend on the process generating them. One possibility is to try to model and estimate this process together with the dynamic panel model (2).

(b) Inconsistency of the OLS Estimators of the Dummy Variable, or Fixed-Effects, Model.

The ordinary least squares estimates of both the coefficient of the lagged dependent variable and the exogenous variable are inconsistent in the fixed effects model. As is well-known, the fixed effects model is equivalent to taking deviations from individual (country) means and then estimating an ordinary OLS regression:

(5) $$y_{it} - \bar{y}_{i.} = \beta(x_{it} - \bar{x}_{i.}) + \gamma(y_{i,t-1} - \bar{y}_{i,-1}) + \nu_{it}, \text{ where}$$

$$\nu_{it} = \varepsilon_{it} - \bar{\varepsilon}_{i.}\ .$$

Although $\sigma_{x\varepsilon} = 0$,

(6) $$\begin{aligned}\sigma^2_{y(-1)\varepsilon} &= \underset{N \to \infty}{p\lim} \frac{1}{T}\sum_t \left(y_{i,t-1} - \bar{y}_{i,-1}\right)\left(\varepsilon_{it} - \bar{\varepsilon}_{i-1}\right) \\ &= -\frac{1}{T^2}\frac{T-1-T\gamma+\gamma^T}{(1-\gamma)^2}\sigma^2_\varepsilon \neq 0.\end{aligned}$$

Thus, the OLS estimates of both β and γ in the fixed effects model are inconsistent, although as $T\to\infty$, the inconsistency disappears. But for finite, typically small T, it remains. (See Nickell, 1981, p.1424). For T = 10 and $\gamma = 0.5$, for example, the "bias" of the OLS estimate of γ, say c, is proportional to -0.16, the factor of proportionality being the OLS estimate of the variance of c from the within regression. It is always negative, implying that the bias of the OLS estimate of β, say b, is therefore upward. This conclusion holds regardless of whether one assumes the true model is fixed- or random-effects.

Although the inconsistency will be small when T is moderate to large, small values of T are typically the case. Nonetheless, Nerlove (1971) suggested using the fixed effects model to estimate ρ for FGLS, in contrast to the earlier suggestion of Balestra and Nerlove (1966), hereinafter BN, of a consistent instrumental variable approach. BN also suggested but did not implement a method based on estimating ρ from the pooled and fixed-effects regressions. Rejection of instrumental variables by Nerlove (1971) was based on the instability of the results in Monte Carlo trials. Since the OLS estimates of the parameters from pooled or fixed-effects regressions are inconsistent, the estimates of ρ based on this regression will not be either; hence, the FGLS estimates computed using them will not generally be consistent. In the results reported here, an estimate of ρ is derived from the estimates of residual variance from both the fixed-effects and the pooled regressions, as suggested by B and N (1966), and is not consistent.

Many authors (e.g., Greene, 1993, pp. 475-477, Judge, et al., pp. 484-488), hereinafter GJ, suggest basing an estimate of ρ on the cross-section regression of the overall means and either the pooled or fixed-effects regression. This suggestion, unfortunately often leads to negative estimates of ρ and unwarranted rejection of the model. These estimates are also inconsistent. The GJ suggestion is, unfortunately, utilized in most computer packages for implementing FGLS for panel data or obtaining starting values for ML, and often leads to the adoption of badly biased fixed-effects OLS when a negative estimate of ρ is obtained.

The GJ suggestion is to regress the group means of the independent variable on the group means of the dependent variables:

(7) $$\bar{y}_{i.} = \alpha + \beta \bar{x}_{i.} + w_i \text{, where } w_i = \mu_i + \bar{\varepsilon}_{i.} .$$

The variance of w_i is $\sigma_\mu^2 + \frac{\sigma_\varepsilon^2}{T}$. The purely cross-sectional variation of the individual means gives us information on both the slope and the overall constant in the regression. This is often called the ***between groups*** regression. In many panel data problems purely cross-sectional variation may dominate, but this variation may not give us much information about the true value of the slope of the independent variable if the regression also contains a lagged value of the dependent variable. The residual SS/N = RSSB/N from this regression estimates $\sigma_\mu^2 + \frac{\sigma_\varepsilon^2}{T}$. But it will not be a very good estimate if the regression is estimated by OLS, since (7) will tend to fit too well if cross-section variation dominates the data.[14] σ_μ^2 is then estimated as

$\sigma_{w}^2 - \frac{\sigma_\varepsilon^2}{T}$, where an estimate of σ_ε^2 can be obtained from the fixed-effects regression. If T is large, the estimated value of σ_μ^2 is not likely to be negative no matter how well the between groups regression fits. But if T is small, and particularly if the regression contains a lagged value of the dependent variable on the right-hand side, the chances of obtaining a negative, and therefore unacceptable, estimate of ρ are high irrespective of the validity of the model.[15]

(c) Generalized Least Squares and Feasible GLS.

The means or between regression and the fixed-effects regression both contain information about the parameters of the model: The means regression reflects purely cross-sectional variation; whereas the fixed-effects regression reflects the individual variation over time. GLS combines these two types of information with weights which depend on the characteristic roots of $Euu' = \sigma^2 \Omega$. The individual means themselves are weighted by the reciprocal of the square root of $\xi = 1 - \rho + T\rho$, while the deviations from these means are weighted by the reciprocal of the square root of $\eta = 1-\rho$. A representative transformed observation is

$$y_{it}^* = \xi^{-1/2}\bar{y}_{i.} + \eta^{-1/2}(y_{it} - \bar{y}_{i.}), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T.$$

---

[14] For example, when a lagged value of the dependent variable is included as one of the explanatory variables, its mean may be very close to the mean of the unlagged variable; then the fit of (7) may be nearly perfect. The estimated residual variance may be close to zero in this case. In general, if there is a lot of associated cross-sectional variation, the residual of this relationship may be very small. If combined with the estimate of σ_ε^2 obtained from the within regression, the implied estimate of σ_μ^2 may well turn out to be negative (see Greene, pp. 474-476). But this does not imply that the model is misspecified. Balestra and Nerlove (1966, p.607) suggest estimating σ_μ^2 from the fixed-effects model as the "variance" of the implied constant terms: $\sigma_\mu^2 = \frac{1}{N}\sum_i\left(\bar{y}_{i.} - \bar{y}_{..} - \hat{\beta}(\bar{x}_{i.} - \bar{x}_{..})\right)^2$, where $\hat{\beta}$ is the OLS estimate of β in that regression. This suggestion is the one implemented in Nerlove(1971). Alternatively, if a regression with dummy variables for each individual, overall constant suppressed, has been estimated, it suffices to compute the variance, sum of squared deviations from the mean value divided by N, to estimate σ_μ^2.

[15] In footnote 21, section 5, I compare three different estimates of ρ based on the above OLS regressions: (1) the original suggestion of Balestra and Nerlove (1966); (2) an estimate based only on the fixed-effects regression (Nerlove, 1971); and the suggestion of Judge (1988) and Greene (1993), which is the basis for most computer packages doing panel econometrics. The first, and earliest, suggestion generally yields results closer to the maximum likelihood estimates than the others. The second yields estimates which are generally considerably higher. The last yields results which are far too low and often found to be negative, although not in the results reported here. I have presented estimates of ρ obtained by all three methods and compared them with estimates of ρ obtained from conditional and from unconditional maximum likelihood below in the Table in the footnote.

Thus y^*_{it} is a <u>weighted</u> combination (weighted by the reciprocals of the square roots of the characteristic roots of Ω) of individual means of the original observations $\bar{y}_i.$ and deviations from individual means $(y_{it} - \bar{y}_i.)$. The other variables are similarly transformed to x^*_{it} and $y^*_{it}(-1)$. GLS amounts to running the OLS regression:

(8) $$y^*_{it} = \alpha^* + \beta x^*_{it} + \gamma y^*_{it-1} + \nu_{it}.$$

ν_{it} is the transformed disturbance. Note that the constant has a different interpretation.

Let $\theta^2 = \eta / \xi = (\xi^{-1/2} / \eta^{-1/2})^2$ be the relative weight of the between variation to the within variation. Note that this weight tends to 0 from above as $T \to \infty$, i.e., the within information dominates. For T small, $\theta^2 < 1$, so that the between variation is allowed to play a greater role. When the intraclass correlation, ρ, is close to one, the unobserved, residual cross-section variation is large relative to the unobserved individual variation. $\theta^2 = \dfrac{1}{1 + T\dfrac{\rho}{1-\rho}} = \dfrac{1}{1 + T\dfrac{\sigma^2_\mu}{\sigma^2_\varepsilon}}$ is smaller for fixed T than when ρ is small. Between information gets a lower relative weight when ρ is large than when ρ is small, which corresponds to the usual weighting of data from sources with varying degrees of error.

To obtain an estimate of ρ for use in a feasible GLS, I prefer to obtain both an estimate of σ_ε^2 from a fixed-effects model and then an estimate of σ^2 from the pooled regression, as indicated above. Although this estimate is not consistent, it is never negative and, empirically it gives, at least the appearance of, a tighter upper bound to the true value of γ than the pooled regression does and a closer approximation to the ML estimate.

(d) Bounds for the Coefficient of the Lagged Dependent Variable.

As Maddala (1971) has pointed out, the GLS estimates with $\lambda = 1/\theta^2$ can be considered members of a more general class of estimators obtained through different choices of λ. Let $\hat{\gamma}(\lambda)$ be the estimator of γ obtained by solving the GLS normal equations for an arbitrary value of λ. Sevestre and Trognon (1996, pp. 130-133) show that for the case in which $\beta = 0$, the purely autoregressive case, the following inequality holds:

(9) $$\underset{\text{fixed-effects}}{p\lim\hat{\gamma}(0)} < \gamma < \underset{\text{GLS}}{p\lim\hat{\gamma}(\lambda)} < \underset{\text{OLS pooled}}{p\lim\hat{\gamma}(1)} < \underset{\text{means}}{p\lim\hat{\gamma}(\infty)}.$$

Remarkably, the GLS estimate is inconsistent even when a consistent estimate of ρ is used to compute FGLS estimates. The problem is that the lagged dependent variable is correlated even with the transformed disturbance.

Since $p\lim\hat{\gamma}(\lambda)$ is a continuous function of λ, there exists a value λ* in the interval [0, $1/\theta^2$] for which $p\lim\hat{\gamma}(\lambda)=\gamma$. Sevestre and Trognon (1983) show that this value is

(10)
$$\lambda^* = K(1-\rho)/\{\frac{(1-\gamma^T)E(y_{i0}\mu_i)}{(1-\gamma)\sigma^2}+K\xi\},$$
$$\text{where } K=\frac{T-1-T\gamma+\gamma^T}{T(1-\gamma)^2}, \text{ and } \rho,\ \xi, \text{ and } \sigma^2 \text{ are as before.}$$

They also show that when β ≠ 0, the estimate $\hat{\gamma}(\lambda)$ behaves almost the same as in the purely autoregressive case. Since the λ* estimate is consistent when there are no exogenous variables, it remains so when there are. The trick is to obtain a consistent estimate of λ* which can be accomplished by finding an appropriate instrumental variable for y_{-1}. Even in this case the results depend heavily on the distribution of the estimate of λ*.

In the dynamic error-components model, not only are the OLS pooled regression estimates, the fixed-effect or within estimates, and the between estimates inconsistent, but so are the GLS estimates using the true value of ρ. However, the method of instrumental variables may be used to obtain a feasible member of the λ-class of estimates which is consistent. (See Sevestre and Trognon, 1996.) Unfortunately, this estimate may have a very large variance, as demonstrated in Nerlove (1971).

Nonetheless, the fixed-effects and the pooled regressions may be used to bound the true value of γ even when exogenous regressors are also included. Empirically, I have found that FGLS appears to provide an even tighter bound, although since FGLS is also based on an inconsistent estimate of ρ, there is no guarantee that this is in fact an upper bound.

(e) Maximum Likelihood Conditional on the Initial Value of the Lagged Dependent Variable.

When the likelihood function for the model (2) with $u_{it} = \mu_i + \varepsilon_{it} \sim N(0, \sigma^2\Omega)$ is derived in the usual way from the product of the densities of y_{it} conditional on x_{it} and y_{it-1}, the joint density is conditional on y_{i0}. This likelihood function can be written in terms of the earlier notation introduced as

(11) $$\log L(\alpha,\beta,\gamma,\sigma_\mu^2,\sigma_\varepsilon^2 | y_{11},\ldots,y_{NT};x_{11},\ldots x_{NT};y_{10},\ldots,y_{N0})$$

$$= -\frac{NT}{2}\log 2\pi - \frac{NT}{2}\log\sigma^2 - \frac{N}{2}\log\xi - \frac{N(T-1)}{2}\log\eta$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}^{*} - \alpha\xi^{-1/2} - \beta x_{it}^{*} - \gamma y_{it-1}^{*}\right)^2 ,$$

where y*, x* and y^*_{-1} are the transformed variables. Since

$\xi = \dfrac{\lambda}{1-(\lambda-1)/T}$ and $\eta = \dfrac{1}{1-(\lambda-1)/T}$, logL can be expressed as a function solely of λ, σ^2, α, β, and γ. Trognon (1978) shows that, when the exogenous variable x is generated by a first-order autoregression with white noise input, w~ wn (0, σ_w^2 I), also assumed in the Monte Carlo experiments reported in Nerlove (1971),

(12) $$x = \delta x_{-1} + w \quad ,$$

maximization of the conditional likelihood function (12) yields boundary solutions $\hat{\rho} = 0$, which, unlike interior maximum likelihood solutions, are inconsistent, for a considerable, and indeed likely, range of parameter values. In particular, there is a value of γ in (2),

$$\gamma^* = \frac{(T-3)^2 - 8}{(T+1)^2} ,$$

such that when $\gamma < \gamma^*$ there exists an interior maximum of (11) which yields consistent ML estimates, but that when $\gamma \geq \gamma^*$ there are values of ρ for which the conditional likelihood function (2) is maximized at the boundary $\rho = 0$, i.e., for the OLS estimates of the pooled regression , which we know to be inconsistent. The problem is that when T is small the permissible range of γ, the coefficient of the lagged dependent variable is implausible (e.g., negative or very small). For example, for T = 5, $\gamma^* = -0.11$, while for T = 10, $\gamma^* = 0.34$. When $\gamma \geq \gamma^*$, whether or not an interior maximum with consistent ML estimates occurs depends on the value of ρ: For $\rho < \rho^*$ boundary maxima occur where

$$\rho^* = \left(\frac{T-1}{T+1}\right)^2 \frac{\beta^2\sigma_w^2}{\sigma^2} \frac{1-\gamma}{(\gamma-\gamma^*)(1-\gamma\delta)^2} .$$

For example, when T = 5, $\beta = 1.0$, $\gamma = 0.75$, $\delta = 0.5$, and $\frac{\sigma_w^2}{\sigma^2} = 1.0$, $\gamma^* = -0.11$ and the critical value of ρ is $\rho^* = 0.31$. That means that any true value of the intraclass correlation less than 0.31 is liable to produce a boundary solution to (11) $\rho = 0$ and inconsistent estimates of all the parameters. Using these results. Trognon (1978) is able to replicate the Monte Carlo results reported in Nerlove (1971).

Even though ML may yield inconsistent estimates when the nonnegligible probability of a boundary solution is taken into account, it is nonetheless true that the likelihood function summarizes the information contained in the data about the parameters.[16] From a conventional. Neyman-Pearson 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. That only the maximum and the Hessian at the maximum are all the matters from a conventional 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 and because of the central limit theorem in its many forms many estimators, including ML estimators in regular cases, tend to normality in distribution. So the problem of possible inconsistency of the ML estimates should not concern us unduly from the standpoint of likelihood inference. It is the whole shape of the likelihood function which expresses what the data have to say about the model and its parameters which matters.[17] For this reason, sections of some of the multidimensional likelihood functions are also presented in the next section. When first

[16] Although clearly implied in what Fisher wrote in the 1920's (Fisher, 1922 and 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) written largely in reaction to both the classical Neyman-Pearson (frequentist) and the Bayesian approaches to inference. A good recent discussion is Lindsey (1996).

[17] The principle of likelihood inference and its application to dynamic panel models is elaborated in Nerlove(1997). A maximum at the boundary conveys perfectly valid information about the parameter in question, just as does a near-plateau solution at which the assymptotic standard errors derived from the information matrix are huge. More importantly the existence of two or more local maxima at not very

differences are taken to eliminate a linear deterministic trend, the individual-specific time invariant effects become differences in the trend slopes. This makes the interpretation of the model in first-difference form different than that in levels. Moreover, the time- and individual varying disturbance is now likely to be serially correlated, a fact which needs to be taken into account in the formulation of the unconditional likelihood function. A parallel set of results for the country-specific trends model is presented below.

(f) Unconditional Likelihood and Unconditional Maximum Likelihood.

While it is not guaranteed that a boundary solution to the likelihood equations is obtained, which would yield ML estimates which are inconsistent, it is apparent, as suggested above, that in panels with a short time dimension the initial values provide important information about the parameters of the model, and to condition on them is to neglect this information.

It is not, in fact difficult to obtain the unconditional likelihood function once the marginal distribution of the initial values is specified. The problem is a correct specification of this distribution. If $|\gamma| \geq 1$ or the processes generating the x_{it} are not stationary, it will not, in general be possible to specify the marginal distribution of the initial observations. I will assume that, possibly after some differencing, both the y_{it} and the x_{it} are stationary. The derivation of the unconditional likelihood function in the case in which deterministic or stochastic trends are included is contained in Nerlove (1997).[18]

different likelihood values but widely separated values of the parameters, such as I have obtained in the case of regional Indonesian data, is even more revealing.

[18] Adding trend, t, to (2)

(2') $y_{it} = \alpha + \beta x_{it} + \gamma y_{it-1} + \tau_i t + \mu_i + \varepsilon_{it}$, i=1,...N, t=1,...T, and differencing,

(2'') $\Delta y_{it} = \beta \Delta x_{it} + \gamma \Delta y_{it-1} + \tau_i + \Delta \varepsilon_{it}$, i=1,...N, t=1,...T,

where Δ denotes the first-difference operator and τ_i is the individual-specific trend coefficient, assumed to have mean zero (enforced by eliminating any overall constant in the differences by deducting the sample means). Thus, not only is the meaning of ρ altered, but if ε_{it} did not contain a unit root to start with it will now; in particular, if ε_{it} is not serially correlated to start with, it will follow a first-order moving average process with unit root. The variance -covariance matrix of the new disturbances $\tau_i + \Delta\varepsilon_{it}$ is now block diagonal with blocks:

$$A = \sigma^2 \begin{bmatrix} 1 & a & b & \ldots b \\ a & 1 & a & b.. \\ b & a & 1 & a\ldots \\ \vdots & \vdots & \vdots & \ldots 1 \end{bmatrix} \text{ where } \sigma^2 = \sigma_\tau^2 + \sigma_\varepsilon^2, a = \frac{\sigma_\tau^2 - \sigma_\varepsilon^2}{\sigma^2}, \text{ and } b = \frac{\sigma_\tau^2}{\sigma^2}.$$

Under this assumption, the dynamic relationship to be estimated is stationary and $|\gamma| < 1$. Consider equation (3)[19] with the intercept eliminated, for y_{i0} and the infinite past:

$$(13)\ y_{i0} = \sum_{j=1}^{\infty} \gamma^j \beta x_{i,-j} + \frac{1}{1-\gamma}\mu_i + v_{i0}\ \text{, where } v_{it} = \gamma v_{it-1} + \varepsilon_{it}\ .^{20}$$

If $\beta = 0$, so that the relationship to be estimated is a pure autoregression for each y_{it}, the vector of initial values $y_0 = (y_{10}, \ldots y_{N0})'$ has a joint normal distribution with means 0 and variance-covariance matrix

$[\frac{\sigma_\mu^2}{(1-\gamma)^2} + \sigma_v^2]I_N = \left(\frac{\sigma_\mu^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2} \right)I_N$. The unconditional likelihood is therefore

$$(14)\qquad \log L(\gamma, \rho, \sigma_\mu^2, \sigma_\varepsilon^2 | y_{11}, \ldots, y_{NT}; \ldots; y_{10}, \ldots, y_{N0})$$

$$= -\frac{NT}{2}\log 2\pi - \frac{NT}{2}\log\sigma^2 - \frac{N}{2}\log\xi - \frac{N(T-1)}{2}\log\eta$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}^{\bullet} - \gamma y_{it-1}^{\bullet}\right)^2$$

$$-\frac{N}{2}\log\left(\frac{\sigma_\mu^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2}\right) - \left[\frac{1}{2\left(\frac{\sigma_\mu^2}{(1-\gamma)^2} + \frac{\sigma_\varepsilon^2}{1-\gamma^2}\right)}\right]\sum_{i=1}^{N} y_{i0}^2$$

The characteristic roots of A give the necessary transform and Jacobian. This is taken into account in the formulation of both the conditional and the unconditional likelihood functions. As indicated, however, differencing is unnecessary when the initial values are conditioning.

[19]For a particular time period T and the infinite past

$y_{iT} = \gamma^\infty y_{i-\infty} + \sum_{j=0}^{\infty}\gamma^j \beta x_{i-j} + \frac{1-\gamma^\infty}{1-\gamma}\mu_i + v_{iT}$, where $v_{iT} = \sum_{j=0}^{\infty}\gamma^j \varepsilon_{iT-j}$. Since $1 \geq |\gamma|$ and

$v_{iT} = \sum_{j=0}^{\infty}\gamma^j \varepsilon_{iT-j}$ is the MA form of a first-order autoregression with white noise input, equation (13) follows.

[20] If all variables are expressed as deviations from their overall means, there is no need to include an intercept; if not, μ_i should be replaced by $\alpha + \mu_i$.

This likelihood function can easily be concentrated: To maximize, express σ_μ^2, σ_ε^2, ξ and η in terms of ρ and γ.

For given ρ and γ in the interval [0,1), concentrate the likelihood function with respect to σ^2. It follows that

$$\hat{\sigma}^2(\gamma,\rho) = \frac{RSS^*(\gamma,\rho)}{N(T+1)} \quad where\ RSS^*(\gamma,\rho) = \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}^* - \gamma y_{it-1}^*)^2 + (\sum_{i=1}^{N} y_{i0}^2 \Big/ \left[\frac{\rho}{(1-\gamma)^2} + \frac{1-\rho}{1-\gamma^2}\right].$$

Thus, the concentrated LF is

$$\begin{aligned}
\log L^*(\gamma,\rho) = &-\frac{N(T+1)}{2}\log 2\pi - \frac{N}{2}\log\xi - \frac{N(T-1)}{2}\log\eta \\
&-\frac{N(T-1)}{2}\log\left\{\frac{RSS^*(\gamma,\rho)}{N(T-1)}\right\} - \frac{N}{2}\left\{\frac{\rho}{(1-\gamma)^2} + \frac{1-\rho}{1-\gamma^2}\right\} \\
&-\left(\tfrac{1}{2RSS^*/N(T+1)}\right)\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}^* - \gamma y_{it-1}^*)^2 - \sum_{i=1}^{N} y_{i0}^2 / \{(2/N(T+1))\left[\frac{\rho}{(1-\gamma)^2} + \frac{1-\rho}{1-\gamma^2}\right]RSS^*\}
\end{aligned}$$

Maximizing L* is quite a bit more complicated than the usual minimization of the sum of squares in the penultimate term because RSS*, in that term, depends on $\sum_{i=1}^{N} y_{i0}^2$, as well as on ρ and γ, which enter the final terms as well. When $\beta \neq 0$, things are more complicated still. But more important than finding the maximum of L* is its shape above the γ-ρ plane. It is apparent from the results presented below that there may be significant trade-offs between γ and ρ without large effects on the value of the likelihood.

Various alternative specifications of the likelihood function are considered in the literature are reported and analyzed in Sevestre and Trognon (1996, pp. 136-138).[21] Considerable simplification, however, can be obtained if, following Nerlove (1971), we are willing to assume that x_{it} follows a well-specified common stationary time-series model for all individuals i. The first term in (13) is

[21] One interesting possibility discussed by Sevestre and Trognon(1996, p. 136-138) is to choose y_{i0} a linear function of some *observed* individual-specific time-invariant exogenous variables and a disturbance which is decomposed as the sum of the individual-specific disturbances μ_i and a remainder. The first-order equations for maximizing the likelihood then take on a simple recursive form when $\beta = 0$, and permit other simplification when $\beta \neq 0$. But if we knew some individual-specific time-invariant observed variables influenced behavior why not incorporate them directly in (2), the equation to be estimated?

$\varphi_{i0} = \beta \sum_{j=0}^{\infty} \gamma^j x_{i,-j}$. Hence, for any stationary processes x_{it} ,which may be serially correlated.

$$\frac{\varphi_{it}}{\beta} = \gamma \frac{\varphi_{it-1}}{\beta} + x_{it}$$

with variances

$$\sigma^2_{\varphi_i} = \frac{\beta^2 \sigma^2_{x_i}}{1-\gamma^2} \quad . \tag{15}$$

If we suppose that the variance of the x_{it} is the same for all i, then the random variable

$$\phi_{it} = \sum_{j=0}^{\infty} \gamma^j \beta x_{it-j}$$

has a well defined variance which is the same for all i and a function of β, γ, and σ^2_x . This then enters the final term in the unconditional likelihood (14), which now becomes:

(16) $\log L(\beta, \gamma, \sigma^2_\mu, \sigma^2_\varepsilon | y_{11}, \ldots, y_{NT}; x_{11}, \ldots x_{NT}; y_{10}, \ldots, y_{N0})$

$$= -\frac{N(T+1)}{2}\log 2\pi - \frac{NT}{2}\log \sigma^2 - \frac{N}{2}\log \xi - \frac{N(T-1)}{2}\log \eta$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y^*_{it} - \beta x^*_{it} - \gamma y^*_{it-1}\right)^2$$

$$-\frac{N}{2}\log\left(\frac{\beta^2\sigma^2_x}{1-\gamma^2} + \frac{\sigma^2_\mu}{(1-\gamma)^2} + \frac{\sigma^2_\varepsilon}{1-\gamma^2}\right) - \left[\frac{1}{2\left(\frac{\beta^2\sigma^2_x}{1-\gamma^2} + \frac{\sigma^2_\mu}{(1-\gamma)^2} + \frac{\sigma^2_\varepsilon}{1-\gamma^2}\right)}\right]\sum_{i=1}^{N} y^2_{i0} .$$

Concentrating the likelihood function to permit a one- or two-dimensional grid search is no longer possible. Nor is it possible to graph the likelihood surface with respect to variations in all of the parameters. Although "slicing" the likelihood function along any hyperplane in the parameter space can reveal the trade-offs between any pair of parameters. If gradient or search procedures yield and interior maximum, the ML estimates obtained are consistent as long as the random variables $\phi_{it} = \sum_{j=0}^{\infty} \gamma^j \beta x_{i,t-j}$ have well-defined variances and covariance's, which they will if the x_{it} are generated by stationary a process. It doesn't really matter what this process is as long as it is stationary. Besides, since the x_{it} are assumed to be exogenous, we really have no basis on which to model their determination and are likely to misspecify this part of the model. In this sense we ought to prefer this kind of "almost full-information" maximum likelihood. Still we have to assume something about the variance of the x process in order to proceed. I suggest estimating σ_x^2 from the sample data.

To generalize these results to the case in which their are several explanatory variables in addition to the lagged value of the dependent variable, assume that X_{it} follows a stationary VAR process and replace βx^*_{it} by $X^*_{it}\beta$ and $\beta^2 \sigma_x^2$ by $\beta' \Sigma_{XX} \beta$ in the above formula.

4. Empirical Evidence on the Comparative Performance of Different Panel Data Methods

In order to examine the effects of the econometric methods employed on the finding of growth rate convergence or the lack of it, I initially used 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 at ftp://nber.harvard.edu/pub/. This is the same data set which has been used in dozens of previous studies. Following Islam (1995), s and n were computed as quinquennial means over the preceding 5-year span for the 5 years 1965, 1970, 1975, 1980, 1985; y was taken as the value reported in that year and in 1960 for the lagged value applicable to 1965. The results of the six methods applied to these data are reported in Table 1 for the usual undifferenced model. Table 2 reports the results for the country-specific trends model which requires differencing to reduce the process to stationarity. In this case, the conditional and unconditional likelihood functions are defined for the first differences of the original data.

I have listed the regression methods in the order in which the corresponding estimates of γ appear in the inequality of Sevestre and Trognon (equation 9 above). These estimates are followed by the maximum likelihood estimates conditional on the initial values y_{i0} or Δy_{i1} and the ML estimates unconditional on the initial values, assuming stationarity of both the processes generating the exogenous variable and real GDP per capita. In a substantive study of growth rate convergence, it would clearly be important to include additional explanatory variables such as, for example, the stock of human capital, also available at the NBER internet site, infrastructure investment, and so forth. However, my focus here is on properties of alternative estimators and for this purpose, omission of relevant variables simply increases the unexplained disturbance variance and thus heightens the contrast among alternative estimators.

Turning now to the regression estimates presented in Table 1 (Tables 1 and 2 appear at the end of the paper): Consider the first four methods. The estimates of γ for the 94-country sample range from a low of 0.72 (fixed-effects regression) to a high of 0.98 (country means regression) with pooled OLS and FGLS falling in between.[22] For the OECD countries the range is 0.76 to 0.93. The implied speed of convergence thus ranges from 90% in 35 years to 90% in 570 years. None could be characterized as evidence of reasonably rapid convergence. All of the estimates of γ satisfy the Trognon-Sevestre inequality, although the regressions contain an exogenous explanatory variable in contrast to the case

[22] In the following Table, I present the three estimates of ρ discussed above as possible candidates for the transformation involved in FGLS for the 94-country sample and the model in levels. It is argued above that the Greene-Judge estimate is sharply biased downwards and prone to be negative; similarly, the argument Nickell gives with reference to the downward bias in the coefficient of the lagged dependent variable in a fixed-effects regression suggests that the other coefficients will be biased upwards, including the variance of the estimated fixed effects, Coupled with a downward bias in the estimate of the residual variance in the fixed-effects regression, this provides an explanation of the extremely high estimates obtained by the Nerlove (1971) method. It is interesting to note that the Balestra-Nerlove estimate, while substantially higher than the GJ estimate (it can never be negative) is, nonetheless, not too far out of line with the estimates of ρ obtained from the conditional likelihood function for the OECD countries and for both the conditional and unconditional likelihood functions for the 94-country sample.

Method	94-countries	22-countries
Balestra-Nerlove(1966)	0.2678	0.4027
Nerlove(1971)	0.7790	0.7038
G-J(1983/88)	0.0983	0.0804
Conditional ML	0.1133	0.4796
Unconditional ML	0.1288	0.7700

considered by Sevestre and Trognon. Pooled OLS and FGLS also stand in the order predicted by the Sevestre-Trognon results. While it is tempting to infer that FGLS provides a tighter upper bound to the true value of γ than the pooled OLS regression estimate. the temptation should be resisted. The FGLS estimates are doubly inconsistent: they are based on an inconsistent estimate of ρ reflecting the inconsistency of the estimates of the residual variance and the fixed effects depending on which regressions they are derived from. Not only is the estimated value of β sensitive to the method of estimation but the estimate of α, the elasticity of output with respect to capital stock in the production function is extremely so, reflecting the dependence of the estimated value on the coefficient of the lagged dependent variable, γ. This parameter should estimate approximately (1 - the share of labor in the real GDP). It is clear that all of the estimates of α are wide of the mark. If therefore one were to infer policy implications from this parameter, it could be seriously misleading.

The most interesting estimates are those for conditional and unconditional maximum likelihood presented as methods 5 and 6 in Table 1 and 2 for the level model and the first-difference model, respectively . For the model in levels and the 22-country OECD sample, these estimates differ quite a bit from one another, although unconditional ML is not far from the fixed-effects OLS regression, while conditional ML yields results close to FGLS using the Balestra-Nerlove (1966) first-round estimate of ρ. For the 94-country sample, the conditional and the unconditional ML estimates differ little from one another. They are close to the pooled OLS regression estimates (a consequence of the fact that the estimated value of ρ is small although significantly different from zero), but are both quite different than any of the inconsistent regression estimates. The estimates of β are quite insensitive to the method used, presumably because the estimates of γ are not very different; consequently the implied estimates of α are similar, albeit different for the two samples. While the results for the first-difference model are quite different from those for the levels model, the same pattern of relation between conditional and unconditional estimates emerges.

To understand better the relation between the conditional and the unconditional ML estimates, consider the log of the ratio of the unconditional to the conditional likelihood, i.e. the marginal density of y_{i0}:

log {unconditional / conditional likelihood } =

$$-\frac{N}{2}\log 2\pi - \frac{N}{2}\log\left(\frac{\beta^2\sigma_x^2}{1-\gamma^2} + \frac{\rho\sigma^2}{(1-\gamma)^2} + \frac{(1-\rho)\sigma^2}{1-\gamma^2}\right) - \left[\frac{1}{2\left(\frac{\beta^2\sigma_x^2}{1-\gamma^2} + \frac{\rho\sigma^2}{(1-\gamma)^2} + \frac{(1-\rho)\sigma^2}{1-\gamma^2}\right)}\right]\sum_{i=1}^{N} y_{i0}^2.$$

Let the sample variance of y_{i0} be vary_0 and let

$$\varphi^2 = \left(\frac{\beta^2\sigma_x^2}{1-\gamma^2} + \frac{\rho\sigma^2}{(1-\gamma)^2} + \frac{(1-\rho)\sigma^2}{1-\gamma^2}\right).$$

Then

log {unconditional / conditional likelihood } = $f(\varphi^2)$ =

$$-\frac{N}{2}\log 2\pi - \frac{N}{2}\log(\varphi^2) - \left[\frac{N\,\mathrm{var}\,y_0}{2\varphi^2}\right].$$

The maxima of the two likelihood functions will occur at about the same values of the parameters on which φ depends when $df(\varphi^2)$ is close to zero, which occurs at φ^2 = vary_0. At the unconditional ML estimates for the levels model, for example, for the 94-country sample, at $\varphi^2 = 0.91$ and vary_0 =0.80, while, for the 22-country sample, φ^2 =0.25 and vary_0 = 0.26.

Table 2 presents parallel results for the first-difference model. Once again the first four estimates of γ fall in the order to be expected from the Trognon-Sevestre inequality, although they are all lower than for the levels model, in the first three cases much lower, implying much more rapid convergence to equilibrium. The estimates of all the parameters are much different for the 22-country sample and quite variable. Perhaps the most interesting findings, however, are for the conditional and unconditional ML estimates. Once the estimates of ρ for the 94-country sample are quite close to one another and those for the 22-country sample far apart, but now there is a remarkable reversal of the magnitudes of ρ and γ as between the 94-country sample and the 22-country sample: For the former ρ is about one-half of the estimated value for γ, but in the case of the 22-country sample ρ is only a small fraction of the estimated value of γ.

Further insight into the nature of the conditional and unconditional likelihood functions for the two samples can be obtained graphically. Having eliminated the constant term by taking deviations from the overall means of all variables, we are left with four parameters: ρ, γ, β and σ^2. Figure 1 plots the unconditional likelihood function for the 94-country sample, levels model. Figure 2 plots the likelihood function for the 22-country sample, levels model. Likelihood functions are plotted in Figures 3 and 4 for the first-difference model, respectively for the 94- and 22-country samples. (Figures appear at the end of the paper.) I have plotted both three-dimensional likelihood surfaces for pairs of variables and two-dimensional contours. "Slices" are taken at the likelihood maximizing values for the parameters not plotted. These plots clearly reveal the implications of the data for the "interactions" between pairs of parameters. Although there are $\binom{5}{2} = 10$ possible pairs to consider, I focus on the crucial pairs: ρ vs. γ and β vs. σ^2.

Although the likelihood reaches a unique maximum in every case, which is quite well-defined, it is clear that there are significant trade-offs between each pair of parameters. In the case of the 22 OECD countries, the unconditional ML estimates are precisely determined. As suggested above, this is because for small N, the weight of the initial observations and the parameters determining them is more substantial than for large cross-sectional samples. As indicated above, the likelihood function is sufficient for the parameters of the model and provides useful insight into what the data tell us about these parameters quite apart from the values that maximize it.

5. Conclusions

The principal conclusion that can be drawn from this analysis is that, in panel data econometrics, method matters -- a lot. Although, using a highly simplified Solow/Swan model without human capital stocks or infrastructure, I have found estimates of the adjustment parameter significantly different than one in every case, indicating convergence. All of the estimates for the model in levels, however, are so close to one, always greater than 0.7, that convergence to within 90% of equilibrium in less than one generation is effectively ruled out. This can hardly be called "convergence" in any relevant sense. Moreover, the estimates range from 0.72 to 0.98, suggesting a convergence range of from 33 to over 500 years, with most clustering around 0.8, underscoring the importance of choice of econometric method. When the model is estimated in first-difference form, the estimates of γ are much lower, indicating rapid convergence in the case of the 94-country sample. The method of choice, unconditional ML, yields well-defined and reasonable estimates in every case. Much of the variation in estimates of the speed of convergence appears to be due to trade-offs between the crucial parameter ρ, which measures the importance of unobserved cross-sectional variation relative to total residual variation, and γ, which measures the speed of adjustment. For this reason, it is especially important to introduce other relevant variables, such as infrastructure investment and human capital stock, in order to reduce the importance of *unobserved* cross-sectional variation..

A second important finding is that the Sevestre-Trognon inequality, proved only for the case $\beta = 0$, and then only asymptotically, holds for all the examples presented. Indeed, fixed-effects OLS always yields estimates of the adjustment parameter at the extreme low end of the range of estimates obtained. The "bias" of fixed-effects models in the estimation of dynamic panel models is apparent. In this context, the use of such methods biases a test for convergence, or more appropriately rapid convergence, towards finding it. Fixed-effects models, however, are widely used, in part because they are the basis for two-round FGLS estimators, and because computer packages for panel data analysis incorporate an extremely misguided suggestion for estimating ρ, guaranteed to yield extremely low or even negative values of this parameter. These packages should be avoided, and, if they are used and do yield a negative estimate, it should not be concluded that the model is misspecified or that fixed-effects are a preferable alternative. Fixed-effects OLS remains badly biased in a dynamic context irrespective of whether the packaged routines fail.

I do find, however, that FGLS, using the Balestra-Nerlove (1966) estimate of ρ, which can never be negative, always lie between the fixed-effects OLS estimates and the pooled OLS estimates, which are known to yield upwardly biased estimates of γ. It is not appropriate to conclude that these FGLS estimates, however, represent a tighter upper bound to the true value of γ, since they are doubly inconsistent estimates and may lie below the true value. This is underscored by the finding that both conditional and unconditional ML yield different estimates of ρ and γ, sometimes higher and sometimes lower than FGLS. The interaction between ρ and γ is crucial in this regard.

Finally, maximum likelihood, unconditional on the initial observations, assuming them to be stationary and generated by the same dynamic process we are trying to estimate and assuming the exogenous variables also to be stationary, is feasible and indeed a viable alternative to conventional regression methods or conditional ML. Use of such methods will, however, generally involve removal of the overall means of all variables prior to analysis and omission of a constant term and may also involve differencing to remove deterministic or stochastic trends. formulation of the unconditional likelihood function is somewhat more complicated in the case of differenced variables but, as demonstrated, quite feasible nonetheless. The unconditional and simpler conditional ML method may yield similar results under certain circumstances, but cannot generally be expected to do so.

REFERENCES

Arellano, M., and S. Bond, 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies, 58*: 277-297.

Balestra, P., and M. Nerlove, 1966. "Pooling Cross-Section and Time-Series Data in the Estimation of a Dynamic Economic Model: The Demand for Natural Gas." *Econometrica, 34*:585-612.

Barnard, G. A., "Statistical Inference," *Jour. Royal Statistical Society, Ser. B 11:* 115-149 (1949).

Barnard, G. A., "The Theory of Information," *Jour. Royal Statistical Society, Ser. B 13:* 46-64 (1951).

Barnard, G. A., "The Use of the Likelihood Function in Statistical Practice," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1:* 27-40 (1966).

Barnard, G. A., "The Bayesian Controversy in Statistical Inference," *Journal of the Institute of Actuaries, 93:* 229-69 (1967).

Barnard, G. A., Jenkins, G. M and Winsten, C. B., 1962. "Likelihood Inference and Time Series," (with discussion), *Journal of the Royal Statistical Society, Series A, 125:*321-375).

Bernard, A. B., and S. N. Durlauf, 1996. "Interpreting Tests of the Convergence Hypothesis." *Journal of Econometrics, 71*:161-173.

Bernard, A. B., and S. N. Durlauf, 1995. "Convergence in International Output." *Journal of Applied Econometrics, 10*:97-108.

Barro, R. J., 1991. "Economic growth in a Cross-Section of Countries." *Quarterly Journal of Economics, 106*:407-443.

Barro, R. J., and Jong-Wha Lee, 1993. "International Comparisons of Educational Attainment. " *NBER Working Paper No. 4349.*

Barro, R. J., and X. Sala-I-Martin, 1995. *Economic Growth*. New York: McGraw-Hill.

Baumol, W., 1986. "Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show." *American Economic Review, 76:* 1072-1085.

Binder, M., and M. H. Pesaran, 1996. "Stochastic Growth." *Working Paper No. 19-18, Department of Economics, University of Maryland.*

Birnbaum, A., 1962 " On the Foundations of Statistical Inference." (with discussion) *Journal of the American Statistical Association, 57*:269-306.

Caselli, F., G. Esquivel and F. Lefort, 1996. "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics," *Journal of Economic Growth, 1*: 363-389.

Crépon, B., and J. Mairesse, 1996. "The Chamberlain Approach." Pp. 323-391 in L. Mátyás and P. Sevestre, *op.cit.*

Chamberlain, G., 1984 . "Panel Data." Pp. 1247-1313 in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics, II.* Amsterdam: Elsevier.

Diebold, F. X., and M. Nerlove, 1990. "Unit Roots in Economic Time Series: A Selective Survey." Pp. 3-69 in T.B. Fomby and G.F. Rhodes, eds., *Advances in Econometrics, Vol. 8: Co-Integration, Spurious Regressions, and Unit Roots.* Greenwich, CT: JAI Press.

de la Fuente, A., 1997. "The Empirics of Growth and Convergence: A Selective Review," *Journal of Economic Dynamics and Control, 21*: 23-73.

Edwards, A. W. F., *Likelihood.* Cambridge: University Press, 1972.

Fisher, R. A., "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society of London, Series A, 222:*309-368 (1922).

Fisher, R. A., "Theory of Statistical Estimation," *Proceedings of the Cambridge Philosophical Society*, *22*:700-725 (1925).

Greene, W., 1993. *Econometric Analysis,* Second Edition. New York: Macmillan Publishing.

Holtz-Eakin, D, W. Newey, and H. Rosen, 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica, 56:* 1371-1395.

Islam, N., 1995a. "Growth Empirics: A Panel Data Approach." *Quarterly Journal of Economics, 110*:1127-1170.

Islam, N., 1995b. "Convergence: Variations in Concept and Results." Unpublished. International Institute for Advanced Studies, Cambridge, MA.

Knight, M., N. Loayza and D. Villanueva, 1993. "Testing the Neoclassical Growth Model," *IMF Staff Papers, 40*: 512-41.

Judge, G., R. C. Hill, W. Griffiths, H. Lütkepohl, and T-C. Lee, 1988. *Introduction to the Theory and Practice of Econometrics,* Second Edition. New York: John Wiley & Sons.

Kiviet, Jan F., "On the Bias, Inconsistency and Efficiency of Various Estimators in Dynamic Panel Models," *Journal of Econometrics, 68*:53-78 (1995).

Lee, K., M. H. Pesaran, and R. Smith, 1996. "Growth and Convergence: A Multicountry Empirical Analysis of the Solow Growth Model," unpublished.

Lindsey, J. K., 1996. *Parametric Statistical Inference.* Oxford: Clarendon Press.

Loayza, N., 1994. "A Test of the International Convergence Hypothesis Using Panel Data." *Policy Research Working Paper No. 1333.* The World Bank.

Mankiw, N. G., D. Romer and D. N. Weil, 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics, 108*:407-437

Mátyás, L. and P. Sevestre, 1996. *The Econometrics of Panel Data: Handbook of Theory and Applications.* 2nd. ed.1996. Boston: Kluwer Academic Publishers, 1992.

Maddala, G. S., 1971. "The Use of Variance Components Models in Pooling Cross-Section and Time Series Data." *Econometrica, 39*:341-358.

Mundlak, Y., 1978. "On the Pooling of Cross-Section and Time-Series Data," *Econometrica, 46*: 69-86.

Nerlove, M., 1971. "Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross-Sections." *Econometrica, 39*:359-382.

Nerlove, M., 1996. "Growth Rate Convergence, Fact or Artifact? An Essay in Panel Data Econometrics." Paper presented to the Sixth Conference on Panel Data Econometrics, Amsterdam, 28-29 June 1996.

Nerlove, M., 1997. "Likelihood Inference for Dynamic Panel Models." In preparation.

Nerlove, M., and P. Balestra, 1996. "Formulation and Estimation of Econometric Models for Panel Data." Pp. 3-22 in L. Mátyás and P. Sevestre, *op.cit.*

Nerlove, M., and L. K. Raut, 1997. "Growth Models with Endogenous Population: A General Framework," Pp. 1117-1174 in *The Handbook of Family and Population Economics*, ed. by M.R. Rosenzweig and O. Stark. New York: Elsevier Scientific Publishers.

Nickell, S., 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica, 49*:1417-1426.

Quah, D., 1996. "Empirics for Economic Growth and Convergence." Forthcoming, *European Economic Review.*

Sevestre, Patrick, and Alain Trognon, "Propriétés de grands échantillons d'une classe d'estimateurs des modéles autoregréssives à erreurs compossées," *Annales de l'INSEE, 50*:25-49 (1983).

Sevestre, Patrick, and Alain Trognon, 1996. "Linear Dynamic Models," pp. 120-144 in Mátyás and Sevestre, *op. cit.*

Solow, R.M., 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics, 70*:65-94.

Swan, T.W., 1956. "Economic Growth and Capital Accumulation." *Economic Record, 32*:324-361.

Trognon, Alain. 1978. "Miscellaneous Asymptotic Properties of Ordinary Least Squares and Maximum Likelihood Methods in Dynamic Error Components Models," *Annales de l'INSEE,30-31:*631-657.

TABLE 1: PARAMETER ESTIMATES FOR THE MODEL IN LEVELS, ALTERNATIVE ECONOMETRIC ANALYSES

METHOD OF ANALYSYS	94-COUNTRY SAMPLE	22-COUNTRY SAMPLE
1. Fixed Effects OLS		
γ	0.7204 (0.0211)	0.7645 (0.0166)
β	0.1656 (0.0172)	0.1634 (0.0510)
Implied α	0.3719 (0.0278)	0.4096 (0.0783)
Residual Variance	0.0113	0.0020
2. Feasible GLS		
Estimate of ρ used*	0.2675	0.4027
γ	0.9130 (0.0119)	0.8282 (0.0156)
β	0.1520 (0.0135)	0.1913 (0.0422)
Implied α	0.6362 (0.0247)	0.5269 (0.0579)
Residual Variance	0.0213	0.0047
3. Pooled OLS		
γ	0.9487 (0.0090)	0.8857 (0.0125)
β	0.1244 (0.0108)	0.1764 (0.0308)
Implied α	0.7080 (0.0271)	0.6067 (0.0452)
Residual Variance	0.0193	0.0041
4. Country Means OLS		
γ	0.9817 (0.0112)	0.9320 (0.0148)
β	0.0919 (0.0138)	0.1493 (0.0343)
Implied α	0.8339 (0.0704)	0.6870 (0.0593)
Residual Variance	0.0047	0.0580
5. Conditional ML		
ρ	0.1133 (0.0497)	0.4796 (0.1584)
γ	0.9339 (0.0122)	0.8189 (0.0245)
β	0.1370 (0.0131)	0.1908 (0.0438)
Implied α	0.6744 (0.0289)	0.5131 (0.0664)
Residual Variance	0.0194 (0.0013)	0.0052 (0.0012)
6. Unconditional ML		
Estimates of σ_x^2 used	0.0826	0.0069
ρ	0.1288 (0.0456)	0.7700 (0.0731)
γ	0.9385 (0.0105)	0.8085 (0.0228)
β	0.1334 (0.0124)	0.1815 (0.0521)
Implied α	0.6846 (0.0277)	0.4865 (0.0791)
Residual Variance	0.0197 (0.0013)	0.0113 (0.0028)

Figures in parentheses are standard errors.
* Estimated by the method suggested in Balestra and Nerlove (1966).

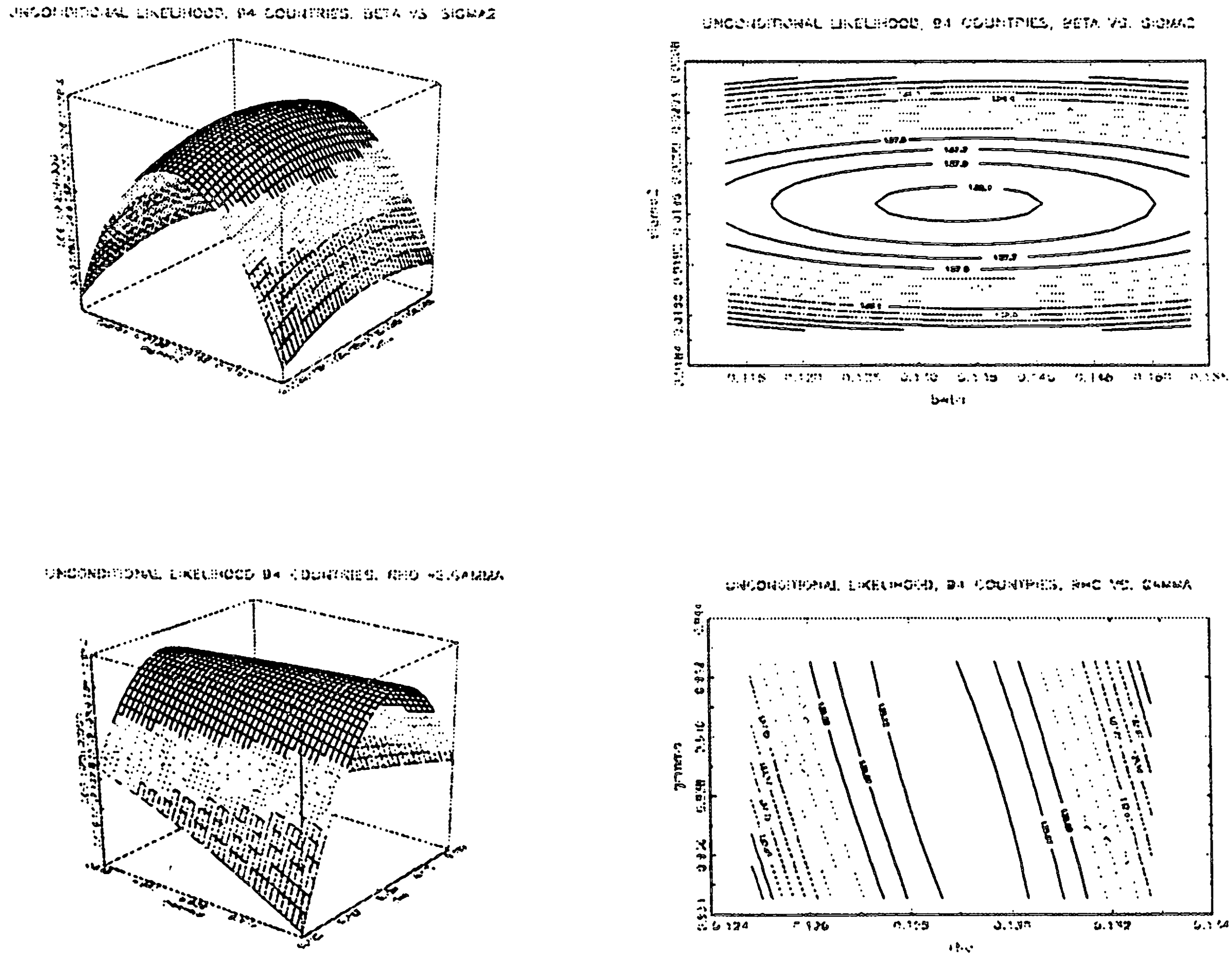
TABLE 2: PARAMETER ESTIMATES FOR THE MODEL IN FIRST DIFFERENCES, ALTERNATIVE ECONOMETRIC ANALYSES

METHOD OF ANALYSYS	94-COUNTRY SAMPLE	22-COUNTRY SAMPLE
1. Fixed Effects OLS		
γ	0.4007 (0.0375)	0.4544 (0.0611)
β	0.1199 (0.0187)	- 0.0126 (0.0637)
Implied α	0.1667 (0.0246)	- 0.0237 (0.1209)
Residual Variance	0.0077	0.0014
2. Feasible GLS		
Estimate of ρ used*	0.4866	0.3628
γ	0.4227 (0.0406)	0.5833 (0.0531)
β	0.1520 (0.0135)	0.1913 (0.0422)
Implied α	0.1864 (0.0259)	0.1322 (0.1218)
Residual Variance	0.0213	0.0047
3. Pooled OLS		
γ	0.7031 (0.0328)	0.6237 (0.0453)
β	0.1632 (0.0195)	0.0845 (0.0586)
Implied α	0.3548 (0.0373)	0.1834 (0.1121)
Residual Variance	0.0141	0.0022
4. Country Means OLS		
γ	0.9178 (0.0471)	0.7215 (0.0572)
β	0.1719 (0.0339)	0.1174 (0.0978)
Implied α	0.6763 (0.1263)	0.2965 (0.1873)
Residual Variance	0.0041	0.0005
5. Conditional ML		
ρ	0.2267 (0.0664)	0.0126 (0.0405)
γ	0.4540 (0.0651)	0.6187 (0.0490)
β	0.1368 (0.0208)	0.0815 (0.0601)
Implied α	0.2004 (0.0358)	0.1762 (0.1159)
Residual Variance	0.0122 (0.0009)	0.0021 (0.0003)
6. Unconditional ML		
Estimate of σ_x^2 used	0.0597	0.0058
ρ	0.2335 (0.0632)	0.0936 (0.0696)
γ	0.4364 (0.0578)	0.7254 (0.0512)
β	0.1340 (0.0201)	0.1478 (0.0727)
Implied α	0.1921 (0.0317)	0.3500 (0.1326)
Residual Variance	0.0120 (0.0008)	0.0027 (0.0004)

Figures in parentheses are standard errors.

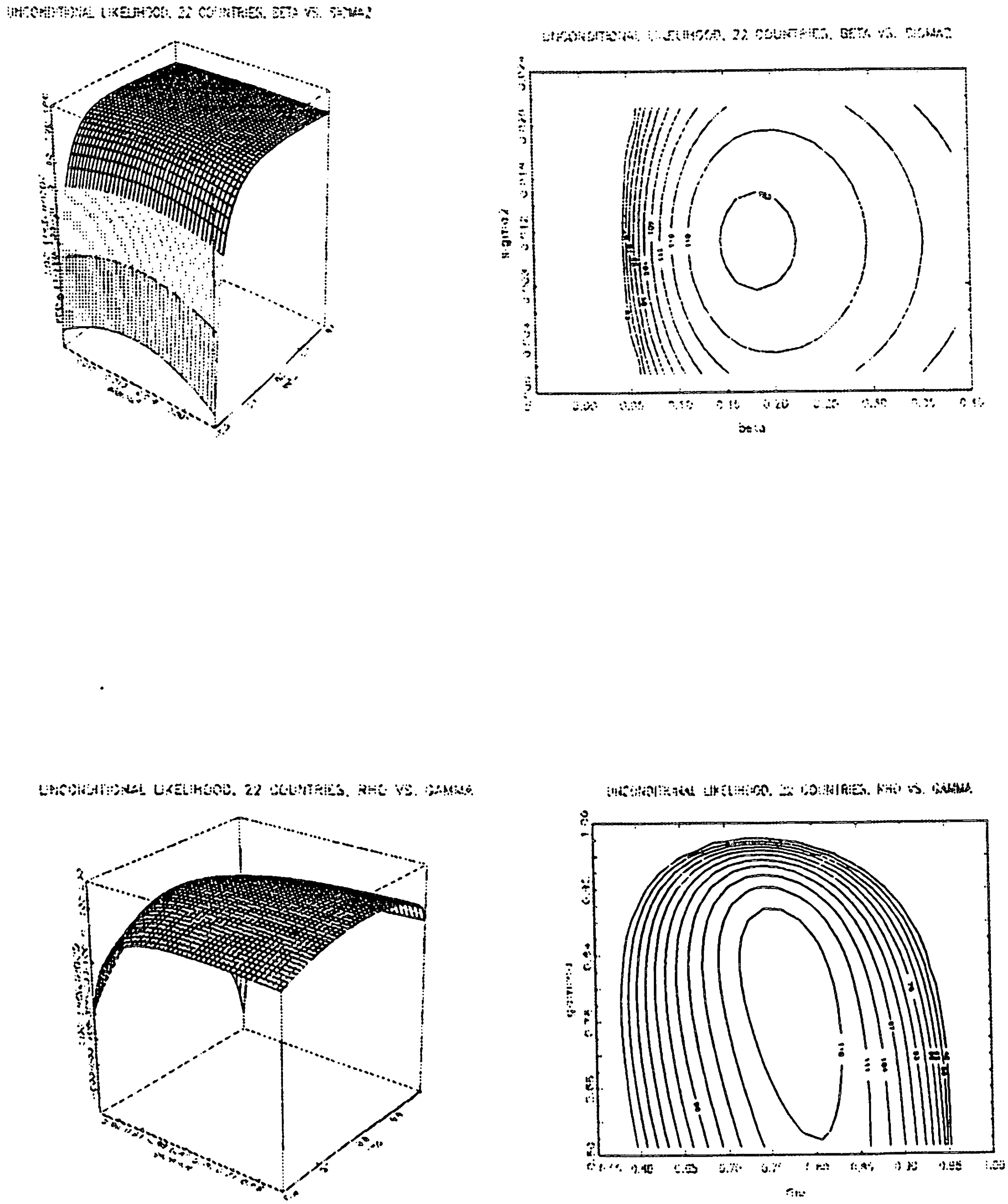
* Estimated by the method suggested in Balestra and Nerlove (1966).

FIGURE 1: UNCONDITIONAL LIKELIHOOD, 94-COUNTRY SAMPLE, LEVELS MODEL



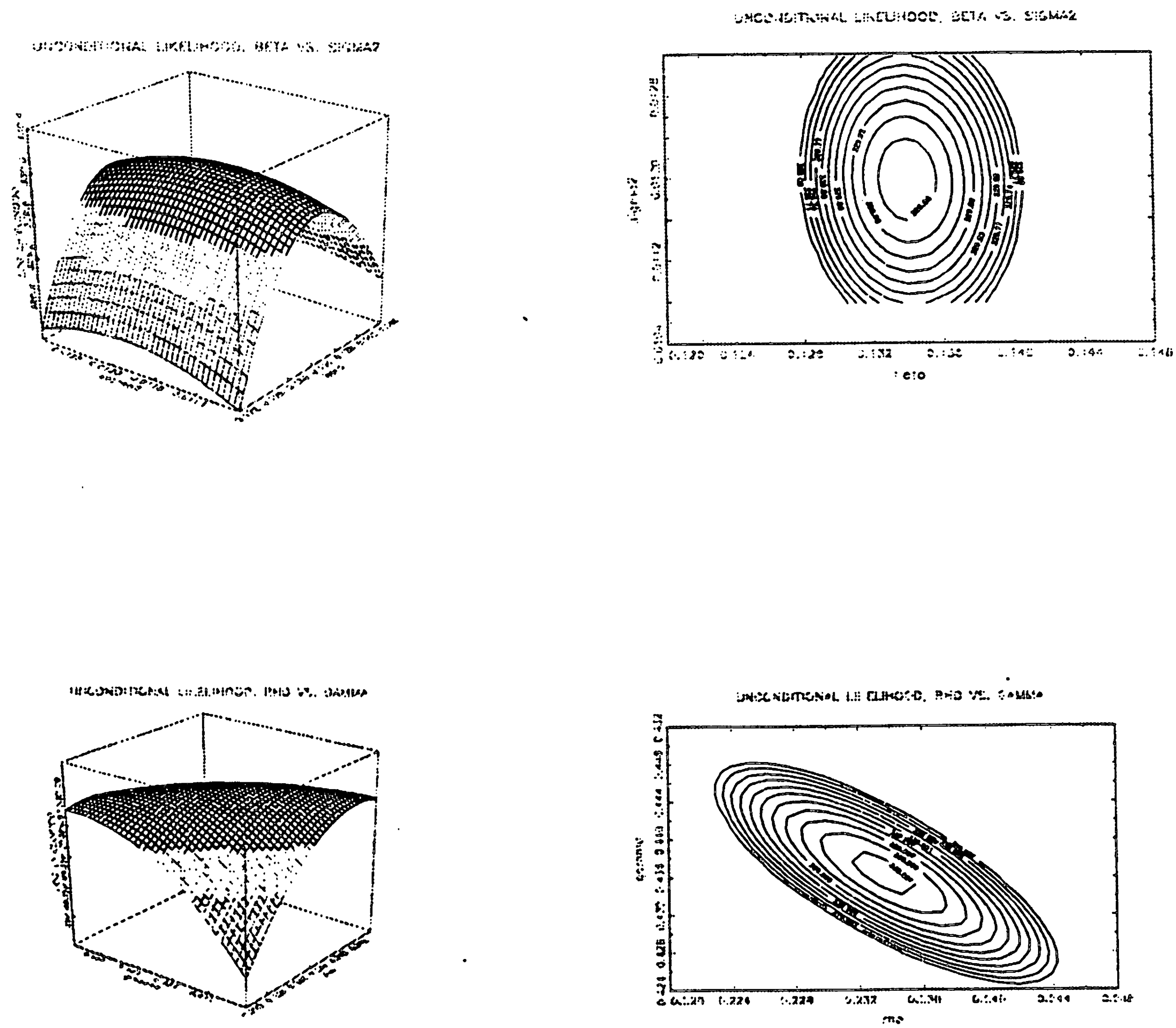
Estimated values: $\rho = 0.130$ (0.0417); $\gamma = 0.938$ (0.0105);
$\beta = 0.134$ (0.0124); $\sigma^2 = 0.0197$ (0.0013).

FIGURE 2: UNCONDITIONAL LIKELIHOOD, 22-COUNTRY SAMPLE, LEVELS MODEL



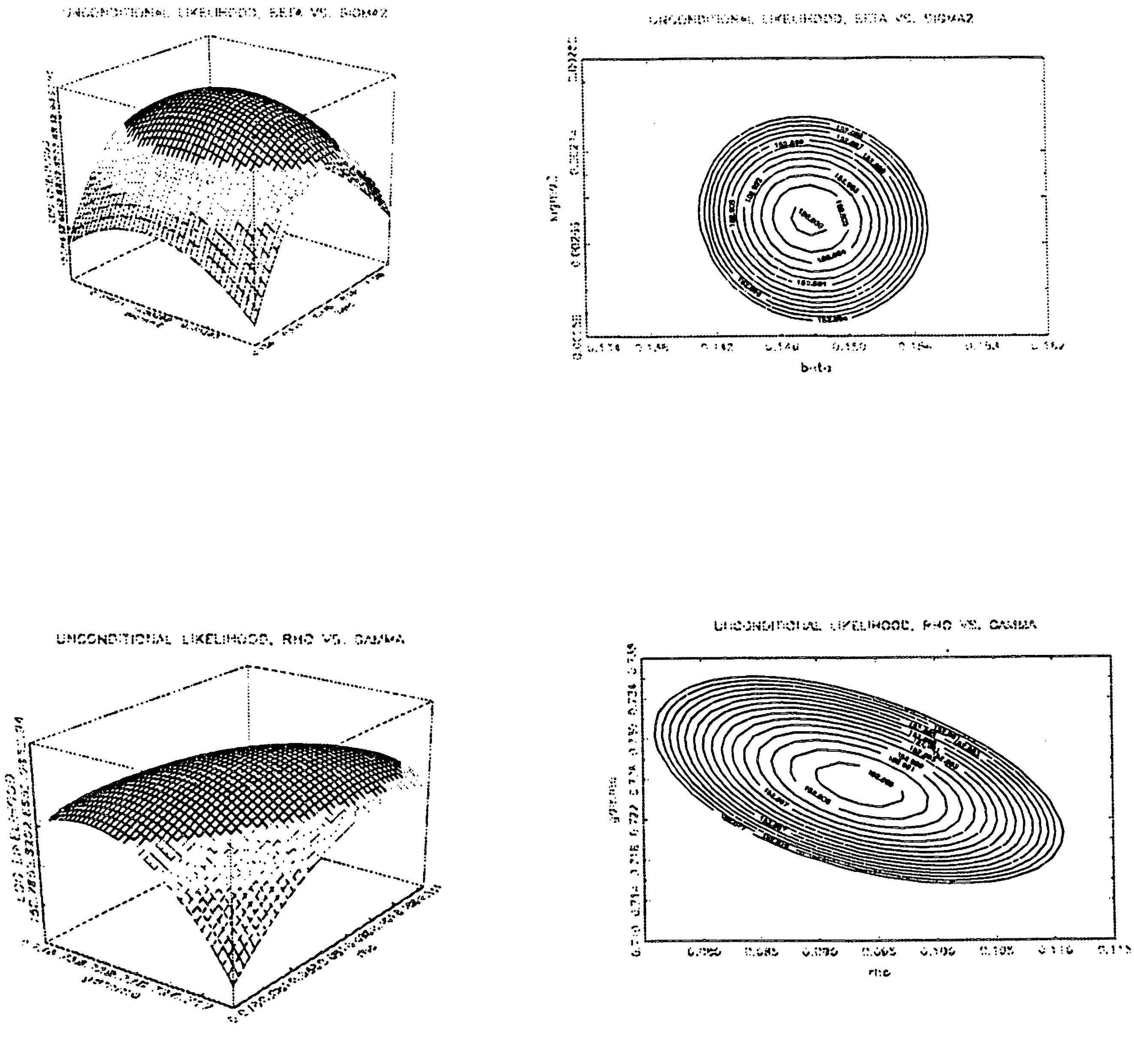
Estimated values: ρ = 0.770 (0.0740); γ = 0.808 (0.0230);
β = 0.180 (0.0520); σ^2 = 0.0113 (0.0029).

FIGURE 3: UNCONDITIONAL LIKELIHOOD, 94-COUNTRY SAMPLE, FIRST-DIFFERENCE MODEL



Estimated values: $\rho = 0.234$ (0.0632); $\gamma = 0.436$ (0.0578);
$\beta = 0.134$ (0.0201); $\sigma^2 = 0.0120$ (0.0008).

FIGURE 4: UNCONDITIONAL LIKELIHOOD, 22-COUNTRY SAMPLE, FIRST-DIFFERENCE MODEL



Estimated values: $\rho = 0.094$ (0.0696); $\gamma = 0.725$ (0.0512);
$\beta = 0.148$ (0.0727); $\sigma^2 = 0.0027$ (0.0004).