TECHNICAL WORKING PAPER SERIES

ESTIMATING DETERMINISTIC TRENDS IN THE PRESENCE OF SERIALLY CORRELATED ERRORS

Eugene Canjels
Mark W. Watson

Technical Working Paper No. 165

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1994

This is a substantially revised and expanded version of "A Note on Estimating Deterministic Trends in the Presence of Serially Correlated Errors" by Watson. We thank Sergio Rebelo for posing a question that motivated this research. We have benefitted from discussions and comments by two referees, Edwin Denson, Michael Horvath, Robert King, James MacKinnon, Gauri Prakash and James Stock. Financial support was provided by the National Science Foundation through grants SES-91-22463 and SBR-94-09629. This paper is part of NBER's research program in Economic Fluctuations. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #165
September 1994

ESTIMATING DETERMINISTIC TRENDS IN THE PRESENCE OF SERIALLY CORRELATED ERRORS

ABSTRACT

This paper studies the problems of estimation and inference in the linear trend model: $y_t=\alpha+\beta t+u_t$, where u_t follows an autoregressive process with largest root ρ, and β is the parameter of interest. We contrast asymptotic results for the cases $|\rho| < 1$ and $\rho=1$, and argue that the most useful asymptotic approximations obtain from modeling ρ as local-to-unity. Asymptotic distributions are derived for the OLS, first-difference, infeasible GLS and three feasible GLS estimators. These distributions depend on the local-to-unity parameter and a parameter that governs the variance of the initial error term, κ. The feasible Cochrane-Orcutt estimator has poor properties, and the feasible Prais-Winsten estimator is the preferred estimator unless the researcher has sharp *a priori* knowledge about ρ and κ. The paper develops methods for constructing confidence intervals for β that account for uncertainty in ρ and κ. We use these results to estimate growth rates for real per capita GDP in 128 countries.

Eugene Canjels
Department of Economics
Northwestern University
Evanston, IL 60208

Mark W. Watson
Department of Economics
Northwestern University
Evanston, IL 60208
and NBER

1 Introduction

Many economic time series display clear trends well represented by deterministic linear or exponential functions of time. The slope of the trend function represents the average growth in the series (or rate of growth, if the series is in logarithms) and is often a parameter of primary interest. Serial correlation in the data complicates efficient estimation and statistical inference about the trend function, and this paper studies trend estimation and inference when this problem is severe.

To be specific, assume that a series can be represented as

$$y_t = \alpha + \beta t + u_t \tag{1}$$

$$(1 - \rho L)u_t = v_t \tag{2}$$

where y_t is the level or log-level of the series, and u_t denotes the deviations of the series from trend. These deviations are serially correlated, with a largest autoregressive root of ρ. The error term v_t is an I(0) process. If the $u_t's$ are jointly normally distributed, and the precise pattern of serial correlation is known, then efficient estimators of α and β can be constructed by GLS, and statistical inference can be conducted using standard regression procedures. In practise, the distribution of the errors and the pattern of serial correlation is unknown, so that GLS estimation and exact inference are infeasible.

Applied researchers typically use one of three feasible estimators, motivated by the asymptotic equivalence of these estimators to the infeasible GLS estimator. If $|\rho| < 1$, so that u_t is $I(0)$, then the feasible GLS estimator is asymptotically equivalent to the infeasible GLS estimator, under general conditions. Moreover, the classic result of Grenander and Rosenblatt (1957) implies that the OLS estimators of α and β are asymptotically equivalent to the GLS estimators. Thus, if u_t is $I(0)$, OLS or feasible GLS applied to the level of y_t is asymptotically efficient. On the other hand, when $\rho = 1$, so that u_t is $I(1)$, α can no longer be consistently estimated by any method, and the OLS estimator of β is no longer asymptotically efficient. In this case, the data should be differenced and the Grenander and Rosenblatt result implies that the sample mean of Δy_t (the OLS estimator of β in the differenced regression) is asymptotically equivalent to the efficient, but infeasible, GLS estimator of β. In summary, if u_t is $I(0)$ then OLS from the levels regression produces the asymptotically efficient estimator, while if u_t is $I(1)$ then the sample mean of Δy_t is the asymptotically efficient estimator.

Inference is just as dependent on the $I(0)/I(1)$ dichotomy. Ideally, in either situation, inference should be carried out using the t-statistic from the infeasible GLS regression. When u_t is $I(0)$, this t-statistic can approximated using the OLS estimator together with a serial correlation robust standard error estimated from the OLS residuals. Alternatively, when $\rho = 1$ and the data are $I(1)$, this t-statistic can be approximated using the sample mean of Δy_t together with a serial correlation robust variance estimated from the first differences of the data. Of course, since most researchers can't know *a priori* whether their data are $I(0)$ or $I(1)$, these results are of limited value. In this paper we study inference problems and the behavior of OLS, first-difference and feasible GLS estimators when the data are either $I(0)$ or $I(1)$ and ρ is unknown.

Our analysis builds on two literatures. The first is the literature on the linear regression model with AR(1) errors exemplified by Cochrane and Orcutt (1949) and Prais and Winsten (1954). The second is the literature on inference in regressions with $I(1)$ variables exemplified by Dickey and Fuller (1979), Durlauf and Phillips (1988) and Elliott, Rothenberg and Stock (1992). Much of the former literature focuses on efficient estimation of regression parameters when the errors follow a stationary AR(1) process, and is directly relevant for our analysis when $|\rho| < 1$ and v_t is *iid*.[1] There are few exact analytic results in this literature because of the dependence of results on the regressors and the nonlinearity introduced by feasible GLS estimation.[2] Moreover, the asymptotic results summarized above rely on $|\rho| < 1$ and are not refined enough to discriminate between OLS and feasible GLS estimators. Thus, the majority of work in this area has relied on Monte Carlo simulations. Equations (1) and (2) have also been extensible studied in the unit root literature, primarily with a focus on tests for the hypothesis that $\rho = 1$. In most of this literature, the regression coefficients α and β are nuisance parameters and ρ is the parameter interest.[3] One of the purposes of this paper is to highlight what this analysis says about the feasible estimators of β and statistical inference.

We begin our analysis in Section 2 by presenting results on the asymptotic distributions of estimators of β. These include the OLS, first-difference, infeasible GLS and three different, but commonly used, feasible GLS estimators. We avoid the sharp $|\rho| < 1$ and $\rho = 1$ dichotomy in the asymptotic distributions by using local-to-unity asymptotics, with the hope that these provide better finite sample approximations. The asymptotic results for $|\rho| < 1$ and $\rho = 1$ are not new: they are reported here for completeness and because, particularly when $\rho = 1$, the results may not be widely appreciated by applied researchers. In any event, the local-to-unity results are the most relevant, since in most econometric applications the errors are highly serially correlated, although perhaps not characterized by an exact unit root. These results

[1] There is large literature on this topic, including Beach and MacKinnon (1978), Chipman (1979), Kadiyala (1968), Maeshiro (1976 and 1979), Magee (1987), Park and Mitchell (1980), Rao and Griliches (1969), Spitzer (1979), and Thornton (1987).

[2] Two exceptions directly relevant for our analysis are Prais and Winsten (1954) and Chipman (1979). The first paper studies equations (1) and (2) when $\alpha = 0$ and v_t is *iid*, and calculates the relative efficiency of the OLS and first-difference estimators as a function of ρ and the sample size, T. Chipman (1979) relaxes the assumption on α and calculates the greatest lower bound of the efficiency of the OLS estimator for all T and $\rho \leq 1$. We discuss the Chipman (1979) analysis in more detail in Section 2.2.1

[3] A notable exception is Durlauf and Phillips (1988), which is discussed in more detail in Section 2.2.1.

show sharp differences in the relative efficiencies of the estimators and four conclusions emerge from the analysis. First, the Cochrane-Orcutt estimator performs very poorly when ρ is large. Second, the OLS estimator is more robust to variations in ρ than the first-difference estimator. Third, the variance of the initial error term has an important effect on the relative efficiencies of the estimators. Finally, the asymptotic results suggest that the feasible Prais-Winsten estimator is the best estimator in most applied situations. Section 2 concludes with a small finite-sample experiment that indicates that the asymptotics provide reasonable approximations to the finite-sample relative efficiencies.

Section 3 studies the problem of statistical inference about β. Existing Monte Carlo evidence suggests that methods relying on $I(0)$ asymptotic approximations greatly understate the uncertainty in β when $|\rho| < 1$ but large. This leads to confidence intervals that are much too small and hypothesis tests with sizes that are too large. Asymptotic approximations that rely on $\rho = 1$ have analogous problems. This section uses the local-to-unity asymptotic approximations from Section 2 to construct bounds tests and conservative confidence intervals building on methods developed in Dufour (1990) and Cavanagh, Elliott and Stock (1993).

In Section 4 we apply the methods to estimate and construct confidence intervals for real per-capita GDP growth rates for one hundred and twenty-eight countries using post-war data. Consistent with the analysis in Section 2, we find large differences between the Cochrane-Orcutt and other estimators for many of the countries. There are smaller, but economically important differences in the other estimators, and this highlights the importance of estimator choice. Finally, for most countries, the high degree of serial correlation and short sample leads to wide confidence intervals for β.

Finally, we offer a summary and some conclusions in Section 5, and the appendix contains proofs and other detailed calculations.

2 Estimators

2.1 The Model

The statistical model for the observations $\{y_t\}_{t=1}^{T}$ is conveniently summarized in the following assumptions:

1. The data y_t are generated by

$$y_t = \alpha + \beta t + u_t \text{ for } t = 1, \ldots, T. \tag{3}$$

2. The error term u_t is generated by $(1 - \rho_T L)u_t = v_t$, for $t = 2, \ldots, T$.

3. $u_1 = \sum_{i=0}^{[\kappa T]} \rho_T^i v_{1-i}$.

4. $v_t = d(L)\epsilon_t$, with $d(L) = \sum_{i=0}^{\infty} d_i L^i$, and $\sum_{i=0}^{\infty} i \mid d_i \mid < \infty$.

5. The error term ϵ_t is a martingale difference sequence with $E(\epsilon_t^2 \mid \epsilon_{t-1}, \epsilon_{t-2}, \ldots) = 1$ and with $\sup_t E\epsilon_t^4 < \infty$.

Assumption (1) says that the data are generated as a linear trend plus noise; the parameter β is the average trend growth in the series and is the parameter of interest. Assumptions (2) and (3) are written to include both $I(0)$ and $I(1)$ processes. When $\rho_T = \rho$, with $|\rho| < 1$, then u_t is $I(0)$; while when $\rho_T = 1$, then u_t is $I(1)$. More generally, when $\rho_T = (1 + \frac{c}{T})$, then u_t follows a "local-to-unity" $I(1)$ process, with $c = 0$ corresponding to an exact unit root and values of $c \neq 0$ generating data that are less ($c < 0$) or more ($c > 0$) persistent then the exact unit root process.[4]

Assumption (3) incorporates a range of assumptions about the initial condition u_1, depending on the value of κ and ρ_T. For example, when $\kappa = 0$, then $u_1 = v_1$, so that the initial value is assumed to be an $O_p(1)$ random variable. When $\kappa > 0$, then u_1 is $O_p(T^{1/2})$ when u_t is $I(1)$, but is $O_p(1)$ when u_t is $I(0)$. When $\rho_T = \rho$, with $|\rho| < 1$ and $\kappa T \to \infty$, then u_1 is drawn from the unconditional distribution of u_t, and the process is covariance stationary.[5]

Assumption (5) implies that the functional central limit applies to the partial sums of ϵ_t, i.e., $T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} \epsilon_t \Rightarrow W(s)$, where $W(s)$ is a standard Wiener process.[6] Assumption (4) insures that the functional central limit theorem also applies the partial sums of v_t, specifically $T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} v_t \Rightarrow d(1)W(s)$.

2.2 Asymptotic Properties of Estimators

2.2.1 OLS, First-Difference and GLS Estimators

Let $\hat{\beta}_{OLS}$ denote the OLS estimator of β in (1), let $\hat{\beta}_{FD} = (T-1)^{-1} \sum_{t=2}^{T} \Delta y_t$ denote the first-difference estimator, and let $\hat{\beta}_{GLS}$ denote the infeasible GLS estimator that corrects for non-zero ρ_T. Specifically, $\hat{\beta}_{GLS}$ is the OLS estimator in the transformed regression

$$y_t - \rho_T y_{t-1} = (1 - \rho_T)\alpha + \beta[t - \rho_T(t-1)] + u_t - \rho_T u_{t-1}, \quad t = 2, 3, \ldots, T. \tag{4}$$

together with

$$\sigma^{-1} y_1 = \sigma^{-1}\alpha + \sigma^{-1}\beta + \sigma^{-1} u_1, \tag{5}$$

where $\sigma^2 = (1 - \rho_T^{2([\kappa T]+1)})/(1 - \rho_T^2)$ for $\rho_T \neq 1$ and $\sigma^2 = [\kappa T] + 1$ for $\rho_T = 1$ For simplicity, the GLS estimator ignores the $I(0)$ serial correlation associated with $d(L)$.

[4]These "local-to-unity" processes have been used extensively to study local power properties of unit root tests, construct confidence intervals for autoregressive parameters for highly persistent processes, and more generally, to study the behavior of statistics whose distribution depends on the persistence properties of the data. Some notable examples are Bobkoski (1983), Cavanagh (1985), Cavanagh, Elliot and Stock (1993), Chan and Wei (1987), Chan (1988), Phillips (1987), and Stock (1991).

[5]See Elliott (1993) for related discussion of the initial error in the I(1) model.

[6]A range of alternative assumptions will also suffice; see Phillips and Solo (1992) for discussion.

This allows us to focus on the major source of serial correlation, $\rho_T \neq 0$, and leads to no loss of asymptotic efficiency for the models considered here (Grenander and Rosenblatt (1957)).

In large samples, the behavior of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$ and $\hat{\beta}_{GLS}$ is summarized in Theorems 1 and 2:

Theorem 1 *(Behavior of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$ and $\hat{\beta}_{GLS}$ with $I(0)$ Errors):*
Under assumptions (1)-(5) with $\rho_T = \rho$, and $|\rho| < 1$:
(a) $T^{\frac{1}{2}}(\hat{\beta}_{OLS} - \beta) \xrightarrow{L} N(0, V_1)$, *where* $V_1 = 12(1-\rho)^{-2}d(1)^2$.
(b) $T(\hat{\beta}_{FD} - \beta)$ *converges in distribution to a random variable with zero mean, variance* $V_2 = \sum_{i=0}^{\infty} f_i^2 + var(u_1)$, *where* $f_i = \sum_{j=0}^{i} \rho^{(i-j)} d_j$. *The limiting distribution of* $T(\hat{\beta}_{FD} - \beta)$ *depends on the distribution of the ϵ's, and so in general is non-normal.*
(c) $T^{\frac{1}{2}}(\hat{\beta}_{GLS} - \beta) \xrightarrow{L} N(0, V_1)$, *where V_1 is specified in (a).*

Proof. *Part (a) and (c) follow from a straightforward application of the central limit theorem. To show part (b), note that* $T(\hat{\beta}_{FD} - \beta) = u_T - u_1$, *from which the result follows immediately.*$\square$

Theorem 2 *(Behavior of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$ and $\hat{\beta}_{GLS}$ with $I(1)$ Errors):*
Let $S_c(\tau) = (-2c)^{-1}(1 - e^{2\tau c})$. *Then under assumptions (1)-(5), with* $\rho_T = (1 + \frac{c}{T})$:
(a) $T^{\frac{1}{2}}(\hat{\beta}_{OLS} - \beta) \xrightarrow{L} N(0, R_1)$, *where*

$$R_1 = d(1)^2 c^{-5}[18(c-2)^2 e^{2c} + 72c(c-2)e^c + 12c^3 + 54c^2 + 72c - 72]$$

$$+ d(1)^2 144 S_c(\kappa)[\frac{ce^c + c - 2(e^c - 1)}{2c^2}]^2.$$

(b) $T^{\frac{1}{2}}(\hat{\beta}_{FD} - \beta) \xrightarrow{L} N(0, R_2)$, *where*

$$R_2 = d(1)^2[S_c(1) + (1 - e^c)^2 S_c(\kappa)].$$

(c) $T^{\frac{1}{2}}(\hat{\beta}_{GLS} - \beta) \xrightarrow{L} N(0, R_3)$, *where*

$$R_3 = d(1)^2[\frac{S_c(\kappa)c^2 + 1}{(S_c(\kappa)c^2 + 1)(1 - c + \frac{1}{3}c^2) - S_c(\kappa)(\frac{1}{2}c^2 - c)^2}]$$

Proof. *See Appendix.*

Corollary 3 *(Behavior of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$ and $\hat{\beta}_{GLS}$ when $\rho = 1$):*
Under assumptions (1)-(5), with $\rho_T = 1$:
(a) $T^{\frac{1}{2}}(\hat{\beta}_{OLS} - \beta) \xrightarrow{L} N(0, \frac{6}{5}d(1)^2)$.
(b) $T^{\frac{1}{2}}(\hat{\beta}_{FD} - \beta) \xrightarrow{L} N(0, d(1)^2)$.
(c) $T^{\frac{1}{2}}(\hat{\beta}_{GLS} - \beta) \xrightarrow{L} N(0, d(1)^2)$.

We highlight five features of these results. First, $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$ and $\hat{\beta}_{GLS}$ converge to β faster in the $I(0)$ model than the $I(1)$ model. This results obtains because the variance of the errors is bounded in the $I(0)$ model and increases linearly with t in the $I(1)$ model. Sampson (1991) discusses the implication of this result for long-run forecast confidence intervals.

Second, the averaging in $\hat{\beta}_{OLS}$ in the $I(0)$ and $I(1)$ cases and in $\hat{\beta}_{FD}$ in the $I(1)$ case leads to asymptotically normal estimators. In contrast, since $T(\hat{\beta}_{FD} - \beta) = [T/(T-1)](u_T - u_1)$, no such averaging occurs for $\hat{\beta}_{FD}$ in the $I(0)$ case, so that $\hat{\beta}_{FD}$ is not asymptotically normally distributed in general. (See Quah and Wooldridge (1988) and Schmidt (1993) for related discussion.)

Third, $\hat{\beta}_{GLS}$ is the asymptotically efficient estimator regardless of the value of ρ and it corresponds to the BLUE estimator when $d(L) = d$, a constant. The efficiency of the FD and the OLS estimator relative to the GLS estimator differs dramatically in the $I(0)$ and $I(1)$ cases. When the errors are $I(0)$, then $\hat{\beta}_{FD}$ converges to β more slowly than does $\hat{\beta}_{GLS}$, and thus has an asymptotic relative efficiency of 0. In this case, $\hat{\beta}_{OLS}$ is asymptotically efficient, the familiar result from Grenander and Rosenblatt (1957). When the errors are $I(1)$, $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$ and $\hat{\beta}_{GLS}$ converge at the same rate and the relative efficiency depends on the parameters c and κ. Figure 1 plots the asymptotic relative efficiencies (defined as the ratio of the asymptotic variances of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FD}$ to the asymptotic variance of $\hat{\beta}_{GLS}$) in the $I(1)$ model for a range of values of c and κ. When $c = 0$, both $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FD}$ are invariant to u_1 and so their variances and the relative efficiency do not depend on κ. In this case $\hat{\beta}_{FD}$ is asymptotically efficient and $\hat{\beta}_{OLS}$ has an efficiency of 5/6. This result is derived in Durlauf and Phillips (1988), who study the properties of trend estimators in the model with $\rho = 1$ (equivalently, $c = 0$). When c is sufficiently negative, $\hat{\beta}_{OLS}$ dominates $\hat{\beta}_{FD}$ for all values of κ. The intersection point of the $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FD}$ relative efficiency curves depends on κ. For example, when $\kappa = 0$, $\hat{\beta}_{FD}$ is efficient relative to $\hat{\beta}_{OLS}$ for values of $-18.6 \le c \le 1.2$, and $\hat{\beta}_{OLS}$ dominates $\hat{\beta}_{FD}$ for c outside this range. When $\kappa = 1.0$, the range narrows to $-7.6 \le c \le 0.9$.

Fourth, when $\kappa = 0$, so that u_1 is $O_p(1)$, the relative efficiency of both $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FD}$ increases monotonically with c. The relatively poor performance of these estimators when u_1 is $O_p(1)$ has been noted elsewhere, notably by Elliott, Rothenberg and Stock (1992) in the context of unit root tests. On the other hand, when $\kappa > 0$, so that u_1 is $O_p(T^{\frac{1}{2}})$, the relative efficiency of $\hat{\beta}_{OLS}$ is U-shaped, with a minimum that depends on the specific value of κ. For example, when $\kappa = 1$, the minimum relative efficiency of $\hat{\beta}_{OLS}$ occurs at $c = -3.006$ where it takes on the value of 0.7535. As $\kappa \to \infty$, the minimum relative efficiency of $\hat{\beta}_{OLS}$ is .7538 and occurs at $c = -3.076$, a result that was also derived by Chipman (1979) using methods different from those employed here.[7]

[7]Chipman (1979) also shows that, when $d(L) = d$, this asymptotic relative efficiency value is the greatest lower bound for the relative efficiency of $\hat{\beta}_{OLS}$ for all $n \ge 2$. Because of a slight numerical error in Chipman's paper, his reported numerical results are different from those reported here.

Finally, when the errors are $I(1)$, the variances of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$, and $\hat{\beta}_{GLS}$ depend on c and κ in important ways. For example, Figure 2 plots the variance of $\hat{\beta}_{GLS}$ as a function of c and κ. As c increases, the persistence of the errors increases and so does the associated variance of $\hat{\beta}_{GLS}$. Similarly, as κ increases, the variance of u_1 increases, leading to an increased variance in $\hat{\beta}_{GLS}$.

2.2.2 Feasible GLS Estimators

The efficient GLS estimator relies on two parameters, ρ and κ, whose values are typically unknown. In this section we analyze feasible analogues of $\hat{\beta}_{GLS}$. The parameter ρ is easily estimated from the data, and as we show below, replacing ρ with an estimate has little effect on $\hat{\beta}_{GLS}$. On the other hand, it is impossible to construct accurate estimates of κ, since this parameter only affects the data through the variance of the single observation u_1. We therefore analyze three feasible GLS estimators that differ in their treatment of the initial observation. We find large differences in the relative performance of these estimators across different values of κ.

To focus attention on the parameter κ, we begin by analyzing the estimators assuming that ρ is known; a simple modification of these results yields the results for unknown ρ. As above, the GLS estimators ignore the serial correlation associated with the $I(0)$ dynamics in $d(L)$, since the Grenander-Rosenblatt (1957) results imply that OLS or GLS treatment of $d(L)$ has no asymptotic effect on the estimators of β that we consider. Let $\hat{\beta}_{CO}$ denote the Cochrane-Orcutt (1949) GLS estimator that ignores the levels information in the first observation; that is, $\hat{\beta}_{CO}$ denotes the OLS estimator of β in equation (4). Let $\hat{\beta}_{CC}$ denote the GLS estimator constructed under the assumption that $u_0 = 0$. This assumption is often made in the unit root literature (see, e.g., Elliott, Rothenberg and Stock (1992)) and is referred to as the "conditional case." Thus, $\hat{\beta}_{CC}$ is the OLS estimator of β from (4) together with:

$$y_1 = \alpha + \beta + u_1. \tag{6}$$

Finally, let $\hat{\beta}_{PW}$ denote the Prais-Winsten (1954) estimator; that is, the OLS estimator of β from (4) together with:

$$(1-\rho_T^2)^{1/2}y_1 = (1-\rho_T^2)^{1/2}\alpha + (1-\rho_T^2)^{1/2}\beta + (1-\rho_T^2)^{1/2}u_1. \tag{7}$$

The Prais-Winsten estimator is defined for $\rho_T \leq 1$, and we limit our discussion to this situation. In the notation introduced in the last section, $\hat{\beta}_{CC}$ corresponds to the GLS estimator constructed using $\kappa = 0$, and $\hat{\beta}_{PW}$ is the limiting value of the GLS estimator as $\kappa \to \infty$.

When $\rho_T = \rho$, with $|\rho| < 1$ (*i.e.*, u_t is $I(0)$), each of the GLS estimators is asymptotically efficient and the large sample distribution is given in Theorem 1.

(Specifically, the value of c that we report ($c = -3.07558$) is a more accurate estimate of the root to his polynomial (3.3) than the value reported in his paper ($c = -3.09485$).)

Thus, we need only consider the behavior of the estimators in the $I(1)$ model, and this is done in the following lemma:

Lemma 4 *(Behavior of GLS Estimators with $I(1)$ Errors):*
Under assumptions (1)-(5), with $\rho_T = (1 + \frac{c}{T})$:
(a) $T^{1/2}(\hat{\beta}_{CO} - \beta) \xrightarrow{L} N(0, G_1)$, *where*

$$G_1 = \frac{12d(1)^2}{c^2}, \text{ for } c \neq 0, \text{ and}$$

$$G_1 = d(1)^2, \text{ for } c = 0.$$

(b) $T^{1/2}(\hat{\beta}_{CC} - \beta) \xrightarrow{L} N(0, G_2)$, *where*

$$G_2 = \frac{d(1)^2}{1 - c + \frac{1}{3}c^2}[1 + S_c(\kappa)\frac{(c - \frac{1}{2}c^2)^2}{1 - c + \frac{1}{3}c^2}].$$

(c) $T^{1/2}(\hat{\beta}_{PW} - \beta) \xrightarrow{L} N(0, G_3)$, *where*

$$G_3 = \frac{d(1)^2}{(1 - \frac{1}{2}c + \frac{1}{12}c^2)^2}[c^2 S_c(\kappa) + 1 + \frac{1}{12}c^2].$$

Proof. See Appendix.

Part (a) of the lemma implies that G_1, the limiting variance of $T^{\frac{1}{2}}(\hat{\beta}_{CO} - \beta)$, is discontinuous at $c = 0$. This occurs because the regression constant term, α, becomes unidentified as $c \to 0$. For values of c close to zero, α is very poorly estimated, and the collinearity between the two regressors $(1, t)$ in equation (4) means that $\hat{\beta}_{CO}$ is also a poor estimate of β. When $c = 0$, α disappears from equation (4) and so this source of variance disappears from $\hat{\beta}_{CO}$. Figure 3 shows the efficiency of each of the estimators relative to $\hat{\beta}_{GLS}$. The Cochrane-Orcutt estimator, $\hat{\beta}_{CO}$, performs very poorly for small values of c regardless of the value of κ. This result is consistent with a large literature on the poor performance of the Cochrane-Orcutt estimator with trending regressors and ρ close to unity.[8]

The relative performance of the other two estimators depends on the values of κ and c. When $\kappa = 0$, $\hat{\beta}_{CC}$ is the asymptotically efficient estimator; while $\hat{\beta}_{PW}$ is the efficient estimator as $\kappa \to \infty$. From Figure 3, $\hat{\beta}_{PW}$ is approximately efficient even when κ is very small. For example, for $\kappa = .01$ the relative efficiency of $\hat{\beta}_{PW}$ is larger than 0.73 for all values of c; for $\kappa = .05$ the relative efficiency is larger than .92; and for all values of $\kappa \geq .10$, $\hat{\beta}_{PW}$ is essentially efficient. While $\hat{\beta}_{CC}$ is efficient when $\kappa = 0$, this efficiency gain disappears quickly for moderate values of c as κ increases.

[8] See Prais and Winsten (1954), Maeshiro (1976,1978), Beach and MacKinnon (1978), Park and Mitchell (1980), Thornton (1987) and Davidson and MacKinnon (1993, Section 10.6).

We are now ready to discuss the feasible GLS estimators with ρ_T unknown. These estimators are calculated like their infeasible counterparts, using an estimator of ρ_T in equations (4) and (7). These estimators will be denoted as $\widehat{\beta}_{FCO}$, $\widehat{\beta}_{FCC}$, and $\widehat{\beta}_{FPW}$. Analysis of these estimators is complicated by the fact that they implicitly depend on the estimator for ρ_T, and a variety of estimators of ρ_T have been suggested. For $\widehat{\beta}_{FCO}$ the non-linear least squares estimator is often employed, and this estimator is studied by Nagaraj and Fuller (1991) for the model with general regressors. Their analysis can be simplified here because of the special structure of the regressors: equation (4) together with assumption (2) can be combined as:

$$y_t = a + bt + \rho_T y_{t-1} + v_t, \text{ for } t = 2, 3, \ldots, T, \tag{8}$$

where $a = \alpha(1-\rho_T) + \beta\rho_T$ and $b = \beta(1-\rho_T)$. Thus, $\widehat{\beta}_{FCO}$ can be formed from the OLS estimators from equation (8) as $\widehat{\beta}_{FCO} = \widehat{b}/(1-\widehat{\rho}_T)$ for $\widehat{\rho}_T \neq 0$ and $\widehat{\beta}_{FCO} = \widehat{a}$ for $\widehat{\rho}_T = 1$, where $\widehat{a}$, $\widehat{b}$, and $\widehat{\rho}_T$ are the OLS estimators of the coefficients in equation (8). Equivalently, $\widehat{\beta}_{FCO}$ can be constructed as the OLS estimator of β in (4) using $\widehat{\rho}_T$ in place of ρ_T. Since the asymptotic distribution $T(1-\widehat{\rho}_T)$ is readily deduced when $\rho_T = (1+\frac{c}{T})$, (see Stock (1991), for example), the asymptotic distribution of $T^{\frac{1}{2}}(\widehat{\beta}_{FCO} - \beta)$ can also be readily deduced.

The problem is more complicated when analyzing $\widehat{\beta}_{FCC}$ and $\widehat{\beta}_{FPW}$, since these estimators are generally based on iterative schemes for estimating ρ_T, α, and β. Iterative schemes are often used to construct $\widehat{\beta}_{FCO}$ as well. Since the limiting distribution of $\widehat{\rho}_T$ depends in important ways on the precise way the data are "detrended" (for example, see Schmidt and Phillips (1992) and Elliott, Rothenberg and Stock (1992)), the limiting distribution of $\widehat{\beta}_{FCC}$, and $\widehat{\beta}_{FPW}$ will depend on the precise specification of the iterations. Rather than present results for specific versions of these estimators, we present limiting representations of $\widehat{\beta}_{FCC}$, and $\widehat{\beta}_{FPW}$ written as functions of $\widehat{c} = \operatorname{plim} T(1-\widehat{\rho}_T)$. Different estimators of ρ_T will lead to different limiting random variables $\widehat{c}$ and different asymptotic distributions for the estimator of β. A specific example is contained in Durlauf and Phillips (1988, Theorem 4.1), who derive the limiting distribution of $\widehat{\beta}_{FCO}$ when $c = 0$ and $\widehat{c}$ is constructed from the Durbin-Watson statistic calculated from the levels OLS regression.

Before presenting the limiting distributions for the feasible GLS estimators, it is useful to introduce some additional notation. The error term in the feasible GLS version of (4) is $\widehat{v}_t = u_t - \widehat{\rho}_T u_{t-1}$, and the limiting values of the feasible GLS estimators can be written in terms of initial condition u_1 and partial sums of $\widehat{v}_t$. In the appendix we show that $T^{-\frac{1}{2}} u_1 \Rightarrow \widetilde{W}_c(\kappa) \sim N(0, S_c(\kappa))$, where $S_c(\kappa)$ is defined in Theorem 2; we also show that $T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} \widehat{v}_t \Rightarrow \widehat{W}(s)$ where $\widehat{W}(s)$ is a functional of $W(s)$ and $\widetilde{W}_c(\kappa)$.

With this notation established, we now present the limiting distribution of the feasible GLS estimators:

Theorem 5 *(Behavior of Feasible GLS Estimators):*
Suppose that assumptions (1)-(5) are satisfied, $\rho_T = (1+\frac{c}{T})$, *and* $\text{plim}(\hat{\rho}_T - 1) = \hat{c} \neq 0$. *Then:*
(a) $T^{\frac{1}{2}}(\hat{\beta}_{FCO} - \beta) \Rightarrow \hat{c}^{-1} 12 \int_0^1 (\frac{1}{2} - s) d\widehat{W}(s)$,
(b) $T^{\frac{1}{2}}(\hat{\beta}_{FCC} - \beta) \Rightarrow [1 - \hat{c} + \frac{1}{3}\hat{c}^2]^{-1}[(\hat{c} - \frac{1}{2}\hat{c}^2)\widetilde{W}_c(\kappa) - \int_0^1 (\hat{c}s - 1) d\widehat{W}(s)]$, *and*
(c) $T^{\frac{1}{2}}(\hat{\beta}_{FPW} - \beta) \Rightarrow [1 - \frac{1}{2}\hat{c} + \frac{1}{12}\hat{c}^2]^{-1}[\hat{c}\widetilde{W}_c(\kappa) - \int_0^1 (1 + \frac{1}{2}\hat{c} - \hat{c}s) d\widehat{W}(s)]$.

Proof. *See Appendix.*

This theorem allows us to offer practical advice about choice of estimators. First, notice that $\hat{c}$ appears in the denominator of the limiting representation of $T^{\frac{1}{2}}(\hat{\beta}_{FCO} - \beta)$. For most commonly used estimators of ρ, $\hat{\rho}_T$ can take on values arbitrarily close to 1 with positive probability, so that $\hat{c}$ can be very close to zero. This means that $\hat{\beta}_{FCO}$ can be very badly behaved, since realizations of $\hat{c}$ close to zero will often lead to extreme realizations of $\hat{\beta}_{FCO}$. On the other hand, $\hat{\beta}_{FCC}$ and $\hat{\beta}_{FPW}$ are better behaved, since $[1 - \hat{c} + \frac{1}{3}\hat{c}^2] > 0$ and $[1 - \frac{1}{2}\hat{c} + \frac{1}{12}\hat{c}^2] > 0$ for all values of $\hat{c}$. This can be seen in Figure 4 which plots the limiting probability densities of $T^{\frac{1}{2}}(\hat{\beta}_{FCO} - \beta), T^{\frac{1}{2}}(\hat{\beta}_{FCC} - \beta)$ and $T^{\frac{1}{2}}(\hat{\beta}_{FPW} - \beta)$, for the case with $c = 0$, $\kappa = 1$, and $d(1) = 1$.[9] Also plotted is the probability density of the exact (infeasible) GLS estimator (which in this case is the standard normal). The estimators $\hat{\beta}_{FCC}$, and $\hat{\beta}_{FPW}$ have probability distributions very close to the infeasible efficient estimator. On the other hand, the distribution of $\hat{\beta}_{FCO}$ is much more disperse, with thicker tails than the other distributions. For example, the limiting probability that $|T^{\frac{1}{2}}(\hat{\beta}_{FCO} - \beta)|$ exceeds 2 is approximately 20%; while the corresponding values for $\hat{\beta}_{FCC}$ and $\hat{\beta}_{FPW}$ are approximately 5%. Figure 4 suggests that little is lost using in using either $\hat{\beta}_{FCC}$ and $\hat{\beta}_{FPW}$ in place of the infeasible efficient estimator, at least for this value of c and κ, and that $\hat{\beta}_{FCO}$ performs poorly. Additional calculations (not shown) indicate that the relative efficiencies of $\hat{\beta}_{FCC}$ and $\hat{\beta}_{FPW}$ are close to their infeasible analogues for a wide range of values of c and κ.

Table 1 summarizes many of the results in this section by presenting the average mean squared error for the different feasible estimators and different values of κ, averaged over different ranges of c.[10] As a benchmark, the first row of the table shows results for the efficient, but infeasible, GLS estimator. The next two rows are the OLS and first-difference estimators, followed by two of the feasible GLS estimators. (Since the asymptotic mean squared error of $\hat{\beta}_{FCO}$ does not exist, this estimator is not included in the table.) The last row of the table shows results for a "pre-test"

[9] The densities for the feasible GLS estimators are estimates based on 5000 draws from approximations to the asymptotic distributions (constructed using T=500). The estimators $\hat{\beta}_{FCO}$ and $\hat{\beta}_{FCC}$ were constructed using $\hat{\rho}_T$ constructed as the OLS estimator of (8). The Prais-Winsten estimator used $\min(1, \hat{\rho}_T)$.

[10] These MSE's were estimated using the simulations described in footnote 9.

estimator ($\hat{\beta}_{PT}$) constructed from the OLS and FD estimator. Figure 1 provides the motivation for this estimator. Since the OLS estimator dominates the first-difference estimator for large negative values of c and is dominated by the first-difference estimator for small values of c, the pre-test estimator corresponds to the OLS estimator when $\hat{c}$ is large and negative and corresponds to the FD estimators when $\hat{c}$ is close to zero. Specifically, $\hat{\beta}_{PT} = \hat{\beta}_{OLS}$ when $\hat{c} < \bar{c}$ and $\hat{\beta}_{PT} = \hat{\beta}_{FD}$ when $\hat{c} \geq \bar{c}$, where $\bar{c}$ is pre-specified threshold. The results shown in the table are for $\bar{c} = -15$, a value that produced good results over the range of values of κ and c that we considered.

Table 1 and the figures shown above suggest five conclusions:

(i) The infeasible GLS estimator $\hat{\beta}_{CO}$ performs very poorly for values of c close to 0. This poor performance is inherited by the feasible GLS estimator. For all values of $c \neq 0$ and for all values of κ, this estimator is dominated by $\hat{\beta}_{OLS}$. Thus, this estimator should not be used and is ignored in the remaining discussion.

(ii) For very small values of c (say, $-2 \leq c \leq 0$), $\hat{\beta}_{FD}$ is the preferred estimator with a mean squared error approximately 5% lower than $\hat{\beta}_{FCC}$ and $\hat{\beta}_{FPW}$. For this range of values of c, the OLS estimator, $\hat{\beta}_{OLS}$, has a relative efficiency of approximately .75. The pre-test estimator performs well, and is 1%-2.5 % less efficient than $\hat{\beta}_{FD}$, depending on the value of κ.

(iii) For values of c in the range $-10 \leq c \leq -2$, the relative performance of the estimators depends critically on the value of the initial error, parameterized by κ. When $\kappa = 0$, $\hat{\beta}_{FCC}$ dominates the other estimators; $\hat{\beta}_{FPW}$ is the preferred estimator when $\kappa \geq .10$. When $\kappa = .05$ the feasible GLS estimators and $\hat{\beta}_{FD}$ are comparable.

(iv) For values of $-30 \leq c \leq -10$ and when $\kappa = 0$, $\hat{\beta}_{FCC}$ is the preferred estimator. When $\kappa \geq 0.05$, the variance of $\hat{\beta}_{FCC}$ is more than twice as large as the variance of the best estimator, $\hat{\beta}_{FPW}$. The first difference estimator also performs poorly relative to $\hat{\beta}_{FPW}$ when $\kappa \geq .05$.

(v) Items (ii)-(iv) show clearly that the best estimator depends on the values of c and κ. Neither of these parameters can be consistently estimated from the data, and so a good choice must depend on either prior knowledge or robustness considerations. Our reading of the results suggests that $\hat{\beta}_{FPW}$ is the most robust estimator, with a MSE close to the optimum for all values of the parameters considered. The pretest estimator is a reasonable alternative to $\hat{\beta}_{FPW}$; it has slightly better performance when c close to 0 but somewhat worse performance for large negative c.

2.3 Small Sample Properties of Estimators

The asymptotic results summarized in Theorems 1, 2 and 5 are potentially useful for two reasons. First, the asymptotic relative efficiencies can provide a criterion for choosing among the estimators even in finite samples. Second, the asymptotic distributions provide a basis for constructing confidence intervals and carrying out hypothesis tests. In this section we evaluate the first of these uses, and ask whether the $I(0)$ and $I(1)$ asymptotic variances provide a useful guide for choosing among the estimators in small samples. In the following section, we discuss confidence intervals and statistical inference.

Table 2 shows the exact relative efficiencies of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$, $\hat{\beta}_{FCC}$, $\hat{\beta}_{FPW}$, and $\hat{\beta}_{PT}$ for the model with $d(L) = d$, $\varepsilon_t \sim NIID(0,1)$, for various values of T, ρ, and for $\kappa = 0$ (panel A) and $\kappa = 1.0$ (panel B).[11] Also shown in the table are the relative efficiencies implied by the $I(1)$ asymptotics, calculated using $c = T(\rho - 1)$. The $I(0)$ asymptotic relative efficiencies are not shown because they do not vary with T, ρ or κ; from Theorem 1 they are 1.00 for $\hat{\beta}_{OLS}$, $\hat{\beta}_{FCC}$, $\hat{\beta}_{FPW}$, and $\hat{\beta}_{PT}$ and 0.00 for $\hat{\beta}_{FD}$. In all cases, the $I(0)$ asymptotic relative efficiency suggests indifference between the four estimators $\hat{\beta}_{OLS}$, $\hat{\beta}_{FCC}$, $\hat{\beta}_{FPW}$ and $\hat{\beta}_{PT}$, and suggests that these estimators are preferred to $\hat{\beta}_{FD}$.

When $\rho = 0.5$, the finite sample results in Table 2 suggest that $\hat{\beta}_{OLS}$, $\hat{\beta}_{FCC}$, and $\hat{\beta}_{FPW}$ are essentially efficient for all of the sample sizes considered. These estimators are significantly better than $\hat{\beta}_{FD}$. The pre-test estimator has a relative efficiency intermediate between $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FD}$ when $T = 30$, and very close to $\hat{\beta}_{OLS}$ for larger values of T. Thus the $I(0)$ relative efficiency predictions are quite accurate when $\rho = 0.5$. The predictions based on the $I(1)$ asymptotic relative efficiencies are off the mark. The $I(1)$ asymptotics suggests that $\hat{\beta}_{FCC}$ strongly dominates the other estimators when $\kappa = 0$ and is strongly dominated by both $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FPW}$ when $\kappa = 1$. On the other hand, the estimator with the largest $I(1)$ asymptotic relative efficiency coincides with the largest finite sample relative efficiency, even when $\rho = 0.5$.

For all of the other values of ρ that are considered (0.8, 0.9, 0.95, 1.0), the rankings implied by the $I(1)$ asymptotic relative efficiencies are more accurate the $I(0)$ rankings. Indeed in all cases studied in the tables, the estimator with the largest $I(1)$ asymptotic relative efficiency has the largest finite sample relative efficiency as well. Thus, this experiment suggests that the $I(1)$ asymptotic relative efficiencies provide a useful criterion for ranking estimators in typical econometric settings.

3 Confidence intervals

3.1 Construction of confidence intervals.

In this section we discuss methods for constructing confidence intervals for β. When $\rho < 1$ (so that the errors are $I(0)$) confidence intervals can be constructed in the usual way by inverting the "t-statistic" constructed from any of the asymptotically equivalent estimators $\hat{\beta}_{OLS}$, $\hat{\beta}_{FCO}$, $\hat{\beta}_{FCC}$, $\hat{\beta}_{FPW}$, or $\hat{\beta}_{PT}$. These t-statistics can be formed using an estimator for the variance V_1 in Theorem 1, constructed by replacing ρ and $d(1)$ with consistent estimators. While these confidence intervals are asymptotically

[11] The mean squared errors for $\hat{\beta}_{FCC}$, $\hat{\beta}_{FPW}$, and $\hat{\beta}_{PT}$, were estimated using 10,000 Monte Carlo draws, using $\hat{\rho} = \sum_{t=2}^{T} \hat{u}_t\hat{u}_{t-1} / \sum_{t=2}^{T-1} \hat{u}_t^2$, where $\hat{u}_t$ are the OLS residuals from the regression of y_t onto $(1, t)$. This estimator of ρ is suggested by the simulation results in Park and Mitchell (1980).

valid, they can greatly understate the uncertainty about β when ρ is large and the sample size is small. (See Park and Mitchell (1980) for simulation evidence.) Thus, in most situations of practical interest, confidence intervals based on $I(0)$ approximations are not satisfactory.

An alternative method pursued here is to construct confidence intervals using approximations based on $I(1)$ asymptotics. As we show below, this method yields confidence intervals with coverage rates closer to the nominal size than the $I(0)$ approximations. Unfortunately, the method is also more complicated. The complication arises because in the $I(1)$ model, the asymptotic distribution of the various estimators of β depends on the nuisance parameters c and κ, and these parameters cannot be consistently estimated from the data. Thus, the variances of the estimators cannot be consistently estimated, so that t-statistics will not have the appropriate limiting standard normal distribution. While this problem cannot be circumvented entirely, it is possible to construct asymptotically conservative confidence intervals following the procedures developed by Dufour (1990) and Cavanagh, Elliott and Stock (1993).[12]

Specifically, let $B_\kappa(c)$ denote a $100(1-\alpha_1)\%$ confidence interval for β constructed conditional on a specific value of c and κ. Similarly, let C_κ denote a $100(1-\alpha_2)\%$ confidence interval for c conditional on κ. Assume that $0 \leq \kappa \leq \bar{\kappa}$, where $\bar{\kappa}$ is pre-specified constant. Then the Bonferoni confidence interval, $\bigcup_{0\leq\kappa\leq\bar{\kappa}} \bigcup_{c\in C_\kappa} B_\kappa(c)$, is a conservative $100(1-\alpha_1-\alpha_2)\%$ confidence interval for β.

This confidence interval requires the conditional confidence interval for β, $B_\kappa(c)$, and the marginal confidence interval for c, denoted C_κ. Since $B_\kappa(c)$ conditions on the nuisance parameters c and κ, an asymptotically valid approximation can be constructed using any of the estimators $\widehat{\beta}_{OLS}$, $\widehat{\beta}_{FD}$, $\widehat{\beta}_{CC}$, or $\widehat{\beta}_{PW}$, and their asymptotic variances given in Theorem 1 and Lemma 4. (These variances require $d(1)$, which can be consistently estimated using standard spectral estimators.) The marginal confidence intervals for c, C_κ, can be constructed using the methods developed in Stock(1991).[13]

[12]Dufour (1990) considers the problem of statistical inference in the regression model with Gaussian AR(1) disturbances. He develops "bounds" tests and associated confidence intervals based on exact distributions. Cavanagh, Elliott and Stock (1993) consider testing for Granger-Causality in a regression with a highly serially correlated regressor modeled as a local-to-unity process. They develop bounds tests and associated confidence intervals based on asymptotic distributions.

[13]Stock (1991) considers the case with $\kappa = 0$ and, using our notation, develops methods for constructing confidence sets C_0. However, it is easy to modify his analysis for $\kappa > 0$. Specifically, following Stock, we construct confidence intervals by inverting the Dickey-Fuller t-statistic, $\widehat{\tau}^\tau$. Under the assumption that $\kappa = 0$, Stock shows $\widehat{\tau}^\tau \Rightarrow (\int_0^1 W_c^\tau(s)^2 ds)^{\frac{1}{2}}[c + \int_0^1 W_c^\tau(s)dW(s)/(\int_0^1 W_c^\tau(s)^2 ds)]$, where $W_c^\tau(s)$ is the "detrended" diffusion: $W_c^\tau(s) = W_c(s) - \int_0^1 a_1(r)W_c(r)dr - s\int_0^1 a_2(r)W_c(r)dr$, where the diffusion $W_c(s)$ is defined in the appendix, $a_1 = 4 - 6r$, and $a_2 = -6 + 12r$. These results rely on the fact that $T^{-\frac{1}{2}}u_{[sT]} \Rightarrow d(1)W_c(s)$ when $\kappa = 0$. As shown in the appendix, when $\kappa \neq 0$, $T^{-\frac{1}{2}}u_{[sT]} \Rightarrow d(1)[W_c(s) + e^{sc}\widetilde{W}_c(\kappa)]$, where $\widetilde{W}_c(\kappa) \sim N(0, S_c(\kappa))$ and is independent of $W_c(s)$. Using this, it is straightforward to show that all of Stock's analysis continues to hold, with $W_c(s) + e^{sc}\widetilde{W}_c(\kappa)$ replacing $W_c(s)$ in the above limiting representation for $\widehat{\tau}^\tau$.

In general, this procedure is quite demanding. For each $0 \leq \kappa \leq \bar{\kappa}$, C_κ must be formed, then $B_\kappa(c)$ must be constructed for all $c \in C_\kappa$, and the union taken over all of these confidence sets. There are three special features of the linear trend model that simplify this procedure. First, from Theorem 2, the asymptotic variances of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FD}$ are monotonically increasing in c. Thus, when $B_\kappa(c)$ are formed using t-statistics constructed from $\hat{\beta}_{OLS}$ or $\hat{\beta}_{FD}$, then $\bigcup_{c \in C_\kappa} B_\kappa(c) = B_\kappa(\bar{c})$, where $\bar{c} = \sup_c\{c \in C_\kappa\}$. While this simplification does not necessarily hold for the GLS estimators $\hat{\beta}_{CC}$ and $\hat{\beta}_{PW}$, experiments that we have performed suggest that $\bigcup_{c \in C_\kappa} B_\kappa(c) \approx B_\kappa(\bar{c})$ appears to be a good approximation for confidence sets constructed from these estimators as well. The second simplifying feature is that the distributions of the statistics used to form C_κ change little as κ changes, so that $C_0 \approx C_\kappa$ for all κ.[14] Finally, for all of the estimators, the asymptotic variance is increasing in κ and the limit exists as $\kappa \to \infty$, so that $B_\kappa(c) \subseteq B_\infty(c)$ for all κ. Putting these three results together implies that $\bigcup_{0 \leq \kappa \leq \bar{\kappa}} \bigcup_{c \in C_\kappa} B_\kappa(c) \approx B_\infty(\bar{c})$, where $\bar{c} = \sup_c\{c \in C_0\}$. Thus approximate $100(1-\alpha_1-\alpha_2)\%$ confidence intervals can be formed by (i) choosing the largest value of c in the $100(1-\alpha_2)\%$ confidence interval constructed using the procedure from Stock (1992), and (ii) constructing a $100(1-\alpha_1)\%$ confidence interval for β using this value of c together with $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$, $\hat{\beta}_{CC}$, or $\hat{\beta}_{PW}$ and an associated variance from Theorem 2 or Lemma 4 evaluated at $\kappa = \infty$.

We make two final points before evaluating the small sample properties of this procedure. First, since the variance of all of the estimators is increasing in c, smaller confidence intervals for β can be obtained by constructing 1-sided confidence intervals for c. Second, when the $B_\kappa(c)$ confidence intervals are constructed by inverting the t-statistics for the estimators, the widths of the intervals will be non-random conditional on c and κ. This implies that the narrowest of the confidence intervals (across all estimators) will also have coverage rate exceeding $100(1-\alpha_1-\alpha_2)\%$. Thus, for example, since $\hat{\beta}_{OLS}$ is efficient relative to $\hat{\beta}_{FD}$ when $c < -7.6$ and κ is large, the confidence interval can be constructed using $\hat{\beta}_{OLS}$ when $\bar{c} < -7.6$ and using $\hat{\beta}_{FD}$ when $\bar{c} \geq -7.6$.

3.2 Small sample performance of confidence intervals

Table 3 shows estimated coverage rates for confidence intervals for different values of T and c, calculated as described above. In panel A, the confidence intervals are calculated as the narrowest of the OLS and FD confidence intervals. Panel B shows results for confidence intervals constructed from the Prais-Winsten estimator. The design was much the same as in Section 2.3, i.e., $d(L) = d$ and $\epsilon_t \sim N(0,1)$. Results

[14] When $c = 0$ the distribution of $\hat{\tau}^\tau$ is invariant to κ. This is not strictly true for other values of c, but the distribution changes very little. For example, when $c = -1.0$ the 97.5 percentiles for $\hat{\tau}^\tau$ are -3.72, -3.70, -3.70 and -3.70 when $\kappa = 0.0$, 0.5, 1.0, and 10.0, respectively. The corresponding percentiles are -3.89, -3.84, -3.84, -3.84 for $c = -5.0$; -4.20, 4.20, 4.20, -4.20 for $c = -10.0$; and 4.52, -4.54, -4.54, -4.54 for $c = -20.0$. These percentiles are based on 5,000 simulations with $T = 500$.

are reported for conservative 90%, 95% and 99% confidence intervals constructed with $\alpha_1 = \alpha_2$. Results for non-symmetric α_1 and α_2 are similar and are not reported. The confidence interval for ρ was constructed from the $\hat{\tau}^\tau$ statistic constructed from the regression of y_t onto Δy_{t-1} and $(1, t)$ using the sample $t = 2, \ldots, T$. The sample residual variance from this regression was used as the estimator of $d(1)^2$ in the construction of the confidence intervals for β. Finally, since the Prais-Winsten estimator is defined for $|\rho| \leq 1$ we restricted the upper limit of the confidence interval to $\rho = 1$. For comparability, this restriction was also used in the $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FD}$ confidence intervals.

The coverage rates are close to their nominal level for $c = 0$. When $c < 0$, the confidence intervals are conservative, with coverage rates exceeding the nominal level. This occurs because of the sharp increase in the variance of estimators for small c. So for example, when the true value of $c = -5$, then $c = 0$ is often in the confidence set C_0, the variance of the estimators is much larger when $c = 0$ than when $c = -5$ (see Figure 2) and this leads to a wide confidence interval for β.

4 Economic Growth Rates for the Postwar Period

Table 4 shows estimated annual growth rates of real GDP per capita for 128 countries over the postwar period. The data are annual observations from the Penn World Table (version 5.5) described in Summers and Heston (1991) (series RGDPCH). The data set contains 150 countries, and we limited our analysis to those 128 countries with 20 or more annual observations. The first column of the table shows the country identification number from the Penn World tables, and the next column shows the country name. Columns 3-6 present four estimates of average trend growth ($\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$, $\hat{\beta}_{FCO}$, and $\hat{\beta}_{FPW}$, respectively); column 5 shows the estimate of c used to construct the feasible GLS estimates ($\hat{c}$); column 6 shows the Dickey-Fuller unit-root test statistic ($\hat{\tau}^\tau$) used to construct a confidence interval for c, and columns 7 and 8 present lower and upper limits of the approximated 95% confidence interval for β constructed from the $\hat{\beta}_{PW}$ ($\beta_{\min}$ and $\beta_{\max}$, respectively). The estimate $\hat{c}$ was calculated as explained in footnote 11. The $\hat{\tau}^\tau$ statistic was calculated from the regression of Δy_t onto y_{t-1}, Δy_{t-1} and $(1, t)$ using data from $t = 3, \ldots T$, and the point estimates from this regression were used to estimate $d(1)$. We highlight five features of the results.

First, for the majority of the countries, the different estimators give similar results. For example, for the Congo (country 12) the estimates range from 2.8% ($\hat{\beta}_{FD}$) to 3.4% ($\hat{\beta}_{FCO}$). Second, while the $\hat{\beta}_{FCO}$ estimates are usually similar to the other estimates, they occasionally deviate substantially. For example, the estimates for Suriname (country 81) constructed from $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$,and $\hat{\beta}_{FPW}$ range from 0.4% to 1.4%, while the estimate constructed from $\hat{\beta}_{FCO}$ is -212%. Indeed for 31 of the 128 countries, $\hat{\beta}_{FCO}$ differs from $\hat{\beta}_{OLS}$ by more than 5 percentage points. Third, while the differences in the other three estimators are much smaller, these differences can be quantitatively important. For example, $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$ differ by more than 1% in 5 cases and by more than $\frac{1}{2}$% in 35 cases.

Fourth, the confidence intervals are often wide and include negative values for β. This results from three factors: a small sample size, a large error variance and a high degree of persistence in the annual growth rates. For example, the approximate 95% confidence interval for Algeria (country 1) is $-1.18 \leq \beta \leq 4.10$. For Algeria, the Dickey-Fuller t-statistic is -1.46 which implies that $c = 0$ (i.e., $\rho = 1$) is contained in the 97.5% confidence interval for c. Thus, for this value of c, $\hat{\beta}_{PW}$ corresponds to the first-difference estimator. The mean growth rate for Algeria over the sample period is 1.45% ($= \hat{\beta}_{FD}$) and this is the center of the confidence interval. The standard deviation of the annual growth rates is 7.3%; thus, if the annual growth rates were serially uncorrelated, the standard deviation of the sample mean ($= \hat{\beta}_{FD} = \hat{\beta}_{PW}$) would be 1.33% ($= 7.3\%/\sqrt{30}$). For Algeria, the growth rates are slightly negatively correlated and the estimated standard deviation of $\hat{\beta}_{PW}$ used to construct the confidence interval was 1.18%.

Finally, a few of the confidence intervals are quite narrower. For example, the estimated confidence interval for the UK (series 140) is $2.07 \leq \beta \leq 2.44$. This series is less persistent than most of the others, and the Dickey-Fuller t-statistic is -4.51. This leads to a confidence interval for c with an upper limit of $c = -14.1$ (corresponding to $\rho = 0.66$). From Figure 2, estimates of β are much more precise when $c = -14.1$ than when $c = 0$. Indeed the ratio of the asymptotic standard deviation for $\hat{\beta}_{PW}$ for $c = -14$ and $c = 0$ is 0.2, which approximately corresponds to the difference between the widths of the confidence intervals for β for the UK and the US (country 71).

5 Concluding Remarks

In this paper we study the problems of estimation and inference in the deterministic model. While the structure of the model is very simple, serial correlation in the errors can make efficient estimation and inference difficult. Asymptotic results are presented for $I(0)$ and local-to-unity $I(1)$ error processes, with the latter being the most relevant for econometric applications. The asymptotic distribution of the estimators is shown to depend on two important parameters: (i) the local-to-unity parameter that measures the persistence in the errors and (ii) a parameter that governs the variance of the initial error term.

Three conclusions emerge from our analysis. First, the Cochrane-Orcutt estimator is dominated by the other feasible estimators and should not be used. When the data are highly serially correlated (i.e., the local-to-unity parameter is close to zero), the distribution of the Cochrane-Orcutt estimator has very thick tails, and large outliers are common. Second, the feasible Prais-Winsten estimator is the most robust across the parameters governing persistence and initial variance. This is the preferred estimator unless the researcher has sharp *a priori* knowledge about these parameters.

Finally, inference that ignores uncertainty about ρ or the variance in the initial error term can be seriously flawed and lead to large biases in confidence intervals for trend growth rates. It is not clear how to optimally account for uncertainty in these parameters, but conservative confidence intervals and tests are easily constructed.

A Appendix: Theorem Proofs

A.1 Preliminaries:

From assumption (5), $T^{-\frac{1}{2}}\sum_{t=1}^{[sT]}\epsilon_t \Rightarrow W(s)$; in addition, this result, together with assumption (4) implies $T^{-\frac{1}{2}}\sum_{t=1}^{[sT]}v_t \Rightarrow d(1)W(s)$, where W(s) is a standard Weiner process. Analogously, accumulating the errors backwards from time 0, $T^{-\frac{1}{2}}\sum_{t=-[sT]}^{0}\epsilon_t \Rightarrow \widetilde{W}(s)$ and $T^{-\frac{1}{2}}\sum_{t=-[sT]}^{0}v_t \Rightarrow d(1)\widetilde{W}(s)$, where $\widetilde{W}(s)$ is a standard Weiner process, independent of W(s).

Let $\tilde{u}_t = \sum_{i=0}^{t-2}\rho_T^i v_{t-i}$ with $\rho_T = (1+\frac{c}{T})$. Then $T^{-\frac{1}{2}}\tilde{u}_{[sT]} \Rightarrow d(1)W_c(s)$, where $W_c(s)$ denotes the diffusion process generated by $dW_c(s) = cW_c(s)ds + dW(s)$. Similarly, $T^{-\frac{1}{2}}u_1 = T^{-\frac{1}{2}}\sum_{i=0}^{[\kappa T]}\rho_T^i v_{1-i} \Rightarrow d(1)\widetilde{W}_c(\kappa)$, where $\widetilde{W}_c(\kappa)$ denotes the diffusion process generated by $d\widetilde{W}_c(s) = c\widetilde{W}_c(s)ds + d\widetilde{W}(s)$. Note that $\widetilde{W}_c(\kappa) \sim N(0, S_c(\kappa))$, where $S_c(\kappa) = (-2c)^{-1}(1 - e^{2c\kappa})$. Finally, write $u_t = \tilde{u}_t + \rho_T^{t-1}u_1$, so that $T^{-\frac{1}{2}}u_{[sT]} \Rightarrow d(1)[W_c(s) + e^{sc}\widetilde{W}_c(\kappa)]$.

A.2 Proof of Theorem 2:

A.2.1 Proof of (a):

By direct calculation:

$$T^{\frac{1}{2}}(\hat{\beta}_{OLS} - \beta) = \frac{T^{-\frac{3}{2}}\sum_{t=1}^{T}\frac{t}{T}u_t - (T^{-1}\sum_{t=1}^{T}\frac{t}{T})(T^{-\frac{3}{2}}\sum_{t=1}^{T}u_t)}{T^{-1}\sum_{t=1}^{T}(\frac{t}{T})^2 - (T^{-1}\sum_{t=1}^{T}(\frac{t}{T}))^2}.$$

Thus,

$$T^{\frac{1}{2}}(\hat{\beta}_{OLS} - \beta) = 12T^{-\frac{3}{2}}\sum_{t=1}^{T}u_t(\frac{t}{T} - \frac{1}{2}) + o_p(1)$$

$$\Rightarrow d(1)12\int_0^1 (s - \frac{1}{2})[W_c(s) + e^{sc}\widetilde{W}_c(\kappa)]ds \sim N(0, R_1),$$

where $R_1 = A_1 + A_2$, with

$$A_1 = var\{d(1)12\int_0^1 (s - \frac{1}{2})W_c(s)ds\}$$

and

$$A_2 = var\{\widetilde{W}_c(\kappa)d(1)12\int_0^1 (s - \frac{1}{2})e^{sc}ds\}.$$

To calculate A_1, note that:

$$\int_0^1 (s - \frac{1}{2})W_c(s)ds = \int_0^1 (s - \frac{1}{2})\int_0^s e^{c(s-\tau)}dW(\tau)ds$$

$$= \int_0^1 \{\int_\tau^1 (s-\frac{1}{2})e^{cs}ds\}e^{-c\tau}dW(\tau)$$

$$= \int_0^1 b(\tau)dW(\tau), \text{ with } b(\tau) = \{\int_\tau^1 (s-\frac{1}{2})e^{cs}ds\}e^{-c\tau}.$$

Thus,

$$A_1 = 144d(1)^2 \int_0^1 b(s)^2 ds,$$

and

$$A_2 = 144d(1)^2 S_c(\kappa)[\int_0^1 (s-\frac{1}{2})e^{sc}ds]^2.$$

The first term in R_1 is A_1 after simplification, and the second term is A_2.

A.2.2 Proof of (b):

$$T^{\frac{1}{2}}(\hat{\beta}_{FD}-\beta) = T^{-\frac{1}{2}}u_T - T^{-\frac{1}{2}}u_1 = T^{-\frac{1}{2}}\tilde{u}_T - T^{-\frac{1}{2}}u_1(1-\rho_T^{T-1}) \Rightarrow$$
$$d(1)[W_c(1)-(1-e^c)\widetilde{W}_c(\kappa)] \sim N(0, d(1)^2[S_c(1)+(1-e^c)^2 S_c(\kappa)].$$

A.2.3 Proof of (c):

This GLS estimator is constructed by OLS applied to an equation of the form $y_t = x_t'\delta + e_t$, where $\delta = (\alpha \quad \beta)'$, $x_1 = (\sigma^{-1} \ \sigma^{-1})'$, $x_t = [(1-\rho_T) \ t - \rho_T(t-1)]'$ for $t = 2,\ldots,T$. Let $Q = \sum x_t x_t'$, and $r = \sum x_t e_t$, with elements q_{ij} and r_i for $i,j = 1,2$. Then $(\hat{\beta}_{GLS}-\beta) = (q_{11}q_{22}-q_{12}^2)^{-1}(q_{11}r_2 - q_{12}r_1)$. The various parts of the theorem will be proved by evaluating the relevant expressions for q_{ij} and r_i.

Specifically,

$$q_{11} = \sigma_{u_1}^{-2} + (T-1)(1-\rho_T)^2; \ q_{12} = \sigma_{u_1}^{-2} + (T-1)\rho_T(1-\rho_T) + (1-\rho_T)^2\sum_{t=2}^T t;$$

$$q_{22} = \sigma_{u_1}^{-2} + (T-1)\rho_T^2 + 2\rho_T(1-\rho_T)\sum_{t=2}^T t + (1-\rho_T)^2\sum_{t=2}^T t^2;$$

$$r_1 = \sigma_{u_1}^{-2}u_1 + (1-\rho_T)\sum_{t=2}^T v_t; \quad r_2 = \sigma_{u_1}^{-2}u_1 + \sum_{t=2}^T v_t[t(1-\rho_T)+\rho_T].$$

We consider the cases with $\kappa = 0$ and $\kappa > 0$ in turn.

$\kappa = 0$:

By direct calculation:

$$T^{-1}(q_{11}q_{22}-q_{12}^2) \to (1-c+\frac{1}{2}c^2)$$

$$T^{-\frac{1}{2}}q_{11}r_2 = T^{-\frac{1}{2}}\sum_{t=2}^T v_t(1-c\frac{t}{T}) + o_p(1);$$

$$T^{-\frac{1}{2}}q_{12}r_1 \xrightarrow{p} 0.$$

So that,

$$T^{\frac{1}{2}}(\hat{\beta}_{GLS} - \beta) = \frac{T^{-\frac{1}{2}}\sum_{t=2}^{T} v_t(1 - c\frac{t}{T})}{(1 - c + \frac{1}{3}c^2)} + o_p(1)$$

$$\Rightarrow \frac{d(1)\int_0^1(1 - cs)dW(s)]}{(1 - c + \frac{1}{3}c^2)} \sim N(0, (1 - c + \frac{1}{3}c^2)^{-1}) \tag{9}$$

The result follows by noting that $(1 - c + \frac{1}{3}c^2)^{-1} = R_3$ evaluated at $\kappa = 0$.

$\kappa > 0$:

By direct calculation:

$$(q_{11}q_{22} - q_{12}^2) \rightarrow (S_c(\kappa)^{-1} + c^2)(1 - c + \frac{1}{2}c^2) - (\frac{1}{2}c^2 - c)^2$$

$$T^{\frac{1}{2}}q_{11}r_2 = (S_c(\kappa)^{-1} + c^2)(T^{-\frac{1}{2}}\sum_{t=2}^{T} v_t(1 - c\frac{t}{T})) + o_p(1);$$

$$T^{\frac{1}{2}}q_{12}r_1 = (\frac{1}{2}c^2 - c)(S_c(\kappa)^{-1}T^{-\frac{1}{2}}u_1 - cT^{-\frac{1}{2}}\sum_{t=2}^{T} v_t).$$

Thus,

$$T^{\frac{1}{2}}(\hat{\beta}_{GLS} - \beta) = \frac{-(\frac{1}{2}c^2 - c)S_c(\kappa)^{-1}T^{-\frac{1}{2}}u_1 + T^{-\frac{1}{2}}\sum_{t=2}^{T} v_t[(1 - c\frac{t}{T})(S_c(\kappa)^{-1} + c^2) + (\frac{1}{2}c^3 - c^2)]}{(S_c(\kappa)^{-1} + c^2)(1 - c + \frac{1}{2}c^2) - (\frac{1}{2}c^2 - c)^2} + o_p(1)$$

$$\Rightarrow \frac{d(1)[-(\frac{1}{2}c^2 - c)\widetilde{W}_c(\kappa)S_c(\kappa)^{-1} + \int_0^1[(S_c(\kappa)^{-1} + c^2)(1 - cs) + (\frac{1}{2}c^2 - c)c]dW(s)]}{(c^2 + S_c(\kappa)^{-1})(1 - c + \frac{1}{3}c^2) - (\frac{1}{2}c^2 - c)^2} \sim N(0, R_3),$$

where

$$R_3 = d(1)^2[\frac{c^2 + 1}{(S_c(\kappa)c^2 + 1)(1 - c + \frac{1}{3}c^2) - S_c(\kappa)(\frac{1}{2}c^2 - c)^2}].$$

□

A.3 Proof of Lemma 4:

As in the proof the part (c) of Theorem 2, each of the estimators can be written as the OLS estimator from an equation $y_t = x_t'\delta + e_t$, where $\delta = (\alpha \quad \beta)'$, and the estimators differ in their definition of x_1 and e_1. As above, let $Q = \sum x_t x_t'$, and $r = \sum x_t e_t$, with elements q_{ij} and r_i for $i, j = 1, 2$. Then, for each estimator $(\hat{\beta} - \beta) = (q_{11}q_{22} - q_{12}^2)^{-1}(q_{11}r_2 - q_{12}r_1)$ and the for the proof we evaluate these expressions for each estimator.

A.3.1 Proof of (a):

When $c = 0$, $T^{\frac{1}{2}}(\hat{\beta}_{CO} - \beta) = T^{-\frac{1}{2}} \sum_{t=2}^{T} v_t$, and the result follows directly. For $c \neq 0$,

$$q_{11} = (T-1)(1-\rho_T)^2; \quad q_{12} = (T-1)\rho_T(1-\rho_T) + (1-\rho_T)^2 \sum_{t=2}^{T} t;$$

$$q_{22} = (T-1)\rho_T^2 + 2\rho_T(1-\rho_T)\sum_{t=2}^{T} t + (1-\rho_T)^2 \sum_{t=2}^{T} t^2;$$

$$r_1 = (1-\rho_T)\sum_{t=2}^{T} v_t; \quad r_2 = \sum_{t=2}^{T} v_t[t(1-\rho_T) + \rho_T].$$

Thus,

$$q_{11}q_{22} - q_{12}^2 \to c^2(1 - c + \frac{1}{3}c^2) - (\frac{1}{2}c^2 - c)^2 = \frac{1}{12}c^4;$$

$$T^{\frac{1}{2}} q_{11} r_2 = -c^2 T^{-\frac{1}{2}} \sum v_t (c\frac{t}{T} - 1) + o_p(1);$$

$$T^{\frac{1}{2}} q_{12} r_1 = c^2(1 - \frac{1}{2}c) T^{-\frac{1}{2}} \sum v_t + o_p(1).$$

So that,

$$T^{\frac{1}{2}}(\hat{\beta}_{CO} - \beta) = -(\frac{12}{c^2})[T^{-\frac{1}{2}} \sum v_t (c\frac{t}{T} - 1) + (1 - \frac{1}{2}c) T^{-\frac{1}{2}} \sum v_t] + o_p(1)$$

$$= (\frac{12}{c})[T^{-\frac{1}{2}} \sum v_t (\frac{1}{2} - \frac{t}{T}) + o_p(1) \tag{10}$$

$$\Rightarrow (\frac{12}{c}) d(1) \int_0^1 (\frac{1}{2} - s) dW(s). \tag{11}$$

The result follows by noting that

$$(\frac{12}{c}) d(1) \int_0^1 (\frac{1}{2} - s) dW(s) \sim N(0, G_1),$$

where

$$G_1 = (\frac{12}{c})^2 d(1)^2 \int_0^1 (\frac{1}{2} - s)^2 ds = \frac{12 d(1)^2}{c^2}.$$

A.3.2 Proof of (b):

For $\hat{\beta}_{CC}$,

$$q_{11} = 1 + (T-1)(1-\rho_T)^2; \quad q_{12} = 1 + (T-1)\rho_T(1-\rho_T) + (1-\rho_T)^2 \sum_{t=2}^{T} t;$$

$$q_{22} = 1 + (T-1)\rho_T^2 + 2\rho_T(1-\rho_T)\sum_{t=2}^{T} t + (1-\rho_T)^2 \sum_{t=2}^{T} t^2;$$

$$r_1 = u_1 + (1-\rho_T)\sum_{t=2}^{T} v_t; \quad r_2 = u_1 + \sum_{t=2}^{T} v_t[t(1-\rho_T) + \rho_T].$$

Thus,

$$T^{-1}(q_{11}q_{22} - q_{12}^2) \rightarrow (1 - c + \frac{1}{3}c^2);$$

$$T^{-\frac{1}{2}}q_{11}r_2 = T^{-\frac{1}{2}}u_1 - T^{-\frac{1}{2}}\sum_{t=2}^{T} v_t(c\frac{t}{T} - 1) + o_p(1);$$

$$T^{-\frac{1}{2}}q_{12}r_1 = (1 - c + \frac{1}{2}c^2)T^{-\frac{1}{2}}u_1 + o_p(1).$$

So that,

$$T^{\frac{1}{2}}(\hat{\beta}_{CC} - \beta) = \frac{c(1-\frac{1}{2}c)T^{-\frac{1}{2}}u_1 - T^{-\frac{1}{2}}\sum v_t(c\frac{t}{T} - 1)}{(1 - c + \frac{1}{3}c^2)} + o_p(1) \tag{12}$$

$$\Rightarrow \frac{d(1)}{1 - c + \frac{1}{3}c^2}[c(1 - \frac{1}{2}c)\widetilde{W}_c(\kappa) - \int_0^1 (cs - 1)dW(s)]. \tag{13}$$

The result follows by noting that

$$d(1)(1 - c + \frac{1}{3}c^2)^{-1}[c(1 - \frac{1}{2}c)\widetilde{W}_c(\kappa) - \int_0^1 (cs-1)dW(s)] \sim N(0, G_2),$$

where

$$G_2 = \frac{d(1)^2}{(1 - c + \frac{1}{3}c^2)^2}[(c - \frac{1}{2}c^2)^2 S_c(\kappa) + \int_0^1 (cs-1)^2 ds]$$

$$= \frac{d(1)^2}{1 - c + \frac{1}{3}c^2}[1 + S_c(\kappa)\frac{(c - \frac{1}{2}c^2)^2}{1 - c + \frac{1}{3}c^2}].$$

A.3.3 Proof of (c):

For $\hat{\beta}_{PW}$,

$$q_{11} = (1-\rho_T^2)+(T-1)(1-\rho_T)^2; \quad q_{12} = (1-\rho_T^2)+(T-1)\rho_T(1-\rho_T)+(1-\rho_T)^2\sum_{t=2}^{T} t;$$

$$q_{22} = (1-\rho_T^2)+(T-1)\rho_T^2+2\rho_T(1-\rho_T)\sum_{t=2}^{T} t+(1-\rho_T)^2\sum_{t=2}^{T} t^2;$$

$$r_1 = (1-\rho_T^2)u_1+(1-\rho_T)\sum_{t=2}^{T} v_t; \quad r_2 = (1-\rho_T^2)u_1+\sum_{t=2}^{T} v_t[t(1-\rho_T)+\rho_T].$$

Thus,

$$\begin{aligned} q_{11}q_{22}-q_{12}^2 &\to (c^2-2c)(1-c+\frac{1}{3}c^2)-(\frac{1}{2}c^2-c)^2 \\ &= (c^2-2c)(1-\frac{1}{2}c+\frac{1}{12}c^2); \end{aligned}$$

$$T^{\frac{1}{2}}q_{11}r_2 = -(c^2-2c)T^{-\frac{1}{2}}\sum_{t=2}^{T} v_t(c\frac{t}{T}-1)+o_p(1);$$

$$T^{\frac{1}{2}}q_{12}r_1 = -\frac{1}{2}(c^2-2c)(2cT^{-\frac{1}{2}}u_1+cT^{-\frac{1}{2}}\sum_{t=2}^{T} v_t)+o_p(1).$$

So that,

$$T^{\frac{1}{2}}(\hat{\beta}_{PW}-\beta) = \frac{cT^{-\frac{1}{2}}u_1-T^{-\frac{1}{2}}\sum v_t(c\frac{t}{T}-\frac{1}{2}c-1)}{1-\frac{1}{2}c+\frac{1}{12}c^2}+o_p(1) \tag{14}$$

$$\Rightarrow \frac{d(1)}{1-\frac{1}{2}c+\frac{1}{12}c^2}[c\widetilde{W}_c(\kappa)-\int_0^1(cs-\frac{1}{2}c-1)dW(s)]. \tag{15}$$

The result follows by noting that

$$d(1)(1-\frac{1}{2}c+\frac{1}{12}c^2)[c\widetilde{W}_c(\kappa)-\int_0^1(cs-\frac{1}{2}c-1)dW(s)] \sim N(0,G_3),$$

where

$$\begin{aligned} G_3 &= \frac{d(1)^2}{(1-\frac{1}{2}c+\frac{1}{12}c^2)^2}[c^2S_c(\kappa)+\int_0^1(cs-\frac{1}{2}c-1)^2ds] \\ &= \frac{d(1)^2}{(1-\frac{1}{2}c+\frac{1}{12}c^2)^2}[c^2S_c(\kappa)+1+\frac{1}{12}c^2]. \end{aligned}$$

□

A.4 Proof of Theorem 5:

It is straightforward to verify that the analogues of (6), (8), and (10) continue to hold for the feasible GLS estimators, with $\hat{c}$ replacing c and $\hat{v}_t = u_t - \hat{\rho}_T u_{t-1}$, replacing v_t. The theorem then follows from (7), (9) and (11) using $T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} \hat{v}_t \Rightarrow \widehat{W}(c)$. To see this, and to derive an expression for $\widehat{W}(c)$, write

$$\hat{v}_t = u_t - \hat{\rho}_T u_{t-1} = v_t - (\hat{\rho}_T - \rho_T) u_{t-1}$$

$$= v_t - (\hat{c}_T - c)T^{-1}[\sum_{j=0}^{t-2} \rho_T^j v_{t-1-j} + \rho_T^{t-1} u_1],$$

where $\hat{c}_T = T(1 - \hat{\rho}_T)$. Thus

$$T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} \hat{v}_t =$$

$$T^{-\frac{1}{2}} \sum_{t=1}^{[sT]} v_t - (\hat{c}_T - c)T^{-1} \sum_{t=1}^{[sT]} T^{-\frac{1}{2}} (\sum_{j=0}^{t-2} \rho_T^j v_{t-1-j}) - (\hat{c}_T - c)(T^{-\frac{1}{2}} u_1) T^{-1} \sum_{t=1}^{[sT]} \rho_T^{t-1}]$$

$$\Rightarrow d(1)\widehat{W}(s),$$

where

$$\widehat{W}(s) = W(s) - (\hat{c} - c) \int_o^s W_c(\tau) d\tau - (\hat{c} - c)\widetilde{W}_c(\kappa) \frac{1 - e^{sc}}{-c},$$

and the last line follows from $\hat{c}_T \xrightarrow{p} \hat{c}$, $T^{-\frac{1}{2}} \sum_{j=0}^{[\tau T]} \rho_T^j v_{[\tau T]-j} \Rightarrow W_c(\tau)$, and $T^{-1} \sum_{t=1}^{[sT]} \rho_T^{t-1} - \frac{1-e^{sc}}{-c}$.
□

References

[1] Beach, C.M. and J.G.MacKinnon (1978), "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," *Econometrica*, 46, no. 1, pp. 51-58.

[2] Bobkosky, M.J. (1983), *Hypothesis Testing in Nonstationary Time Series* , Ph.D. thesis, Department of Statistics, University of Wisconsin.

[3] Cavanagh, C.L. (1985), "Roots Local to Unity," manuscript, Department of Economics, Harvard University.

[4] Cavanagh, C., G. Elliott and J. Stock (1993), "Inference in Models with Nearly Nonstationary Regressors," manuscript, Kennedy School of Government, Harvard University.

[5] Chan, N.H. (1988), "On Parameter Inference for Nearly Nonstationary Time Series," *Journal of the American Statistical Association* , 83, pp. 857-862.

[6] Chan, N.H. and Wei, C.Z. (1987), "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics,* 15, pp. 1050-63.

[7] Chipman, John S. (1979), "Efficiency of Least-Squares Estimation of Linear Trend When Residuals are Autocorrelated," *Econometrica*, 47, pp. 115-128.

[8] Cochrane, D. and G.H. Orcutt (1949), "Applications of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, 44, pp. 32-61.

[9] Davidson, R. and J.G. MacKinnon (1993), *Estimation and Inference in Econometrics*, Oxford University Press: New York.

[10] Dickey, D.A. and W.A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, no. 366, pp. 427-31.

[11] Dufour, Jean-Marie (1990), "Exact Test and Confidence Sets in Linear Regressions With Autocorrelated Errors," *Econometrica*, 58, no. 2, pp. 475-494.

[12] Durlauf, S.N. and P.C.B. Phillips (1988), "Trends Versus Random Walks in Time Series Analysis," *Econometrica*, 56, pp. 1333-54.

[13] Elliott, Graham (1993),"Efficient Tests for a Unit Root when the Initial Observation is Drawn from its Unconditional Distribution," manuscript, Harvard University.

[14] Elliott, Graham, Thomas J. Rothenberg and James H. Stock (1992), "Efficient Tests of an Autoregressive Unit Root," NBER Technical Working Paper 130.

[15] Grenander, U. and M. Rosenblatt (1957), *Statistical Analysis of Stationary Time Series*, John Wiley and Sons: New York.

[16] Kadiyala, K.R. (1968), "A Transformation Used to Circumvent the Problem of Autocorrelated Errors," *Econometrica*, 36, pp. 93-96.

[17] Maeshiro, Asatoshi (1976), "Autoregressive Transformation, Trended Independent Variables and Autocorrelated Disturbances," *Review of Economics and Statistics*, 58, pp. 497-500.

[18] Maeshiro, Asatoshi (1979), "On the Retention of the First Observations in Serial Correlation Adjustment of Regression Models," *International Economic Review*, 20, no. 1, pp. 259-265.

[19] Magee, L. (1987), "A Note Cochrane-Orcutt Estimation," *Journal of Econometrics*, 35, pp. 211-18.

[20] Nagaraj, N.K. and W.A. Fuller (1991), "Estimation of the Parameters of Linear Time Series Models Subject to Nonlinear Restrictions," *Annals of Statistics*, 19, no. 3, pp. 1143-54..

[21] Park, R.E. and Mitchell, B.M. (1980), "Estimating the Autocorrelated Error Model with Trended Data," *Journal of Econometrics*, 13, pp. 185-201.

[22] Phillips, P.C.B. (1987), "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika*, 74, pp. 535-47.

[23] Phillips P.C.B. and V. Solo (1992), "Asymptotics for Linear Processes," *Annals of Statistics*, 20, pp. 971-1001.

[24] Prais, S.J. and C.B. Winsten (1954), "Trend Estimators and Serial Correlation," Cowles Foundation Discussion Paper 383.

[25] Quah, D. and J. Wooldridge, (1988), "A Common Error in the Treatment of Trending Time Series," Manuscript, Department of Economics, M.I.T..

[26] Rao, P. and Z. Griliches (1969), "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Auto-Correlated Errors," *Journal of the American Statistical Association*, 64, pp. 253-72.

[27] Sampson M. (1991), "The Effect of Parameter Uncertainty on Forecast Variances and Confidence Intervals for Unit Root and Trend Stationary Time-series Models," *Journal of Applied Econometrics*, 6, no 1, pp. 67-76.

[28] Schmidt, Peter (1993), "Some Results on Testing for Stationarity Using Data Detrended in Differences," *Economics Letters*, 41. pp. 1-6.

[29] Schmidt, P. and P.C.B. Phillips (1992), "LM Tests for a Unit Root in the Presence of Deterministic Trends," *Oxford Bulletin of Economics and Statistics*, 54, pp. 257-87.

[30] Spitzer, J.J. (1979), "Small Sample Properties of Nonlinear Least Squares and Maximum Likelihood Estimators in the Context of Autocorrelated Errors," *Journal of the American Statistical Association*, 74, pp. 41-47.

[31] Stock, James H. (1991), "Confidence Intervals of the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics*, 28, pp. 435-60.

[32] Summers, Robert and Alan Heston (1991), "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988," *Quarterly Journal of Economics*, 106, pp. 327-368.

[33] Thornton, D.L. (1987), "A Note the Effect of the Cochrane-Orcutt Estimator in the AR(1) Regression Model," *Journal of Econometrics*, 36, pp. 369-76.

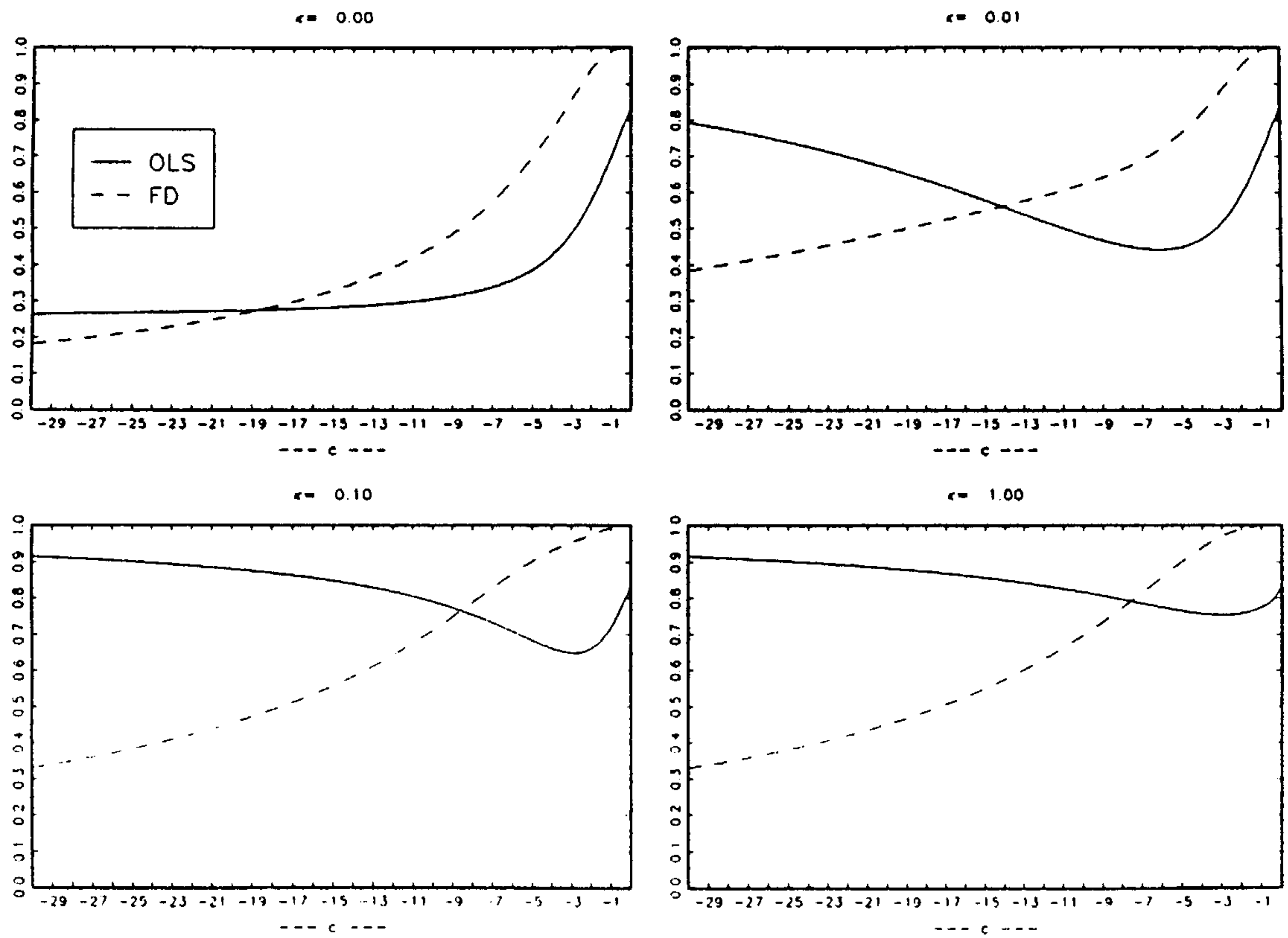

Figure 1

Asymptotic Relative Efficiencies of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FD}$

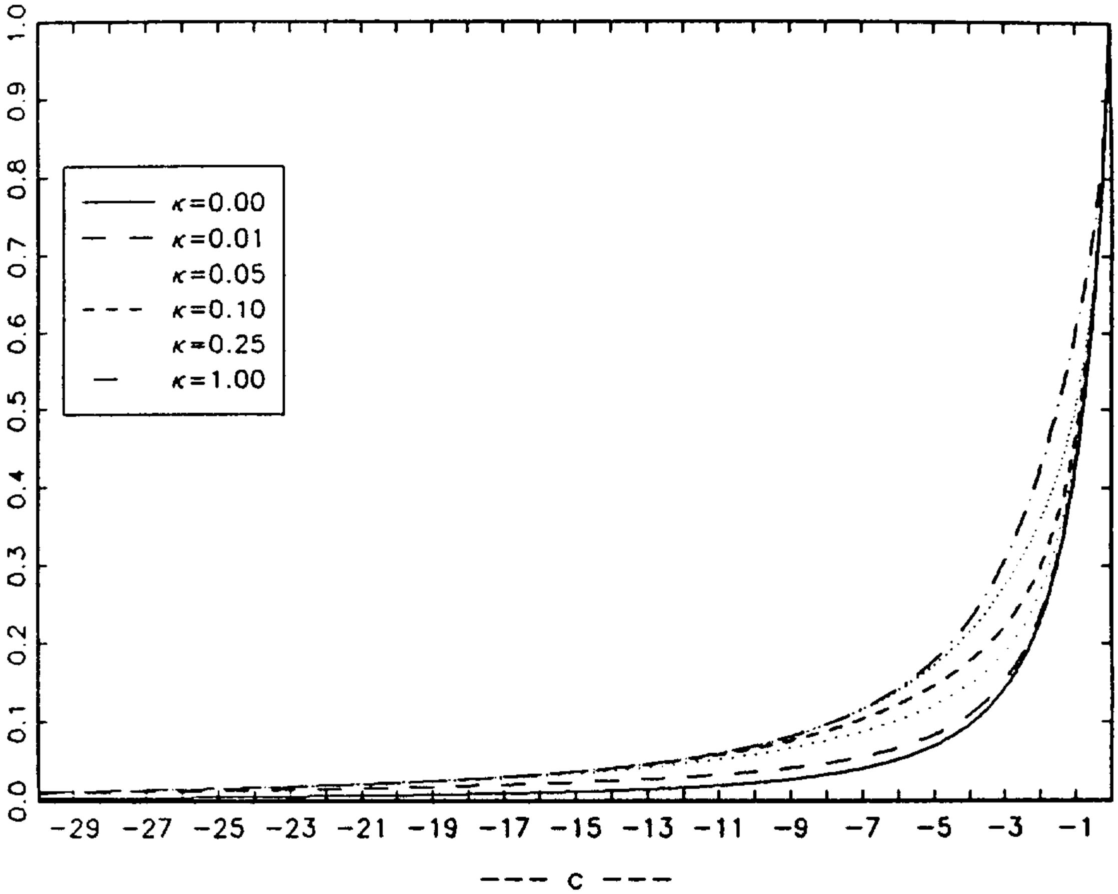

Figure 2

Asymptotic Variance of $\hat{\beta}_{GLS}$

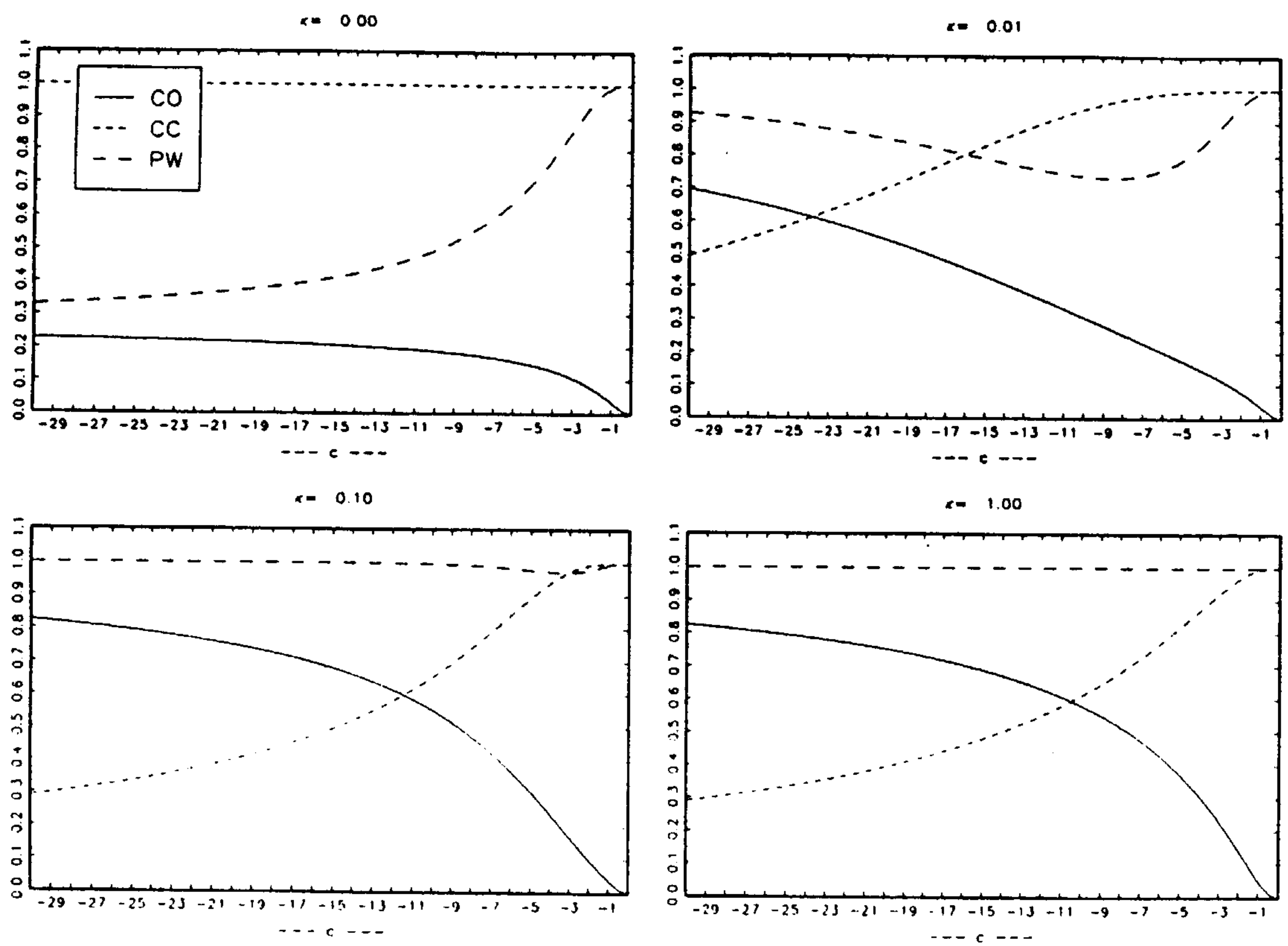

Figure 3

Asymptotic Relative Efficiencies of $\hat{\beta}_{CO}$, $\hat{\beta}_{CC}$ and $\hat{\beta}_{PW}$

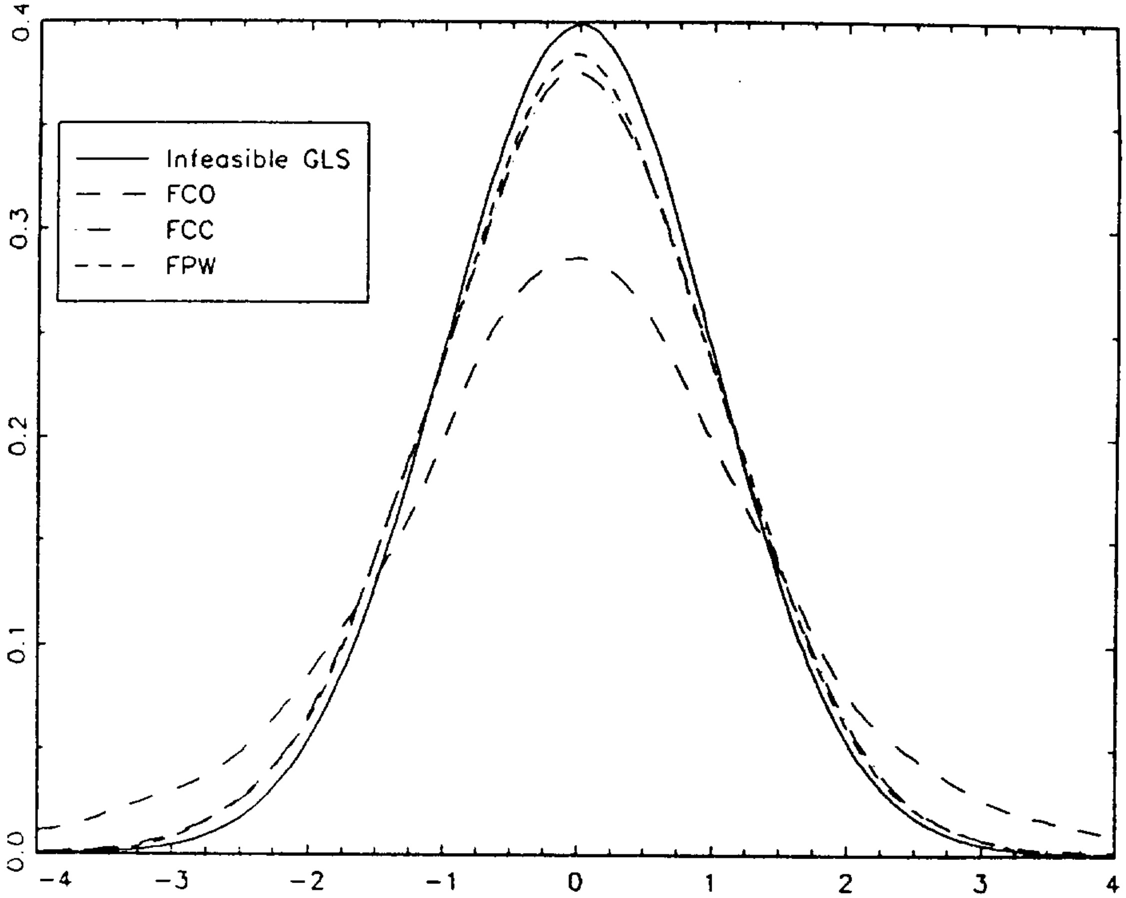

Figure 4

Densities of Feasible GLS Estimators

Table 1
Average Mean Square Error of Estimators

A. Average MSE for -30≤c≤0

Estimator	κ = 0.000	0.010	0.050	0.100	0.250	1.000
$\hat{\beta}_{GLS}$	0.057	0.067	0.081	0.089	0.097	0.105
$\hat{\beta}_{OLS}$	0.108	0.110	0.116	0.120	0.126	0.133
$\hat{\beta}_{FD}$	0.077	0.085	0.102	0.111	0.121	0.129
$\hat{\beta}_{FCC}$	0.065	0.083	0.113	0.128	0.142	0.152
$\hat{\beta}_{FPW}$	0.077	0.081	0.088	0.094	0.101	0.108
$\hat{\beta}_{PT}$	0.082	0.085	0.094	0.102	0.110	0.118

B. Average MSE for -2≤c≤0

Estimator	κ = 0.000	0.010	0.050	0.100	0.250	1.000
$\hat{\beta}_{GLS}$	0.493	0.497	0.512	0.529	0.566	0.634
$\hat{\beta}_{OLS}$	0.678	0.682	0.695	0.710	0.743	0.808
$\hat{\beta}_{FD}$	0.498	0.502	0.516	0.532	0.567	0.635
$\hat{\beta}_{FCC}$	0.595	0.598	0.615	0.641	0.676	0.766
$\hat{\beta}_{FPW}$	0.529	0.531	0.544	0.559	0.590	0.664
$\hat{\beta}_{PT}$	0.509	0.512	0.526	0.541	0.574	0.649

C. Average MSE for -10≤c≤-2

Estimator	κ = 0.000	0.010	0.050	0.100	0.250	1.000
$\hat{\beta}_{GLS}$	0.069	0.083	0.115	0.136	0.158	0.168
$\hat{\beta}_{OLS}$	0.168	0.172	0.186	0.197	0.211	0.219
$\hat{\beta}_{FD}$	0.097	0.107	0.134	0.154	0.179	0.190
$\hat{\beta}_{FCC}$	0.075	0.093	0.135	0.168	0.206	0.220
$\hat{\beta}_{FPW}$	0.106	0.114	0.131	0.146	0.165	0.173
$\hat{\beta}_{PT}$	0.110	0.118	0.141	0.160	0.182	0.192

D. Average MSE for -30≤c≤-10

Estimator	κ = 0.000	0.010	0.050	0.100	0.250	1.000
$\hat{\beta}_{GLS}$	0.008	0.018	0.025	0.026	0.026	0.026
$\hat{\beta}_{OLS}$	0.028	0.028	0.030	0.030	0.030	0.030
$\hat{\beta}_{FD}$	0.027	0.035	0.049	0.052	0.053	0.053
$\hat{\beta}_{FCC}$	0.008	0.028	0.054	0.060	0.063	0.063
$\hat{\beta}_{FPW}$	0.020	0.022	0.025	0.026	0.027	0.027
$\hat{\beta}_{PT}$	0.028	0.029	0.033	0.035	0.035	0.035

Notes: The entries in the table are the mean squared error averaged over the indicated range of c.

Table 2
Relative Efficiencies of Estimators
Exact and I(1) Approximation

A. κ = 0.00

	--- T=30 ---		--- T=50 ---		--- T=100 ---	
	Exact	I(1)	Exact	I(1)	Exact	I(1)
			(i) ρ=0.50			
$\hat{\beta}_{OLS}$	0.891	0.283	0.924	0.268	0.958	0.258
$\hat{\beta}_{FD}$	0.492	0.330	0.335	0.213	0.185	0.113
$\hat{\beta}_{FCC}$	0.958	0.990	1.000	1.000	0.974	1.000
$\hat{\beta}_{FPW}$	0.957	0.390	1.000	0.350	0.975	0.301
$\hat{\beta}_{PT}$	0.705	0.276	0.941	0.272	0.953	0.258
			(ii) ρ=0.80			
$\hat{\beta}_{OLS}$	0.607	0.358	0.631	0.305	0.714	0.273
$\hat{\beta}_{FD}$	0.810	0.632	0.670	0.451	0.466	0.259
$\hat{\beta}_{FCC}$	0.922	0.950	0.937	0.968	0.947	0.997
$\hat{\beta}_{FPW}$	0.838	0.560	0.824	0.454	0.839	0.355
$\hat{\beta}_{PT}$	0.805	0.529	0.664	0.363	0.675	0.260
			(iii) ρ=0.90			
$\hat{\beta}_{OLS}$	0.579	0.488	0.511	0.386	0.498	0.305
$\hat{\beta}_{FD}$	0.912	0.859	0.803	0.698	0.616	0.451
$\hat{\beta}_{FCC}$	0.887	0.917	0.940	0.944	0.925	0.962
$\hat{\beta}_{FPW}$	0.809	0.786	0.772	0.632	0.695	0.442
$\hat{\beta}_{PT}$	0.882	0.815	0.778	0.631	0.560	0.343
			(iv) ρ=0.95			
$\hat{\beta}_{OLS}$	0.678	0.635	0.572	0.529	0.452	0.386
$\hat{\beta}_{FD}$	0.978	0.971	0.923	0.902	0.760	0.698
$\hat{\beta}_{FCC}$	0.889	0.797	0.875	0.899	0.939	0.942
$\hat{\beta}_{FPW}$	0.924	0.874	0.840	0.830	0.697	0.619
$\hat{\beta}_{PT}$	0.978	0.905	0.906	0.860	0.720	0.622
			(v) ρ=1.00			
$\hat{\beta}_{OLS}$	0.860	0.833	0.850	0.833	0.842	0.833
$\hat{\beta}_{FD}$	1.000	1.000	1.000	1.000	1.000	1.000
$\hat{\beta}_{FCC}$	0.907	0.830	0.903	0.859	0.892	0.821
$\hat{\beta}_{FPW}$	0.944	0.965	0.964	0.966	0.977	0.974
$\hat{\beta}_{PT}$	0.969	0.986	0.992	0.993	1.000	0.996

Table 2 (Continued)
Relative Efficiencies of Estimators
Exact and I(1) Approximation

B. κ = 1.00

	--- T=30 ---		--- T=50 ---		--- T=100 ---	
	Exact	I(1)	Exact	I(1)	Exact	I(1)
			(i) ρ=0.50			
$\hat\beta_{OLS}$	0.950	0.856	0.966	0.902	0.982	0.946
$\hat\beta_{FD}$	0.463	0.550	0.308	0.381	0.166	0.213
$\hat\beta_{FCC}$	0.981	0.470	0.960	0.328	1.000	0.174
$\hat\beta_{FPW}$	1.000	0.979	0.979	0.974	1.000	0.994
$\hat\beta_{PT}$	0.735	0.698	0.932	0.859	1.000	0.949
			(ii) ρ=0.80			
$\hat\beta_{OLS}$	0.839	0.774	0.867	0.817	0.915	0.883
$\hat\beta_{FD}$	0.841	0.859	0.667	0.698	0.420	0.451
$\hat\beta_{FCC}$	0.932	0.756	0.870	0.591	0.844	0.393
$\hat\beta_{FPW}$	0.975	0.990	0.958	0.974	1.000	1.000
$\hat\beta_{PT}$	0.834	0.866	0.699	0.721	0.836	0.809
			(iii) ρ=0.90			
$\hat\beta_{OLS}$	0.801	0.753	0.800	0.764	0.842	0.817
$\hat\beta_{FD}$	0.970	0.971	0.895	0.902	0.683	0.698
$\hat\beta_{FCC}$	0.916	0.863	0.870	0.786	0.805	0.621
$\hat\beta_{FPW}$	0.950	0.989	0.942	0.960	0.980	0.994
$\hat\beta_{PT}$	0.942	0.971	0.858	0.869	0.720	0.746
			(iv) ρ=0.95			
$\hat\beta_{OLS}$	0.803	0.764	0.781	0.755	0.782	0.764
$\hat\beta_{FD}$	0.997	0.997	0.983	0.983	0.898	0.902
$\hat\beta_{FCC}$	0.886	0.861	0.876	0.829	0.863	0.786
$\hat\beta_{FPW}$	0.951	0.959	0.948	0.957	0.983	0.979
$\hat\beta_{PT}$	0.994	0.975	0.970	0.962	0.888	0.898
			(v) ρ=1.00			
$\hat\beta_{OLS}$	0.860	0.833	0.850	0.833	0.842	0.833
$\hat\beta_{FD}$	1.000	1.000	1.000	1.000	1.000	1.000
$\hat\beta_{FCC}$	0.936	0.895	0.927	0.879	0.863	0.838
$\hat\beta_{FPW}$	0.989	0.991	1.000	0.965	0.958	0.959
$\hat\beta_{PT}$	1.000	1.000	1.000	0.996	0.980	0.981

Notes: The relative efficiency is the ratio of the variance of the infeasible GLS estimator to the variance of the estimator given in column 1. The columns labeled I(1) are the asymptotic relative efficiencies using $c=T(\rho-1)$. The corresponding I(0) relative efficiencies are 1, 0, 1, 1, 1, 1, respectively for the estimators in column 1 and for all T and $|\rho|<1$.

Table 3
Confidence Interval Coverage Rates (%)

A. Smallest of $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FD}$ Confidence Intervals

(i) $T = 30$

Level	κ	c: 0	-1	-5	-10	-20
90.0	0.0	88.7	96.0	97.8	97.3	94.5
90.0	0.1	88.7	95.6	97.4	96.6	94.5
90.0	1.0	89.2	94.0	97.1	96.8	94.3
95.0	0.0	93.0	97.9	98.8	98.6	96.9
95.0	0.1	92.9	97.9	98.7	98.3	97.1
95.0	1.0	93.6	96.7	98.5	98.5	96.9
99.0	0.0	97.9	99.4	99.7	99.5	99.0
99.0	0.1	98.0	99.5	99.6	99.6	99.1
99.0	1.0	97.9	99.2	99.6	99.7	99.2

(ii) $T = 50$

Level	κ	c: 0	-1	-5	-10	-20
90.0	0.0	90.7	97.3	98.3	98.2	96.8
90.0	0.1	90.8	96.3	97.9	97.9	96.8
90.0	1.0	90.5	95.2	97.9	98.1	96.9
95.0	0.0	94.9	98.7	99.2	99.3	98.6
95.0	0.1	94.8	98.3	99.0	99.1	98.4
95.0	1.0	94.9	97.7	99.0	99.1	98.5
99.0	0.0	98.6	99.7	99.8	99.9	99.7
99.0	0.1	98.7	99.6	99.8	99.8	99.6
99.0	1.0	98.8	99.6	99.8	99.9	99.7

(iii) $T = 100$

Level	κ	c: 0	-1	-5	-10	-20
90.0	0.0	91.4	97.8	98.8	98.6	98.5
90.0	0.1	92.3	97.5	98.4	98.8	98.3
90.0	1.0	91.9	96.4	98.3	98.8	98.3
95.0	0.0	95.6	99.0	99.5	99.5	99.4
95.0	0.1	95.8	99.0	99.3	99.5	99.4
95.0	1.0	95.9	98.5	99.3	99.6	99.4
99.0	0.0	99.1	99.8	99.9	100.0	99.9
99.0	0.1	98.8	99.9	99.9	99.9	99.9
99.0	1.0	99.0	99.8	99.9	100.0	99.9

Table 3 (Continued)
Confidence Interval Coverage Rates (%)

B. Confidence Intervals Constructed from $\hat{\beta}_{PW}$

(i) T = 30

Level	κ	c = 0	-1	-5	-10	-20
90.0	0.0	88.5	96.0	97.7	97.2	93.9
90.0	0.1	88.6	95.4	97.3	96.5	93.6
90.0	1.0	89.0	93.7	97.1	96.5	93.6
95.0	0.0	93.0	97.8	98.7	98.5	96.5
95.0	0.1	92.9	97.8	98.5	98.2	96.5
95.0	1.0	93.4	96.6	98.5	98.4	96.5
99.0	0.0	97.8	99.3	99.7	99.5	98.9
99.0	0.1	97.9	99.5	99.6	99.5	99.0
99.0	1.0	97.9	99.2	99.7	99.6	99.0
(ii) T = 50						
90.0	0.0	90.5	97.1	98.4	98.4	96.5
90.0	0.1	90.7	96.2	97.8	97.9	96.4
90.0	1.0	90.3	95.1	97.8	98.0	96.4
95.0	0.0	94.8	98.6	99.1	99.3	98.5
95.0	0.1	94.7	98.2	99.0	99.1	98.2
95.0	1.0	94.9	97.6	99.0	99.1	98.3
99.0	0.0	98.6	99.7	99.8	99.9	99.7
99.0	0.1	98.7	99.5	99.8	99.8	99.6
99.0	1.0	98.8	99.6	99.7	99.9	99.7
(iii) T = 100						
90.0	0.0	91.3	97.7	98.9	98.9	98.6
90.0	0.1	92.1	97.5	98.4	98.8	98.2
90.0	1.0	91.8	96.3	98.3	98.8	98.2
95.0	0.0	95.6	99.0	99.5	99.6	99.4
95.0	0.1	95.8	98.9	99.3	99.5	99.3
95.0	1.0	95.8	98.5	99.3	99.6	99.3
99.0	0.0	99.1	99.8	99.9	100.0	99.9
99.0	0.1	98.8	99.8	99.9	99.9	99.8
99.0	1.0	99.0	99.8	99.9	99.9	99.9

Notes: The table shows the exact coverage rates (in percent) for conservative confidence intervals constructed with an asymptotic level given in the first column. The confidence intervals in panel A were constructed as the narrowest of the intervals constructed from the OLS and first-difference estimators. The confidence intervals from panel B were constructed from the Prais-Winsten estimator.

Table 4
Annual Real Per-Capita Growth Rates

ID	Country	Smpl Period		$\hat{g}_{OLS}$	$\hat{g}_{FD}$	$\hat{g}_{FCO}$	$\hat{g}_{FPW}$	$\hat{c}$	$\hat{t}^{T}$	g_{min}	g_{max}
1	ALGERIA	1960	1990	2.736	1.450	2.766	2.090	-6.070	-1.484	-1.183	4.102
2	ANGOLA	1960	1989	-2.038	-1.004	-2.759	-1.345	-3.753	-2.298	-6.078	4.071
3	BENIN	1959	1989	-0.403	-0.370	-0.581	-0.397	-15.180	-3.230	-1.840	1.089
4	BOTSWANA	1960	1989	5.839	6.079	6.173	5.906	-10.855	-3.262	2.160	9.998
5	BURKINA FASO	1959	1990	0.859	0.006	-20.376	0.006	0.290	-3.498	-1.783	1.795
6	BURUNDI	1960	1990	0.567	-0.425	2.043	0.041	-5.586	-4.247	-0.347	0.936
7	CAMEROON	1960	1990	2.698	1.888	1.455	2.052	-2.439	-1.816	-1.116	4.894
8	CAPE VERDE IS.	1960	1989	3.880	3.338	4.693	3.545	-4.369	-2.548	-0.557	7.234
9	CENTRAL AFR.R.	1960	1990	-0.486	-0.588	-0.879	-0.542	-5.331	-1.164	-2.051	0.873
10	CHAD	1960	1990	-2.584	-2.010	-2.580	-2.442	-12.200	-2.828	-5.228	1.208
11	COMOROS	1960	1987	-0.044	0.520	-0.780	0.219	-8.490	-3.096	-3.013	4.054
12	CONGO	1960	1990	3.314	2.788	3.360	3.082	-7.081	-2.455	-0.288	5.863
14	EGYPT	1950	1990	3.002	2.385	3.696	2.481	-2.018	-3.725	1.472	3.481
15	ETHIOPIA	1951	1986	0.831	0.669	0.688	0.761	-7.452	-1.814	-0.215	1.553
16	GABON	1960	1990	2.298	2.620	-37.342	2.816	-0.358	-1.413	-2.810	8.150
17	GAMBIA	1960	1990	1.150	0.835	-0.222	1.015	-4.243	-1.112	-1.874	3.744
18	GHANA	1955	1989	-0.266	-0.089	-0.425	-0.205	-8.501	-2.418	-2.260	2.102
19	GUINEA	1959	1989	-0.304	-0.215	-0.590	-0.280	-6.130	-2.108	-1.821	1.481
20	GUINEA-BISS	1960	1990	0.377	1.036	-0.212	0.724	-5.854	-2.089	-2.626	4.898
21	IVORY COAST	1960	1990	1.073	0.833	5.807	0.833	2.305	0.611	-1.518	2.784
22	KENYA	1950	1990	1.170	1.166	1.191	1.177	-17.818	-2.878	-0.648	2.981
23	LESOTHO	1960	1990	4.402	4.053	2.099	4.129	-2.585	-1.881	0.155	7.852
24	LIBERIA	1960	1986	0.682	0.310	449.383	0.310	0.023	0.194	-1.703	2.323
25	MADAGASCAR	1960	1990	-1.962	-1.818	-2.441	-1.880	-5.082	-2.064	-3.456	-0.178
26	MALAWI	1954	1990	1.171	1.226	0.684	1.197	-6.774	-2.016	-0.456	2.908
27	MALI	1960	1990	0.877	0.150	3.057	0.274	-2.117	-2.758	-1.433	1.734
28	MAURITANIA	1960	1990	-0.164	-0.207	-0.641	-0.182	-7.652	-1.869	-2.457	2.843
29	MAURITIUS	1950	1990	1.340	1.398	3.893	1.385	-2.728	-1.677	-1.644	4.441
30	MOROCCO	1950	1990	2.814	2.355	2.804	2.600	-7.017	-2.889	0.083	4.826
31	MOZAMBIQUE	1960	1990	-2.309	-1.426	-7.856	-1.493	-1.207	-2.240	-4.818	1.966
32	NAMIBIA	1960	1989	0.384	0.509	-5.591	0.493	-1.734	-1.409	-2.509	3.528
33	NIGER	1960	1989	-0.415	-0.256	-4.612	-0.288	-2.387	-1.924	-3.373	2.861
34	NIGERIA	1950	1990	1.969	1.337	-8.255	1.475	-2.578	-1.808	-3.502	8.176
35	REUNION	1960	1988	3.764	3.798	2.695	3.784	-4.841	-2.252	2.168	5.429
36	RWANDA	1960	1990	1.974	0.791	1.916	1.366	-5.842	-2.516	-3.566	5.148
37	SENEGAL	1960	1990	0.138	0.204	0.101	0.139	-28.582	-3.862	-0.158	0.484
38	SEYCHELLES	1960	1989	3.896	3.449	4.072	3.783	-11.867	-2.289	1.558	5.340
39	SIERRA LEONE	1961	1990	0.049	0.593	-4.627	0.519	-1.789	-2.238	-1.868	3.055
40	SOMALIA	1960	1989	-0.448	-0.551	-0.387	-0.460	-18.529	-2.486	-3.783	2.680
41	SOUTH AFRICA	1950	1990	1.792	1.343	-16.304	1.344	-0.208	-0.186	-0.201	2.886
42	SUDAN	1971	1990	-0.286	-0.886	-0.373	-0.509	-7.416	-2.518	-5.109	3.337
43	SWAZILAND	1960	1989	1.626	1.985	-4.704	1.932	-1.910	-1.261	-2.135	6.105
44	TANZANIA	1960	1988	1.532	1.686	0.710	1.626	-4.431	-1.819	-0.432	3.803
45	TOGO	1960	1990	1.777	1.797	-1.485	1.783	-2.130	-1.408	-0.565	4.159
46	TUNISIA	1960	1990	3.761	3.222	2.980	3.298	-1.855	-1.123	1.368	5.075
47	UGANDA	1950	1989	-0.188	0.946	-0.031	0.177	-10.558	-2.575	-2.985	4.877
48	ZAIRE	1950	1989	0.339	0.648	19.157	0.648	0.527	-1.250	-1.554	2.850
49	ZAMBIA	1955	1990	-0.613	-0.597	-2.633	-0.602	-3.195	-1.120	-2.085	1.790
50	ZIMBABWE	1954	1990	0.904	1.018	0.785	0.950	-8.292	-2.768	-1.094	3.131

Table 4 (Continued)
Annual Real Per-Capita Growth Rates

ID	Country	Smpl Period		$\hat{\theta}_{OLS}$	$\hat{\theta}_{FD}$	$\hat{\theta}_{FCO}$	$\hat{\theta}_{FPW}$	$\hat{\tau}$	$\hat{\tau}^{*}$	θ_{min}	θ_{max}
52	BARBADOS	1960	1989	3.471	3.619	2.257	3.585	-2.722	-1.572	1.704	5.534
54	CANADA	1950	1990	2.762	2.503	2.849	2.626	-5.874	-2.255	1.447	3.559
55	COSTA RICA	1950	1990	2.310	2.363	0.049	2.355	-1.867	-2.151	0.592	4.134
57	DOMINICAN REP.	1950	1990	2.425	1.978	2.050	2.257	-9.068	-1.127	0.187	3.770
58	EL SALVADOR	1950	1990	1.063	1.009	-4.320	1.011	-0.813	-1.965	-1.430	3.449
60	GUATEMALA	1950	1990	1.239	0.790	-8.392	0.794	-0.318	-0.985	-1.021	2.602
61	HAITI	1960	1989	0.147	-0.331	0.083	-0.081	-6.405	-1.177	-1.881	1.018
62	HONDURAS	1950	1990	1.101	0.788	0.860	0.886	-3.692	-1.664	-0.755	2.332
63	JAMAICA	1953	1989	1.420	2.020	13.278	2.020	0.732	-1.810	-0.845	4.885
64	MEXICO	1950	1990	2.536	2.258	1.738	2.335	-3.248	-1.718	0.374	4.144
65	NICARAGUA	1950	1987	1.021	0.943	-11.828	0.946	-0.869	-1.188	-2.454	4.339
66	PANAMA	1950	1990	2.821	2.181	1.626	2.331	-2.813	-1.202	-0.101	4.464
67	PUERTO RICO	1955	1989	3.649	3.930	0.193	3.911	-1.123	-1.902	1.800	6.080
70	TRINIDAD&TOBAG	1950	1990	2.870	2.598	-4.613	2.615	-1.154	-0.750	-0.037	5.228
71	U.S.A.	1950	1990	1.940	1.894	1.958	1.928	-11.488	-2.778	0.988	2.800
72	ARGENTINA	1950	1990	0.922	0.386	-1.123	0.453	-2.034	-0.307	-1.433	2.184
73	BOLIVIA	1950	1990	1.317	0.632	0.498	0.711	-1.823	-1.319	-1.149	2.412
74	BRAZIL	1950	1990	3.489	2.858	-1.322	2.881	-0.778	-0.813	0.868	5.049
75	CHILE	1950	1990	0.925	1.234	0.801	1.038	-9.391	-3.087	-1.431	3.890
76	COLOMBIA	1950	1990	2.146	1.927	2.202	2.017	-4.963	-1.807	0.873	2.881
77	ECUADOR	1950	1990	2.751	2.165	0.848	2.217	-1.364	-1.174	-0.098	4.428
78	GUYANA	1950	1990	-0.218	-0.998	-0.763	-0.702	-4.521	-1.592	-4.508	2.513
79	PARAGUAY	1950	1990	2.068	1.407	2.638	1.598	-3.378	-2.082	-0.518	3.331
80	PERU	1950	1990	1.406	0.886	6.271	0.888	1.957	-0.187	-2.088	3.859
81	SURINAME	1960	1989	1.398	0.418	-21.105	0.437	-0.540	-0.232	-4.127	4.983
82	URUGUAY	1950	1990	0.372	0.579	0.251	0.437	-10.910	-2.907	-1.442	2.800
83	VENEZUELA	1950	1990	0.439	0.549	6.766	0.549	1.237	-1.338	-1.383	2.491
85	BANGLADESH	1959	1990	1.208	1.392	1.183	1.281	-10.885	-2.813	-1.218	4.000
87	CHINA	1968	1990	5.752	5.984	5.556	5.884	-4.728	-2.310	3.281	8.707
88	HONG KONG	1960	1990	6.264	6.250	6.051	6.261	-11.705	-3.896	5.032	8.881
89	INDIA	1950	1990	1.437	1.794	1.655	1.608	-6.794	-1.473	0.382	3.207
90	INDONESIA	1960	1990	4.471	3.779	-48.723	3.779	0.097	-3.505	1.785	5.773
91	IRAN	1955	1989	1.913	1.528	-55.194	1.528	-0.233	-1.188	-3.875	6.730
92	IRAQ	1953	1987	1.767	0.478	-2.787	0.801	-2.956	-0.791	-5.444	6.397
93	ISRAEL	1953	1990	3.620	3.637	10.801	3.637	0.928	-1.004	1.882	5.583
94	JAPAN	1950	1990	5.781	5.742	14.077	5.742	0.728	-0.743	3.744	7.741
95	JORDAN	1954	1990	3.589	3.110	1.898	3.278	-4.103	-1.845	-0.194	6.415
96	KOREA, REP.	1953	1989	5.944	5.692	-387.470	5.69	0.017	-1.908	3.520	7.864
98	MALAYSIA	1955	1990	4.264	3.871	4.547	4.054	-5.685	-2.818	1.397	6.345
101	MYANMAR	1950	1989	2.411	2.563	2.318	2.448	-13.486	-2.110	0.823	4.503
102	NEPAL	1960	1986	1.924	1.547	2.127	1.800	-9.233	-2.342	-1.244	4.337
104	PAKISTAN	1950	1990	2.355	2.126	2.614	2.234	-5.898	-2.323	0.552	3.700
105	PHILIPPINES	1950	1990	2.001	2.073	-7.998	2.072	-0.459	-2.569	-0.181	4.328
108	SINGAPORE	1960	1990	6.724	6.190	6.347	6.287	-2.224	-1.638	3.531	8.850
109	SRI LANKA	1950	1989	1.854	1.838	2.454	1.844	-4.102	-1.325	0.637	3.039
110	SYRIA	1960	1990	3.702	3.222	2.958	3.489	-7.048	-0.608	0.386	6.057

Table 4 (Continued)
Annual Real Per-Capita Growth Rates

ID	Country	Smpl Period		$\hat{\beta}_{OLS}$	$\hat{\beta}_{FD}$	$\hat{\beta}_{FCO}$	$\hat{\beta}_{PW}$	$\hat{c}$	$\hat{\tau}^{\tau}$	β_{min}	β_{max}
111	TAIWAN	1951	1990	5.653	5.603	6.650	5.613	-2.371	-2.502	4.288	8.018
112	THAILAND	1950	1990	3.822	3.570	37.981	3.571	-0.187	-3.614	1.080	5.151
114	YEMEN	1969	1989	4.727	5.676	3.739	5.329	-3.025	-2.083	2.050	8.402
115	AUSTRIA	1950	1990	3.640	3.664	25.762	3.664	0.176	-0.967	2.578	4.750
116	BELGIUM	1950	1990	2.808	2.767	2.788	2.803	-3.043	-1.366	1.705	3.739
118	CYPRUS	1950	1990	3.970	4.098	3.894	3.894	-16.347	-3.733	2.363	5.788
119	CZECHOSLOVAKIA	1960	1990	3.315	3.041	-0.394	3.081	-1.179	-0.685	1.121	4.861
120	DENMARK	1950	1990	2.644	2.412	2.378	2.487	-3.843	-1.285	1.322	3.501
121	FINLAND	1950	1990	3.434	3.452	3.249	3.441	-8.063	-2.662	2.075	4.828
122	FRANCE	1950	1990	3.080	3.008	11.823	3.008	0.351	-0.555	2.014	4.002
123	GERMANY, WEST	1950	1990	3.189	3.576	5.273	3.578	2.718	-2.881	2.174	4.878
124	GREECE	1950	1990	4.328	3.887	18.217	3.887	0.318	0.130	2.369	5.404
125	HUNGARY	1970	1990	2.234	2.322	8.202	2.322	1.402	-0.836	0.475	4.169
126	ICELAND	1950	1990	3.422	2.969	3.322	3.280	-10.813	-3.143	1.004	4.833
127	IRELAND	1950	1990	3.207	3.102	3.445	3.151	-5.801	-2.702	1.716	4.489
128	ITALY	1950	1990	3.752	3.749	14.779	3.749	0.359	-0.848	2.705	4.783
129	LUXEMBOURG	1950	1990	2.185	2.246	2.287	2.198	-14.278	-3.110	0.843	3.849
130	MALTA	1954	1989	5.496	5.024	8.840	5.104	-2.148	-1.848	3.072	6.878
131	NETHERLANDS	1950	1990	2.763	2.588	1.804	2.611	-1.785	-1.467	1.222	3.854
132	NORWAY	1950	1990	3.346	3.051	3.125	3.138	-3.480	-2.070	2.114	3.988
133	POLAND	1970	1990	0.694	1.242	-10.354	1.224	-0.726	-2.785	-11.785	14.280
134	PORTUGAL	1950	1990	4.320	4.213	2.890	4.228	-1.885	-1.250	2.450	5.875
136	SPAIN	1950	1990	3.786	3.998	-8.252	3.885	-0.386	-0.888	2.287	5.698
137	SWEDEN	1950	1990	2.375	2.312	0.048	2.314	-0.817	-1.170	1.478	3.148
138	SWITZERLAND	1950	1990	2.083	2.210	1.174	2.189	-2.870	-1.690	0.818	3.519
139	TURKEY	1950	1990	2.746	3.144	2.460	2.875	-10.589	-2.534	1.511	4.778
140	U.K.	1950	1990	2.241	2.306	2.249	2.253	-18.815	-4.505	2.072	2.438
141	U.S.S.R.	1970	1989	3.272	3.377	-9.827	3.376	-0.215	-1.981	2.878	4.074
142	YUGOSLAVIA	1960	1990	3.630	2.812	13.477	2.812	1.129	0.919	0.236	5.388
143	AUSTRALIA	1950	1990	2.184	1.870	2.158	2.086	-10.873	-2.065	0.752	2.987
144	FIJI	1960	1990	2.043	2.006	1.853	2.021	-4.629	-1.516	-0.302	4.405
145	NEW ZEALAND	1950	1990	1.674	1.388	1.599	1.559	-8.353	-2.021	0.035	2.742
146	PAPUA N.GUINEA	1960	1990	0.215	0.643	3.106	0.643	3.420	-2.689	-1.229	2.515

Notes: The column labeled ID shows the country ID from the Penn World Tables. The estimators $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$, $\hat{\beta}_{FCO}$, $\hat{\beta}_{PW}$ are described in the text; $\hat{c}$ is an estimate of the local-to-unity paramater, constructed as $T(\hat{\rho}-1)$; $\hat{\tau}^{\tau}$ is the augmented Dickey-Fuller t-statistic; β_{min} and β_{max} are the endpoints of the 95% confidence interval for β constructed using the Prais-Winsten estimator, as described in the text.