GROWTH RATE CONVERGENCE, FACT OR ARTIFACT?

An Essay on the Use and Misuse of Panel Data Econometrics

by
Marc Nerlove

WP 98-21

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

Growth Rate Convergence, Fact or Artifact?

An Essay on the Use and Misuse of Panel Data Econometrics

Marc Nerlove

The 1998 Havlicek Memorial Lecture in Applied Economics

T·H·E OHIO STATE UNIVERSITY

Department of Agricultural, Environmental, and Development Economics

The Ohio State University

HAVLICEK LECTURERS

1994	**Oral Capps, Jr.** Professor, Department of Agricultural Economics Texas A&M University *"Linking Demand Systems to Macroeconomic Models"*
1995	**G. Edward Schuh** Dean and Professor Humphrey Institute of Public Affairs University of Minnesota *"Econometrics and Economic Policy"*
1996	**George G. Judge** Professor, Department of Agricultural and Resource Economics University of California-Berkeley *"Recovering Information from Underdetermined Problems and Limited Economic Data"*
1997	**William G. Tomek** Professor, Department of Agricultural, Resource and Managerial Economics Cornell University *"Structural Econometric Models: Past and Future"* *(with Special Reference to Agricultural Economics Applications)*
1998	**Marc Nerlove** Professor, Department of Agricultural and Resource Economics University of Maryland *"Growth Rate Convergence, Fact or Artifact?"* *An Essay on the Use and Misuse of Panel Data Econometrics*

JOSEPH HAVLICEK, JR.
1934-1993

Joseph Havlicek, Jr. was a teacher, researcher, administrator, and leader in agricultural economics. His enthusiasm and optimism were hallmarks of his character and career. He was intensely interested in his work, excited by his research, and energetic in his teaching. He had the ability to motivate and inspire others -- students, colleagues, administrators, and leaders in the profession. Both his classroom and office represented opportunities for clarification and motivation to all who appeared at his door.

Havlicek was a pioneer in much of his research. He was an early developer of nonlinear econometric systems for modeling interregional competition, among the first to research the economics of waste management, and an innovator in the analysis of consumer and market behavior using systems of equations and in the assessment of investment in agricultural research and extension.

Havlicek grew up on a farm in Mercer County, Ohio, fluent in the Czech language of his parents, who had immigrated from the Ukraine area of Russia just prior to World War I. After earning the Bachelor of Science (1955) and Master of Science (1956) degrees in Agricultural Economics from The Ohio State University, he earned a Ph.D. degree (1959) in both Agricultural Economics and Statistics from North Carolina State University.

North Carolina State had one of the early graduate programs in econometric analysis and Havlicek put this training into practice during his first position (1959-61) in the Agricultural Economics Branch of The Tennessee Valley Authority. His research was in agronomic-economic studies, response surface analysis, and resource economics, including watershed analyses.

By 1961, Purdue University was hiring faculty to establish a graduate program in econometrics and Havlicek joined the team to develop this program as applied to agricultural economics. At Purdue, his research focused on interregional competition, demand and price analysis, and the economics of waste management. He spent 13 years at Purdue before moving to Virginia Polytechnic Institute's agricultural economics department with a joint appointment in the statistics department. At VPI, his research focused on supply and demand analyses of selected products and production inputs, consumption analysis, environmental economics, and returns to investment in agricultural research.

Both at Purdue and VPI, Havlicek taught graduate and undergraduate courses in statistics, econometrics, matrix algebra, demand and price analysis, and marketing. His lecture notes for matrix algebra and for econometrics were reproduced and used not only by his graduate classes, but were distributed upon request to many researchers throughout the country.

Between 1961 and 1982, Havlicek was the major professor for 18 Ph.D. dissertations and 10 Masters theses. Two of these Ph.D. dissertations and four of the M.S. theses won AAEA awards. Throughout his career he served on committees for numerous other graduate students. Many of his students have made significant contributions to the profession, both in the U.S. and abroad.

In 1982, Havlicek went to the University of Maryland to become Chairperson of the Department of Agricultural and Resource Economics. In 1984, he was offered the position of Chairperson of the Department of Agricultural Economics and Rural Sociology at The Ohio State University. The opportunity to lead the department from which he had received his B.S. and M.S. degrees was exciting, and he returned to Ohio. Under Havlicek, the department grew professionally in faculty and received increased national and global recognition. In 1993 he received the OSU Distinguished Affirmative Action Award for his hiring practices to include women and minorities.

Havlicek's leadership abilities were recognized by his professional organizations. He served on the Editorial Council of the *Southern Journal of Agricultural Economics* (1975-77) and the *American Journal of Agricultural Economics* (1981-82), and on the Board of Directors for the American Agricultural Economics Association (1981-84), AAEA Foundation (1987-88), and The Economics Institute, Boulder, Colorado (1985-88). Havlicek was elected and served the Southern Agricultural Economics Association as First Vice President (1980-81), President-Elect (1983-84), President (1984-85), and Past President (1985-86). He was elected and served AAEA as President-Elect (1985-86), President (1986-87), and Past President (1987-88) In 1989, he was named a Fellow of the AAEA. Following his death, AAEA established the Joseph Havlicek Appreciation Club. Its purpose is to bestow travel grants to graduate students and foreign faculty to travel to the AAEA national meetings. In 1995, Havlicek was posthumously awarded the SAEA's Lifetime Achievement Award.

In 1990, Havlicek returned to teaching and research at OSU. He traveled to Taiwan, China, Hungary, Poland, and Czechoslovakia seeking to establish joint research and graduate exchange programs. With the opening of Eastern Europe, Havlicek had the opportunity to teach economic policies under a democracy to farm co-op managers in the Czech Republic and to students of agricultural economics at the Agricultural College of Prague University. Havlicek was able to teach these classes in the Czech language. He became instrumental in establishing a section on the program of the AAEA meetings for European topics and in finding ways for the faculty of Eastern Europe to attend the AAEA national meetings. Among his last responsibilities was work with Prague University to establish programs between that university and OSU. He also was instrumental in arranging a short course on farm loans at OSU for bankers in the Czech and Slovak Republics.

Shortly before his death in March 1993, Havlicek was asked to identify an area in his professional career for which he would like to be honored. His response was price analysis and applied econometrics. All of those who have worked with Havlicek have felt his commitment to applied econometrics. Because of Havlicek's belief that learning should be ongoing from student days throughout the professional's life, the Department of Agricultural Economics and Rural Sociology decided that a memorial lecture series in applied econometrics should be established. Funds for the lectures (to be presented annually in perpetuity) were contributed by Havlicek's colleagues, friends, and the Havlicek family.

The Havlicek Memorial Lecture Series in Applied Econometrics was established in 1993 and the first lecture was given by Professor Oral Capps, Jr., who completed an AAEA award-winning Ph.D. dissertation under Havlicek in 1980.

GROWTH RATE CONVERGENCE, FACT OR ARTIFACT?

An Essay on the Use and Misuse of Panel Data Econometrics

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
http://www.arec.umd.edu/mnerlove/mnerlove.htm

ABSTRACT

The sensitivity of the convergence rate, or the test of no convergence in the standard Barro-Baumol sense, to the econometric method employed is investigated. Two basic models are investigated: The first is the standard model with individual-specific effects for each country; the second is a model in which individual countries have different individual-specific trends in output per capita. All of the results reported support the growth convergence hypothesis conditional on savings and population growth rates in the usual sense but illustrate the rather different estimates of the rates of convergence obtained when different estimation techniques are used. 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 (to within 90% of equilibrium only in excess of 50 years). 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 determinants of growth may be seriously affected. Alternative maximum likelihood methods are developed which utilize the information contained in the initial observations for each country that reflect the operation of the growth process prior to the time at which we began to observe them.

This paper is based on an earlier paper of similar title presented at the Sixth Conference on Panel Data Econometrics, Amsterdam, 28-29 June 1996. The present version has been prepared for the 1998 Havlicek Lecture in Applied Econometrics at Ohio State University, April 16, 1998.

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 (1995) But there is now a vast literature utilizing the data from the Penn World Tables , which I also use here. I do not attempt a comprehensive survey of this literature in this paper.

I thank Hashem Pesaran for helpful discussions, and Robert Barro, Michael Binder, William Greene, G. S. Maddala, and C. Spohr for useful comments . 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 my earlier results . Suh also checked and double checked all derivations and verified that my programs accurately reflected the formulae derived.

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.[1] 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, some authors 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. (When population growth is endogenously determined, this implication of the neoclassical model of economic growth does not necessarily follow; see Nerlove and Raut, 1996.)

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 Islam (1995). 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.[2] 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.

In section 2, I derive a discrete form of the BB-convergence equation, also derived in a different way by Mankiw, et al. (1992), and show that the usual equation for testing the convergence hypothesis can be obtained from it by a simple partial adjustment model rather than the approximation about equilibrium usually employed. In general, no essential restrictions are neglected in this simpler derivation, although in certain contexts the specification may neglect certain cross-equation parameter restrictions. It is the classic Solow/Swan model that, for good or for ill, underlies all recent studies of the determinants of

---

[1] There is a good deal of current discussion regarding the appropriate definition of "convergence." Barnard and Durlauf (1995) give a nice discussion emphasizing the restrictiveness of the approach adopted here (see also Quah, 1996), which is the most prevalent, going back to the earlier work of Barro (1991) and Baumol (1986). Since my main concern is to show that the econometrics matters, I will adopt this notion of convergence in what follows, while admitting that my conclusions about convergence per se apply only to this rather narrow definition. However, the general point that inconsistency in the estimation of the coefficient of the state variable implies inconsistency in the estimation of the effects of other factors is valid in this context and in general.

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

growth and provides the justification for including a lagged value of the dependent variable, initial condition, or other state variable,

Section 3 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.[3] In a dynamic context a single cross-section is best viewed 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 I do examine the effects on the estimated coefficients of the Barro-Lee (1993) estimates of the stock of human capital or a trend variable.

In this section, I look at the problem of implementing the convergence hypothesis in the context of cross-country data over time, and, in particular whether the model ought to be recast so that individual countries are subject to individual specific trends. Because the maximum likelihood methods implemented in the paper require stationarity, I argue that the date should be differenced in order to implement this modification of the basic model. I refer to the two forms as the "levels" model and the "first-differenced" model, respectively.

In section 4, I discuss five existing methods of estimating the rate of convergence (and thus testing for convergence), show that four standard methods yield estimates which satisfy an inequality derived by Trognon and Sevestre (1996), and devise 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 further show that under a mildly restrictive assumption this unconditional maximum likelihood method tends to maximum likelihood conditional on the initial observations as the cross-section dimension of the panel increases. The "biases" in conventional panel methods are reflected in the estimates of the effects of other variables included in differing ways illustrated here.

Finally, in section 5, I apply all six methods to several panel data sets drawn from the Penn World Tables. 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. I argue 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. Results not presented here 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 such as trend or the stock of human capital, making inferences about the determinants of growth problematic unless appropriate econometric methods are used.

Section 6 concludes.

2. The Neoclassical Theory of Growth Convergence

The Solow/Swan model of economic growth is the starting point for all recent empirical analyses of growth convergence and the determinants of growth. In this section, I review briefly the model and dynamize it in a particularly simple approximate way.

a. The Solow/Swan Model

Consider the standard Solow/Swan (Solow, 1956; Swan, 1956) model with exogenous population growth in discrete form: Let Y_t = output, K_t = capital stock, N_t = labor force assumed to be the same as population, S_t = savings, I_t = investment, s = the savings rate, δ = depreciation rate, and $\bar{n}$ = the exogenous rate of growth of population and labor force. Production can be represented by a constant returns to scale function:

$$(1) \qquad Y_t = F(K_t, N_t)$$

[3] 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 do not agree that their formulation of the unconditional likelihood function is equivalent to the one I present below.

or

$$y_t = f(k_t),$$

where $y_t = Y_t/N_t$, $k_t = K_t/N_t$, and $f(k) = F(k, 1)$. Solow/Swan assume that savings equals gross investment and is a constant fraction s of output:

(2) $$I_t = S_t = sY_t.$$

The change in the capital stock equals gross investment minus depreciation:

(3) $$K_{t+1} = (1-\delta)K_t + I_t = sF(K_t, N_t) + (1-\delta)K_t.$$

Population grows exogenously at a rate $\bar{n}$:

(4) $$N_{t+1} = (1+\bar{n})N_t.$$

Thus

(5) $$k_{t+1} = \frac{sf(k_t) + (1-\delta)k_t}{1+\bar{n}} = g(k_t),\ k_0 \text{ given.}$$

The dynamics of the Solow/Swan model is entirely described by the path of k_t, the capital-labor ratio, since population grows exogenously, capital depreciates at a fixed rate, and gross investment is proportional to output.

The existence of stationary solutions to (5), i.e. k^* for which

(6) $$k^* = g(k^*),$$

and the local stability of such solutions depends on the shape of the function g. The conditions which yield a non-negative globally stable steady state solution are the following:

$g'(0) > 1$, $g(0) = 0$,

$g'(k) < 1$, for some $k > 0$, and g is concave.

These properties follow if the production function satisfies:

$f(0) = 0$

$f'(0) > \dfrac{\delta + \bar{n}}{s}$

$f'(k) < \dfrac{\delta + \bar{n}}{s}$, for some $k > 0$,

and f is concave.

A stationary solution k^* is locally stable if $|g'(k^*)| < 1$. Clearly $k^* = 0$ is unstable. Under concavity of f, whenever (6) holds for some $k^* > 0$, then there can be no other $k^* > 0$ for which (6) holds and at that point $|g'(k^*)| < 1$, so the solution is necessarily unique.

b. Growth Convergence in the Neoclassical Model with Exogenous Population Growth and Savings

If $f(k_t)$ is Cobb-Douglas, we can solve explicitly for the time path of y_t:
For $y_t = A_t k_t^{\alpha}$, so that

(7) $$k_t = \left(\frac{y_t}{A_t}\right)^{1/\alpha},$$

and, hence, from (5)

(8) $$y_t = \frac{n+1}{s}k_{t+1} - \frac{1-\delta}{s}k_t = \frac{n+1}{s}\left(\frac{y_{t+1}}{A_{t+1}}\right)^{1/\alpha} - \frac{1-\delta}{s}\left(\frac{y_t}{A_t}\right)^{1/\alpha}.$$

α is the elasticity of output with respect to capital stock; if capital is fully employed and paid its marginal product it is the implied share of capital in total output. A_t is any function of time which may affect the productivity of capital and labor, for example neutral technical change or, more explicitly, investment in human capital or in infrastructure. In the "stationary state"

(9) $$y_t^* = \left(\frac{n+\delta}{s}\right)^{-\alpha/(1-\alpha)} A_t^{1/(1-\alpha)}.$$

Taking logs

(10) $$\log y_t^* = \frac{-\alpha}{1-\alpha}\log(\frac{n+\delta}{s}) + \frac{1}{1-\alpha}\log A_t.$$

Although (8) shows that the rate of convergence to equilibrium is not constant, an approximation is given by a partial adjustment model

(11) $$\log y_t - \log y_{t-1} = (1-\gamma)[\log y_t^* - \log y_{t-1}].$$

This yields the approximate relation to be estimated and the equation employed in recent studies:

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

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. In what follows, I take account of differing s and n over time and cross-sectionally, in some additional analyses I take account of other factors such as investment in human capital or trend, but not infrastructure investment, which might arguably affect A.[4] In some further investigations not reported here, I have included A_t represented by the stock of human capital over time as measured by Barro and Lee (1993) raised to the power φ, $A_t = H_{it}^{\varphi}$, so that the coefficient of $\log H_{it}$ in the implied regression equation is $\frac{(1-\gamma)\varphi}{1-\alpha}$; or, alternatively, a simple linear

[4] As pointed out in a recent paper by Binder and Pesaran (1996), it makes a good deal of difference to the question of convergence at just what point one includes the stochastic disturbance. In this paper I follow the usual practice of tacking it on the end of the equation to be estimated.

trend, with coefficient τ, in place of the entire term $\frac{1-\gamma}{1-\alpha}\log A_t$ might be included.[5] This gives rise to quite a different model, particularly if one argues that the trends ought to be different for different countries. I will refer to (12) where A may differ from country to country but is assumed to be constant over time as the "levels" model. When the term $\frac{1-\gamma}{1-\alpha}\log A_t$ is replaced by a linear trend with slope which may differ from country to country but is assumed to be constant over time, (12) is replaced by

$$(12') \qquad \Delta\log y_t = \frac{\alpha(1-\gamma)}{1-\alpha}\Delta[\log s - \log(n+\delta)] + \nu + \gamma\Delta\log y_{t-1},$$

where ν is the slope of the trend and may differ from country to country. I refer to this model as the "first-difference" model.[6]

3. Empirical Investigations of Convergence and the Rate of Convergence

Equation (12) is 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). 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 initial 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 γ = 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. Details are given in Appendix Table 2.

The estimates of γ 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. 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 γ = 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.[7]

---

[5] The human capital variable constructed by Barro and Lee(1993) is available only for a subsample of countries by quinquennia for the period 1965 - 1985. Essentially the stock of human capital in the population is measured by Barro and Lee as the average schooling in the population as a whole over 25. In the results referred to, this variable is often the wrong sign and never significant.

[6] Differencing the model to achieve stationarity creates some interesting problems with respect to the disturbance term which are discussed below. See especially footnote 12.

[7] 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. Using long annual time series (1865 - 1994) for 16 OECD countries,

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, 1995) 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 initial, or lagged 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 these findings are probably statistical artifacts arising from biases in the econometric methods employed to test the growth convergence hypothesis in the BB sense. This does not mean that this sense is the correct one to employ in the more general context of convergence, as emphasized by Quah(1996), but demonstrates the sensitivity of the conclusions drawn about γ to the econometric method employed, irrespective of the validity of the relationship of such conclusions to more general notions of convergence.

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 studies. Systematic differences in these across countries or regions will systematically bias the conclusions. To the extent that these effects are approximated by differing trends for each country omitting such variables may account for the differences between the first-difference model and the levels model. Because such omitted 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 (1995) 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. Results for the alternative method of unconditional maximum likelihood suggested in Nerlove and Balestra (1996) are presented here.[8]

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

---

Michelacci and Zaffaroni (1997) extend the Solow-Swan model to allow for cross-sectional heterogeneity in the pace of convergence and conclude that the uniform 2% rate of convergence found in much of the empirical literature is a consequence of fractionally integrated nonstationarity with underlying parameter strictly between 0.5 and 1.0. My first-difference formulation is equivalent to assuming a unit root, an hypothesis which is rejected by Micelacci and Zaffaroni.

[8] See also Nerlove (1999). Lee, et al. (1996) also estimate from what they term 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 based on what they refer to as the unconditional density of the dependent variable is equivalent to mine.

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.

In the remainder of this paper, I investigate the sensitivity of the convergence rate, or the test of no convergence, to the econometric method employed as well has 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. In additional research not presented here I also show that the coefficients of other explanatory variables vary considerably 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.

4. Alternative Methods for Estimating Rates of Convergence[9]

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 to show the 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.

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.

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

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. Clearly, however, y_{it-1} cannot be assumed to be uncorrelated with μ_i. It is clear, therefore, that OLS applied to (13) ignoring the component nature of the disturbances $v_{it} = \mu_i + \varepsilon_{it}$, 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.

[9] 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.

(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 (13) as

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

Equation (14) 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}:

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

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 (15), 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 (1).

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 (13).

(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:

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

Although $\sigma_{x\varepsilon}^2 = 0$,

(17) $$\sigma_{y(-1)\varepsilon}^2 = p\lim \frac{1}{T}\sum_t \left(y_{it-1} - \bar{y}_{i-1}\right)\left(\varepsilon_{it} - \bar{\varepsilon}_{i-1}\right)$$
$$= -\frac{1}{T^2}\frac{T-1-T\gamma+\gamma^T}{(1-\gamma)^2}\sigma_\varepsilon^2 \neq 0.$$

Thus, the OLS estimates of both β and γ in the fixed effects model are inconsistent, although as $T \rightarrow \infty$, 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.

(c) Country Means Regression and the Estimation of ρ

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 many computer packages for implementing FGLS for panel data or for 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:

(18) $$\overline{y_{i.}} = \alpha + \beta \overline{x_{i.}} + w_i \text{ , where } w_i = \mu_i + \overline{\varepsilon_{i.}} \text{ .}$$

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 (18) will tend to fit too well if cross-section variation dominates the data.[10] σ_μ^2 is then estimated as

---

[10] 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 (18) 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) and used to obtain FGL

$\sigma_w^2 - \frac{\sigma_\varepsilon^2}{T}$, 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.

(d) 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

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

---

estimates below. 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.

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. 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. Levels model only:

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

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

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

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.

(e) 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:

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

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

(21) $$\lambda^* = K(1-\rho) / \{\frac{(1-\gamma^T)E(y_{i0}\mu_i)}{(1-\gamma)\sigma^2} + K\xi\},$$

$$\text{where } K = \frac{T-1-T\gamma+\gamma^T}{T(1-\gamma)^2}, \text{ and } \rho,\ \xi, \text{ and } \sigma^2 \text{ are as before.}$$

They also show that when $\beta \neq 0$, 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 (13) 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

(22)
$$\log L(\alpha,\beta,\gamma,\sigma_\mu^2,\sigma_\varepsilon^2 | y_{11},\ldots,y_{NT}; x_{11},\ldots x_{NT}; y_{10},\ldots,y_{N0})$$
$$= -\frac{NT}{2}\log 2\pi - \frac{NT}{2}\log\sigma^2 - \frac{N}{2}\log\xi - \frac{N(T-1)}{2}\log\eta$$
$$-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it}^{*} - \alpha\xi^{-1/2} - \beta x_{it}^{*} - \gamma y_{it-1}^{*}\right)^2,$$

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

$\xi = \dfrac{T}{1+\lambda(T-1)}$ and $\eta = \dfrac{\lambda T}{1+\lambda(T-1)}$, 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, $w \sim wn\,(0, \sigma_w^2 I)$, also assumed in the Monte Carlo experiments reported in Nerlove (1971),

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

maximization of the conditional likelihood function (22) 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 (13),

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

such that when $\gamma < \gamma^*$ there exists an interior maximum of (22) which yields consistent ML estimates, but that when $\gamma \geq \gamma^*$ there are values of ρ for which the conditional likelihood function (13) 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 $\dfrac{\sigma_w^2}{\sigma^2} = 1.0$, $\gamma^* = -0.11$ 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 (22) $\rho = 0$ and inconsistent estimate of all the parameters. Using these results, Trognon (1978) is able to replicate the Monte Carlo results reported in Nerlove (1971).[11]

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. (Birnbaum, 1962; Barnard, Jenkins and Winsten, 1962.) For this reason, sections of some of the multidimensional likelihood functions are also presented in the next section. When first 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.[12] I do not attempt to implement this model and approach in this paper but leave the matter for a separate investigation.

(f) Unconditional Maximum Likelihood.

While it is not guaranteed that no boundary solution to the likelihood equations is obtained, yielding inconsistent estimates, it is apparent 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

---

[11] Maddala (1971, pp. 346 - 347) gives a condition for the gradient of the concentrated likelihood function to be positive at a boundary $\rho = 0$ (OLS on the pooled data) for the conditional likelihood function. So if ρ is constrained to the interval [0,1) this implies a local maximum at the boundary 0. Breusch (1987) shows that this condition can be easily checked at the start of his iterative GLS procedure by beginning with the pooled OLS estimates and $\rho = 0$. Unfortunately these results apply only to the likelihood function when no lagged value of the dependent variable is included or when those initial values are conditioned upon. I have not been able to derive a similar result for the unconditional likelihood function below.

[12] Adding trend, t, to (13)

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

(13'') $\Delta y_{it} = \beta \Delta x_{it} + \gamma \Delta y_{it-1} + \tau_i + \Delta\varepsilon_{it}$, 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:

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

The characteristic roots of A give the necessary transform and Jacobian. This should be 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, but then a trend variable must be included as explanatory with as many different slopes as countries. See Nerlove (1999, section 2.2) for an extended discussion of the transformation required to render the time-varying part of the disturbances serially uncorrelated.

$|\gamma| \geq 1$ 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 difficulties associated with the formulation of the unconditional likelihood function in the case in which deterministic or stochastic trends are included are discussed in footnote 12 above.

Under this assumption, the dynamic relationship to be estimated is stationary and $|\gamma| < 1$. Consider equation (14)[13] with the intercept eliminated, for y_{i0} and the infinite past:

$$(24)\quad y_{i0} = \sum_{j=1}^{\infty} \gamma^j \beta x_{i,-j} + \frac{1}{1-\gamma}\mu_i + v_{i0}\,, \text{ where } v_{it} = \gamma v_{it-1} + \varepsilon_{it}\,.^{14}$$

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

$$(25)\quad \log L(\gamma,\rho,\sigma_\mu^2,\sigma_\varepsilon^2 \mid y_{11},\ldots,y_{NT};\ldots;y_{10},\ldots,y_{N0})$$

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

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

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

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

---

[13]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 (24) follows.

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

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

Thus, the concentrated LF is

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

This 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 $\sum_{i=1}^{N} y^2_{i0}$, as well as on ρ and γ, which enter the final terms as well.

When $\beta \neq 0$, things are more complicated still. Various alternative specifications considered in the literature are reported and analyzed in Sevestre and Trognon (1996, pp. 136-138).[15] 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 (24) is

$\varphi_{i0} = \beta\sum_{j=0}^{\infty}\gamma^j x_{i,-j}$. 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

$$\sigma^2_{\varphi_i} = \frac{\beta^2\sigma^2_{x_i}}{1-\gamma^2} \quad . \qquad (26)$$

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

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

---

15 One interesting possibility discussed by Trognon and Sevestre (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 (13), the equation to be estimated?

has a well defined variance which is the same for all i and a function of β, γ, and σ_x^2 . This then enters the final term in the unconditional likelihood (25), which now becomes:

$$(27)\ \log L(\beta,\gamma,\sigma_\mu^2,\sigma_\varepsilon^2 | y_{11},\ldots,y_{NT};x_{11},\ldots x_{NT};y_{10},\ldots,y_{N0})$$

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

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

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

Concentrating the likelihood function to permit a one- or two-dimensional grid search is no longer possible. If gradient 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.

5. Empirical Evidence on Growth Rate Convergence and the Comparative Performance of Different Panel Data Methods

(a) Numerical Results.

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/. The countries are listed in the Appendix Table A-1. 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. Characteristics of the sample are reported in Table 1. The results of the six methods applied to these data or to their first differences are reported in Table 2 and 3. In the case of the latter, an appropriate transformation of the original data is made to eliminate the serial correlation introduced in the time-varying part of the disturbance by the first-difference transformation. I have listed the regression methods in the order in which the corresponding estimates of γ appear in the inequality of

Trognon and Sevestre (equation 20 above). These estimates are followed by the maximum likelihood estimates conditional on the initial values y_{i0} 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, for my purpose here, omission of relevant variables simply increases the unexplained disturbance variance and thus heightens the contrast among alternative estimators.

TABLE 1: COMPARATIVE DESCRIPTIVE STATISTICS FOR THE TWO SAMPLES

Item	94-country sample	22-country sample	Ratio 22/94 values
Variance initial y	0.799	0.256	0.320
Variance final y	0.899	0.222	0.247
Correlation between y0 and y5	0.988	0.090	0.092
Variances about overall means:			
y	1.058	0.204	0.193
z	0.698	0.040	0.058
Pooled variances about country means:			
y	0.045	0.040	0.897
z	0.083	0.007	0.084

Turning now to the regression estimates presented in Tables 2 and 3:

[Insert Table 2 near here.]

Consider the first four methods applied to the levels model. 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. 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 considered by Trognon and Sevestre. Pooled OLS and FGLS also stand in the order predicted by the Trognon-Sevestre 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 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 capital's share 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 2. In the case of 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. The contrast with the 94-country sample is striking: 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. As found earlier, the estimates of β are quite insensitive to the method used, but the estimates of γ are not very different either; consequently the implied estimates of capital's share are similar, albeit different for the two samples.[16]

Turning now to the parameter estimates for the first-difference model presented in Table3.

[Insert Table 3 near here.]

The contrast with the levels model is remarkable; at least in terms of reasonableness, at last we seem to be in the right "ball park." Consider the first four methods applied to the levels model. The estimates of γ for the 94-country sample range from a low of 0.40 (fixed-effects regression) to a high of 0.92 (country means regression) with pooled OLS and FGLS falling in between. For the OECD countries the range is 0.36 to 0.72. The implied speed of convergence thus much more reasonable that obtained for the levels model, although none could be characterized as evidence of reasonably rapid convergence. The estimates of γ no longer satisfy the Trognon-Sevestre inequality. FGLS is now lower than the the estimate obtained by fixed effects OLS. Pooled OLS is greater than FGLS and fixed-effects OLS, as predicted by the Trognon-

[16] What accounts for these remarkable differences between the two samples and for the similarity of the unconditional and conditional ML estimates for the 94-country sample? 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 $\mathrm{var}y_0$ 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 } =

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

This function is clearly decreasing in N and $\mathrm{var}y_0$. Its behavior with respect to φ^2 depends on the relation between $\mathrm{var}y_0$ and φ^2; when $\mathrm{var}y_0 > \varphi^2$, it is increasing; but when $\mathrm{var}y_0 < \varphi^2$, it is decreasing. Thus for a given $\mathrm{var}y_0$ the log of the ratio of the unconditional to the conditional density tends to zero, i.e. the ratio tends to one, as N increases and as φ^2 increases. In other words the unconditional becomes more and more like the conditional likelihood. For $\beta = 0.15$, $\gamma = 0.85$, $\sigma_x^2 = 0.08$, $\rho = 0.8$ and $\sigma^2 = 0.02$, $\varphi^2 = 0.73$. Table 3 presents some descriptive statistics for the two samples. Typically, $\mathrm{var}y_0$ is much less than 0.73, for example, 0.256 for the 22-country sample. On the other hand it is about that value for the 94-country sample. Thus the principle explanation for the similarity of the conditional ML and the unconditional ML estimates for the 94-country sample is the size of the cross-section dimension; similarly, the small sample in the OECD case accounts for the lack of similarity.

Sevestre results. As was the case for the levels model, the estimated value of β is sensitive to the method of estimation, although generally less so, and the estimate of the elasticity of output with respect to capital stock in the production function is more 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 these estimates bring us much closer to what could be considered a reasonable figure.

The most plausible estimates are those for conditional and unconditional maximum likelihood presented as methods 5 and 6 in Table 3. In the case of the 22 country OECD sample, only the estimates differ from one another. Implied speeds of convergence are still, however, quite slow for the OECD countries, but much, much faster for the larger group of 94 countries. Perhaps one can conclude that if differing country-specific trends are taken into account, so-called "beta convergence" obtains at a reasonably rapid rate for a group of diverse countries, but when the relatively homogeneous group of OECD countries is considered, convergence is much more problematic.

(b) Graphical Results.

Further insight into the support for the convergence hypothesis, as modified in this paper, which is given by the likelihood functions for the two samples can be obtained graphically. Since the first-difference model gives by far the most plausible results, I present only graphs for this case. Bear in mind, however, that the convergence concept considered is "beta" convergence with a vengeance. Not only are the results conditioned on differing savings and population growth rates, but I am allowing for differing linear trends among countries. The model is reduced to stationarity by first-differencing the quinquennial averages. The explanatory variables, in this case first differences of savings rates and population growth rates are assumed to be exogenous and to be determined by some sort of stationary process. 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. Although there are $\binom{4}{2} = 6$ possible pairs to consider, I focus on the crucial pairs: ρ vs. γ and β vs. σ^2 . It is important that the likelihood functions as formulated reflect the operation of the process which generated the data before we began to observe them; the appearence of the unconditional likelihood functions is rather different even though the ML estimates are quite close to those given by the unconditional likelihood functions.

Figures 1 and 2 give one-parameter slices and 2-parameter 3-D plots of the likelihood for the pairs ρ-γ and β-σ^2 for the 94- and 22-country samples, respectively. Each Figure consists of two papges with four plots each. The figures are at the end of the paper. The main finding is that the difference between the estimates for the two samples is not great except for a much slower speed of convergence (larger γ) for the more diverse 94-country sample and is well-supported by the shape of the likelihood functions in the two cases.

The likelihood reaches a unique maximum in every case. Except for σ^2, the functions are well-behaved in the vicinity of the maximizing parameter values. Since the value of σ^2 is bounded from below by zero, the graph has the typical shape found in regression problems: a sharp rise from near zero followed by a long slow decline. I find no evidence of double maxima or a boundary maximum of the likelihood function with respect to ρ. There is clearly considerable "trade-off" between ρ and γ for the 94-country sample and to a lesser degree for the 22-country sample.

6. 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 based on analyses of 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. 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. When differing country-specific trend are taken into account and when likelihood methods are employed which take into account the operation of the growth process prior to the point at which the data sample begins, however, the results are dramatically different: convergence to paths conditioned on differing savings and population growth rates and country specific trends is quite rapid for the relatively homogeneous 22-contry sample, although it is still very slow for the more heterogeneous 94-country sample.

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 except for one reversal in the case of the first-difference model.. 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 has been carried out here without significant trauma.

REFERENCES

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., 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.

Breusch, T. S., "Maximum Likelihood Estimation of Random Effects Models," *Journal of Econometrics, 36*: 383 - 389, 1987.

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

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

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

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.

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.

Michelacci, C., and P. Zaffaroni, 1997. "Beta Convergence," unpublished working paper, EM/97/332, Suntory Center, London School of Economics and Political Science.

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., 1994. "Reflections on Agricultural Development Population Growth and the Environment." Presented at the Third Conference on Development Economics, Asian Development Bank, Manila, November 23-25, 1994. Pp. 299 - 347 in M. G. Quibria and J. M. Dowling, *Current Issues in Economic Development*, New York: Oxford University Press, 1996.

Nerlove, M., 1997. "Likelihood Inference for Dynamic Panel Models." Presented at the Seventh Conference on Panel Data Econometrics, 19 - 20 June 1997, Paris. Forthcoming in *l'Annales d'Économie et de Statistique* de l'INSEE.

Nerlove, M., 1998. "Properties of Alternative Estimators of Dynamic Panel Models: An Empirical Analysis of Cross-Country Data for the Study of Economic Growth." Forthcoming in *Analysis of Panels and Limited Dependent Variable Models*, C Hsiao, K. Lahiri, L-F. Lee and M. H. Pesaran, eds., Cambridge: Cambridge University Press.

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 - 1175 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," *European Economic Review, 40:*1353 - 1375.

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.

Appendix:

Data on 94 countries for the period 1960 - 1985 from the Penn World Tables 5.6, publicly available from the NBER web site at ftp://nber.harvard.edu/pub/.

22-Country Sample:

Japan
Austria
Belgium
Denmark
Finland
France
Germany (FRG)
Greece
Ireland
Italy
Netherlands
Norway
Portugal
Spain
Sweden
Switzerland
Turkey
U. K.
Canada
U. S.
Australia
New Zealand

94-country Sample = 22-Country Sample + the Following:

Algeria
Botswana
Cameroon
Ethiopia
Ivory Coast
Kenya
Madagascar
Malawi
Mali
Morocco
Nigeria
Senegal
South Africa
Tanzania
Tunisia
Zambia
Zimbabwe
Costa Rica
Dominican Rep.
El Salvador
Guatemala
Haiti
Honduras
Jamaica
Mexico
Nicaragua
Panama
Trinidad & Tobago
Argentina
Bolivia
Brazil
Chile
Colombia
Ecuador
Paraguay
Peru
Uruguay
Venezuela
Bangladesh
Hong Kong
India
Israel
Jordan
Korea
Malaysia
Burma
Pakistan
Philippines
Singapore
Sri Lanka
Syria
Thailand
Angola
Benin
Burundi
Central African Republic
Chad
Congo
Egypt
Ghana
Liberia
Mauritania
Mauritius
Mozambique
Niger
Rwanda
Somalia
Togo
Uganda
Zaire
Nepal
Papua New Guinea

TABLE 2: 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 Capital Share	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 Capital Share	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 Capital Share	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 Capital Share	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 Capital Share	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 Capital Share	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 3: 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 Capital Share	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 Capital Share	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 Capital Share	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 Capital Share	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 Capital Share	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 Capital Share	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 1.1: UNCONDITIONAL LIKELIHOOD AS A FUNCTION OF RHO

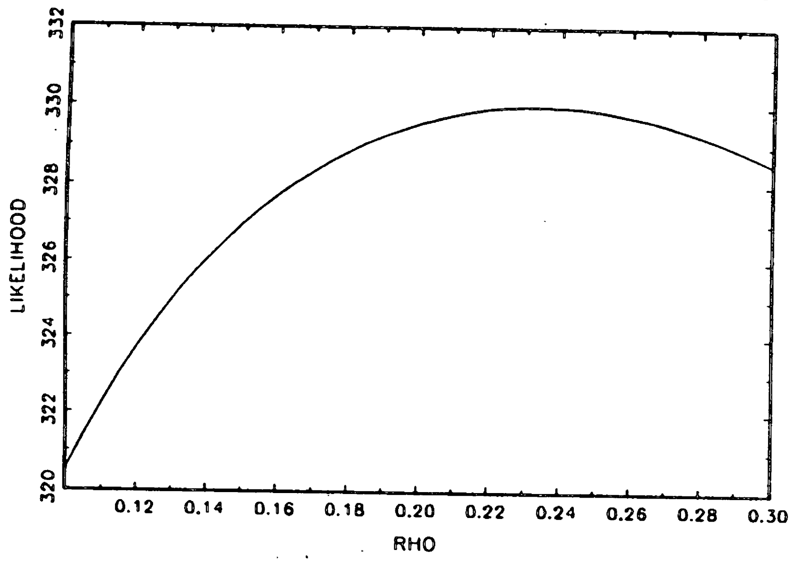

FIGURE 1.2: UNCONDITIONAL LIKELIHOOD AS A FUNCTION OF GAMMA

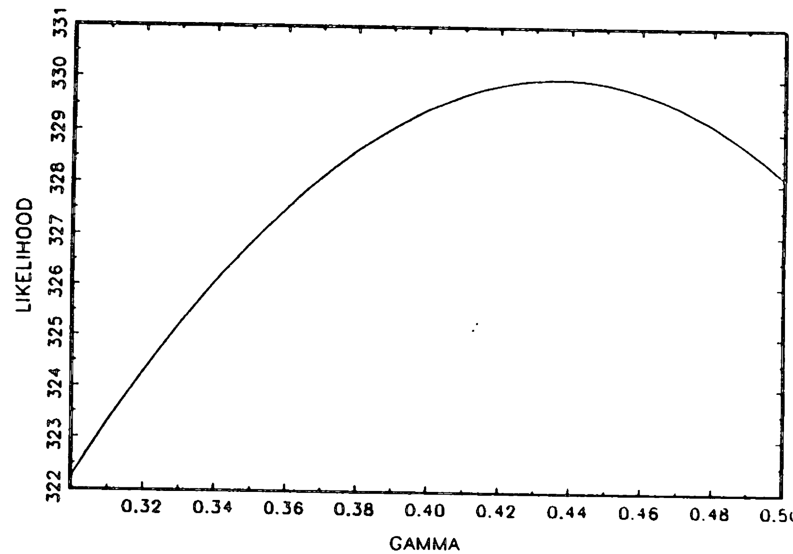

FIGURE 1.3: UNCONDITIONAL LIKELIHOOD, FIRST DIFFERENCES
94 COUNTRIES, RHO VS. GAMMA.

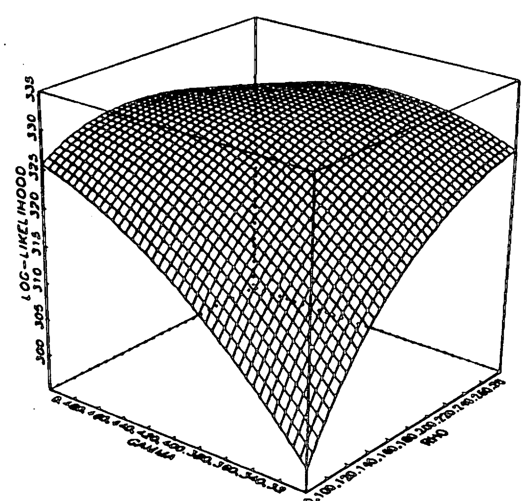

FIGURE 1.4: UNCONDITIONAL LIKELIHOOD, FIRST DIFFERENCES
94 COUNTRIES, RHO VS. GAMMA.

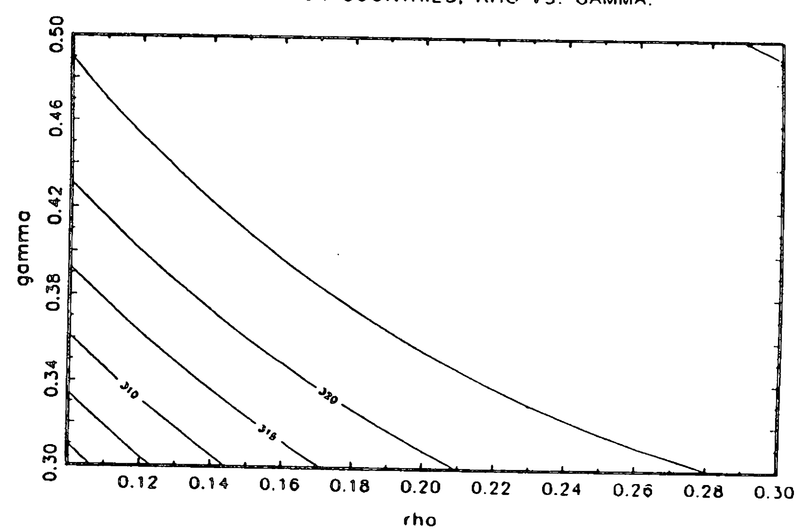

FIGURE 1.5: UNCONDITIONAL LIKELIHOOD AS A FUNCTION OF BETA

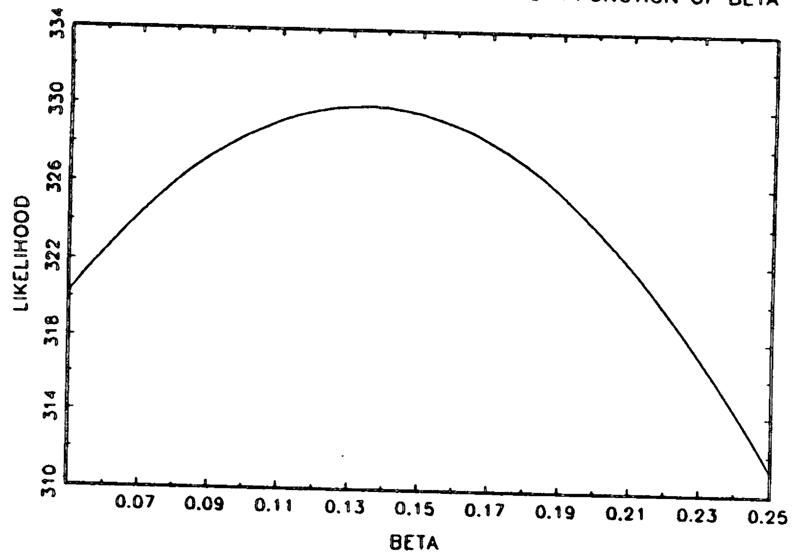

FIGURE 1.6: UNCONDITIONAL LIKELIHOOD AS A FUNCTION OF SIGMA2

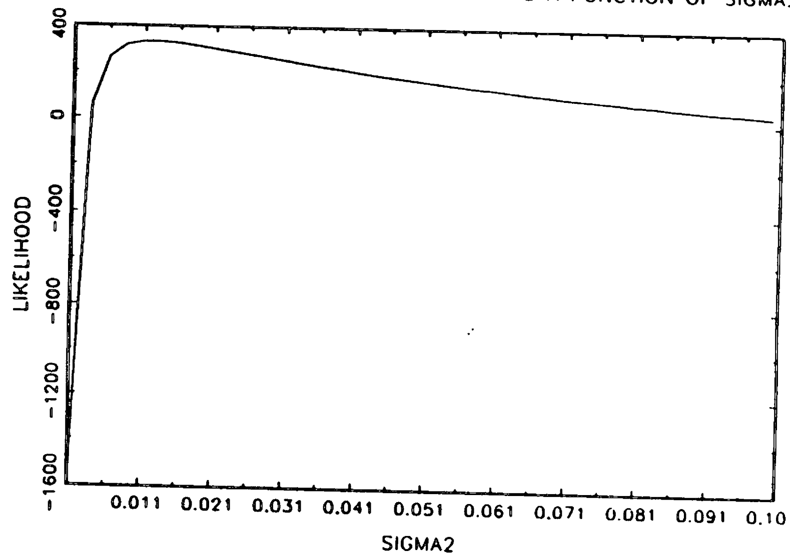

FIGURE 1.7: UNCONDITIONAL LIKELIHOOD, FIRST DIFFERENCES
94 COUNTRIES, BETA VS. SIGMA2.

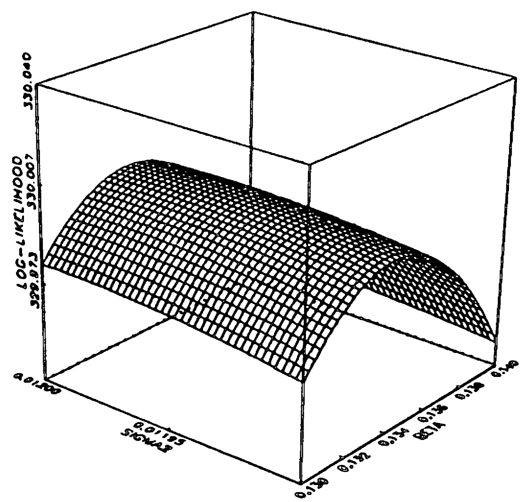

FIGURE 1.8: UNCONDITIONAL LIKELIHOOD, FIRST DIFFERENCES
94 COUNTRIES, BETA VS. SIGMA2.

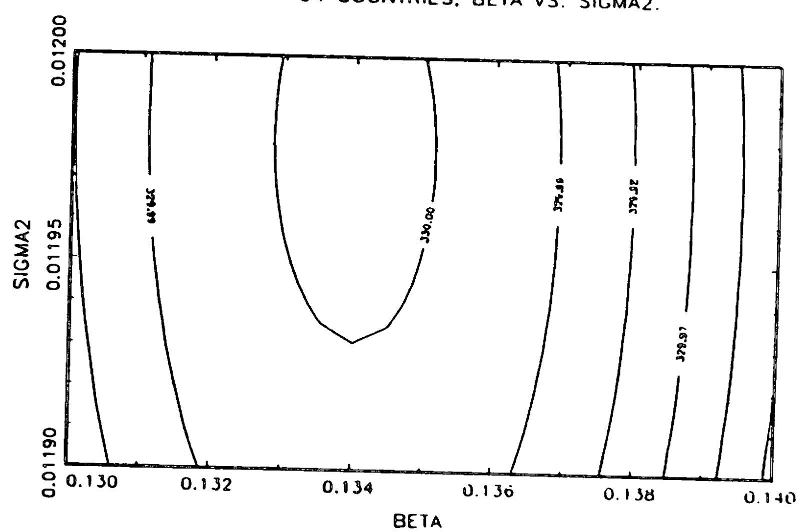

FIGURE 2.1: UNCONDITIONAL LIKELIHOOD AS A FUNCTION OF RHO.
FIRST-DIFFERENCES.
22 OECD COUNTRIES

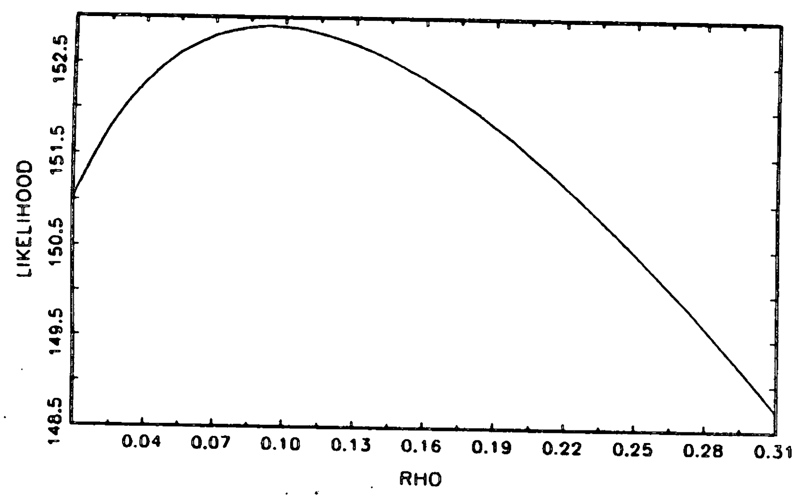

FIGURE 2.2: UNCONDITIONAL LIKELIHOOD AS A FUNCTION OF GAMMA.
FIRST-DIFFERENCS.
22 OECD COUNTRIES

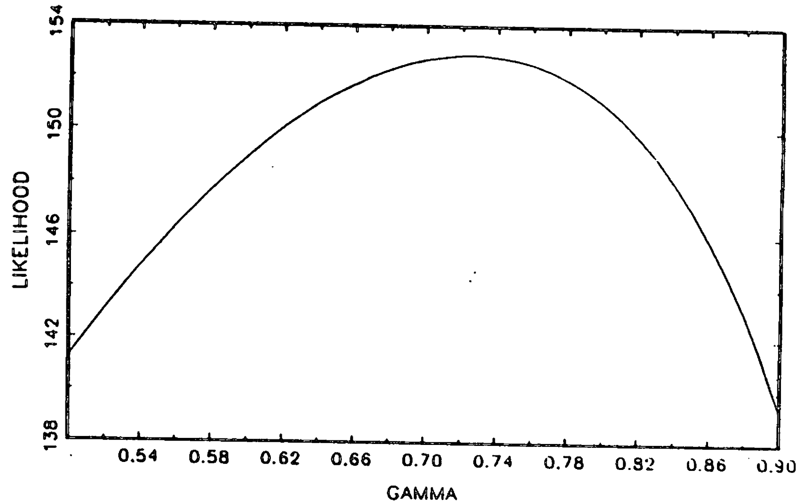

FIGURE 2.3: UNCONDITIONAL LIKELIHOOD.
FIRST DIFFERENCES
22 OECD COUNTRIES. RHO VS. GAMMA.

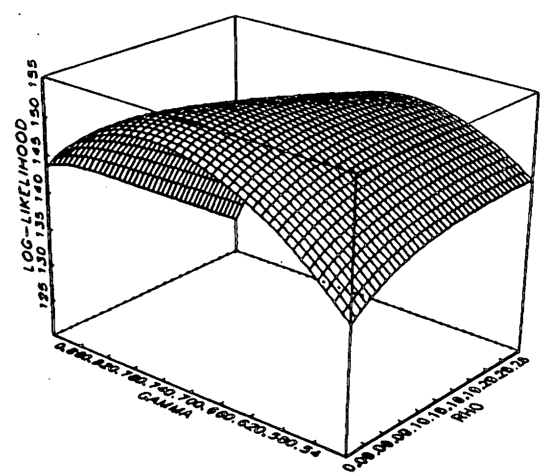

FIGURE 2.4: UNCONDITIONAL LIKELIHOOD.
FIRST DIFFERENCES.
22 OECD COUNTRIES. RHO VS. GAMMA.

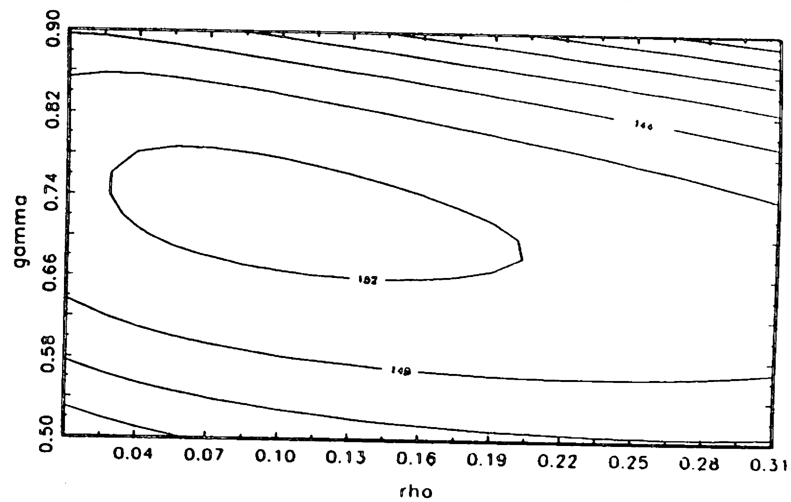

FIGURE 2.5: UNCONDITIONAL LIKELIHOOD AS A FUNCTION OF BETA.
FIRST DIFFERENCES.
22 OECD COUNTRIES

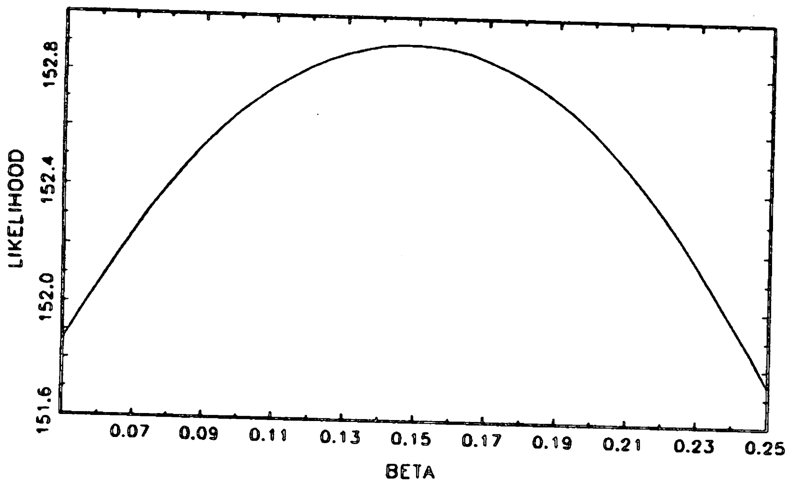

FIGURE 2.6: UNCONDITIONAL LIKELIHOOD AS A FUNCTION OF SIGMA2.
FIRST DIFFERENCES.
22 OECD COUNTRIES.

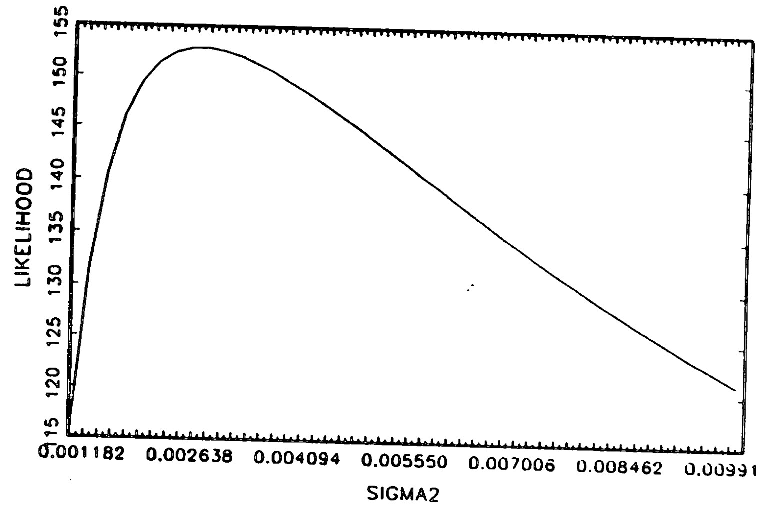

FIGURE 2.7: UNCONDITIONAL LIKELIHOOD.
FIRST DIFFERENCES
22 OECD COUNTRIES, BETA VS. SIGMA2.

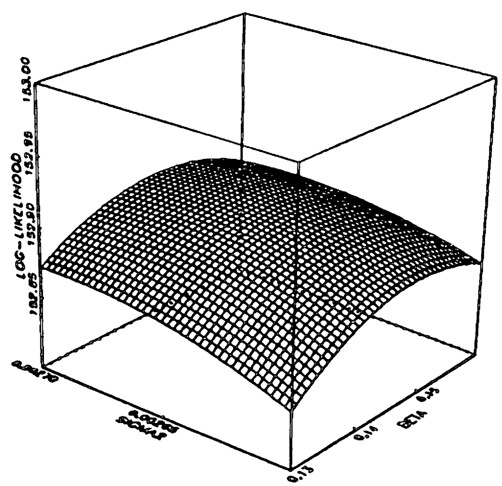

FIGURE 2.8: UNCONDITIONAL LIKELIHOOD.
FIRST DIFFERENCES
22 OECD COUNTRIES, BETA VS. SIGMA2.