ESTIMATING DETERMINISTIC TRENDS IN THE PRESENCE OF SERIALLY CORRELATED ERRORS

Eugene Canjels and Mark W. Watson*

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.

I. 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 the 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 (BUE) estimators of α and β can be constructed by generalized least squares (GLS) and statistical inference can be conducted using standard regression procedures. In practice, the distribution of the errors and the pattern of serial correlation is unknown, so that GLS estimation and exact inference are infeasible.

Received for publication February 28, 1992. Revision accepted for publication September 5, 1995.

This is a substantially revised and expanded version on "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 benefited from discussions and comments by three referees, Edwin Denson, Michael Horvath, Robert King, James MacKinnon, Peter Phillips, Gauri Prakash, and James Stock. Financial support was provided by the National Science Foundation through grants SES-91-22463 and SBR-94-09629.

* New School for Social Research and Princeton University, respectively.

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 ordinary least-squares (OLS) estimators of α and β are asymptotically equivalent to the GLS estimator. 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, whereas 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 be 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 cannot 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 examplified by Dickey and Fuller (1979), Durlauf and Phillips (1988), and Elliott et al. (1996). 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 independently identically distributed (i.i.d.).[1] There are few exact analytic results in this literature because these depend on the specific proper-

[1] There is large literature on this topic, including Beach and MacKinnon (1978), Chipman (1979), Kadiyala (1968), Maeshiro (1976, 1979), Magee

ties of the regressors and because of complications from 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 extensively 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 of 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 II by presenting results on the asymptotic distributions of estimators of β. These include the OLS, first-difference and infeasible GLS estimators, and three different, but commonly used, feasible GLS estimators. We avoid the sharp $|\rho| < 1$ versus $\rho = 1$ 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 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 close to one. 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 II concludes with a finite-sample experiment that indicates that the asymptotics provide reasonable approximations to the finite-sample relative efficiencies.

Section III 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 narrow 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 II to construct bounds tests and conservative confidence intervals building on methods developed in Dufour (1990) and Cavanagh et al. (1995).

In section IV we apply the methods to estimate and construct confidence intervals for real per-capita gross domestic product (GDP) growth rates for 128 countries using postwar data. Consistent with the analysis in section II, 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 the small sample size lead to wide confidence intervals for β.

We offer a summary and some conclusions in section V, and the appendix contains proofs and other detailed calculations.

II. Estimators

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

1. The data y_t are generated by $y_t = \alpha + \beta t + u_t$ for $t = 1, \ldots, T$.
2. The error term u_t is generated by $(1 - \rho_T L)u_t = v_t$ for $t = 2, \ldots, T$.
3. $u_1 = \Sigma_{i=0}^{[\kappa T]} \rho_T^i v_{1-i}$.
4. $v_t = d(L)\epsilon_t$, with $d(L) = \Sigma_{i=0}^{\infty} d_i L^i$, and $\Sigma_{i=0}^{\infty} i|d_i| < \infty$.
5. The error term ϵ_t is a martingale difference sequence with $E(\epsilon_t^2 | \epsilon_{t-1}, \epsilon_{t-2}, \ldots) = 1$ 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 $\rho_T = (1 + c/T)$, 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 than the exact unit root process.[4] We will refer to processes with $\rho_T = (1 + c/T)$ and either $c = 0$ or $c \neq 0$ as $I(1)$ processes.

(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 an intercept is excluded from equation (1) and v_t is i.i.d., and calculates the relative efficiency of the OLS and first-difference estimators as a function of ρ and the sample size; Chipman (1979) includes an intercept in equation (1) 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 IIA.

[3] Two notable exceptions are Durlauf and Phillips (1988), which is discussed in more detail in section IIA, and Lee and Phillips (1994), which studies asymptotic properties of trend estimators in the model without an intercept and with error processes assumed to follow either local-to-unity or long-memory processes.

[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 Bobkosky (1983), Cavanagh (1985), Cavanagh et al. (1995), Chan and Wei (1987), Chan (1988), Phillips (1987), and Stock (1991).

Assumption 3 incorporates a range of assumptions about the initial condition u_1, depending on the values 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., $T^{-1/2} \sum_{t=1}^{[sT]} \epsilon_t \Rightarrow W(s)$, where $W(s)$ is a standard Wiener process.[6] Assumption 4 ensures that the functional central limit theorem also applies to the partial sums of v_t, specifically $T^{-1/2} \sum_{t=1}^{[sT]} v_t \Rightarrow d(1)W(s)$.

A. *OLS, First-Difference, and GLS Estimators*

Let $\hat{\beta}_{OLS}$ denote the OLS estimator of β in equation (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 nonzero ρ_T. Specifically, $\hat{\beta}_{GLS}$ is the OLS estimator in the transformed regression

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

together with

$$\sigma_T^{-1} y_1 = \sigma_T^{-1}\alpha + \sigma_T^{-1}\beta + \sigma_T^{-1} u_1 \tag{4}$$

where $\sigma_T^2 = (1 - \rho_T^{2([\kappa T]+1)})/(1-\rho_T^2)$ for $|\rho_T| \neq 1$ and $\sigma_T^2 = [\kappa T] + 1$ for $|\rho_T| = 1$. For simplicity, the GLS estimator ignores the $I(0)$ serial correlation associated with $d(L)$. 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 (1954) and 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) $T^{3/2}(\hat{\beta}_{OLS} - \beta) \xrightarrow{L} N(0, V_1)$, 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 $V_2 = \sum_{i=0}^{\infty} f_i^2 + \text{var}(u_1)$, 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 nonnormal.

(c) $T^{3/2}(\hat{\beta}_{GLS} - \beta) \xrightarrow{L} N(0, V_1)$, where V_1 is specified in (a).

Proof: Parts (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. □

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 $\rho_T = (1 + c/T)$:

(a) $T^{1/2}(\hat{\beta}_{OLS} - \beta) \xrightarrow{L} N(0, R_1)$, where

$$R_1 = d(1)^2 c^{-5}[18(c-2)^2 e^{2c} + 72c(c-2)e^c + 12c^3 + 54c^2 + 72c - 72] + d(1)^2 144 S_c(\kappa)\left[\frac{ce^c + c - 2(e^c - 1)}{2c^2}\right]^2.$$

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

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

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

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

Proof: See the appendix.

The limiting behavior of the estimators in the exact unit root model follows from evaluating the limiting values of R_1, R_2, and R_3 in Theorem 2 as $c \to 0$. These results are given in:

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) $T^{1/2}(\hat{\beta}_{OLS} - \beta) \xrightarrow{L} N(0, \frac{6}{5}d(1)^2)$.

(b) $T^{1/2}(\hat{\beta}_{FD} - \beta) \xrightarrow{L} N(0, d(1)^2)$.

(c) $T^{1/2}(\hat{\beta}_{GLS} - \beta) \xrightarrow{L} N(0, d(1)^2)$.

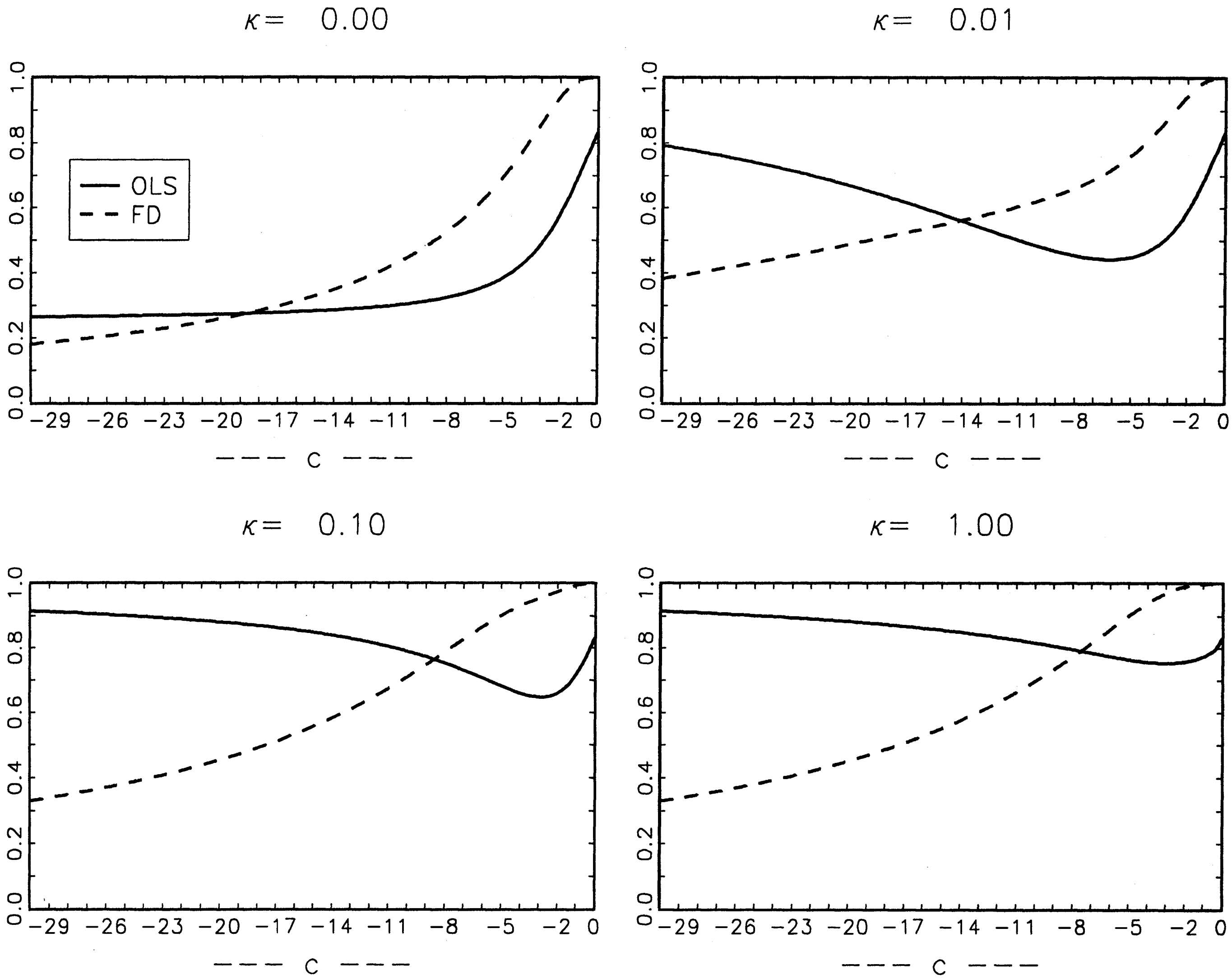
We highlight five features of these results. First, $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$, and $\hat{\beta}_{GLS}$ converge to β faster in the $I(0)$ model than in the $I(1)$ model. This result obtains because the variance of the errors is bounded in the $I(0)$ model and increases linearly with t in 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 $T(\hat{\beta}_{FD} - \beta) = [T/(T-1)] \times (u_T - u_1)$, 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.)

[5] See Elliott (1993) for a 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 a discussion.

FIGURE 1.—ASYMPTOTIC RELATIVE EFFICIENCIES OF $\hat{\beta}_{OLS}$ AND $\hat{\beta}_{FD}$



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 OLS estimators 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 a relative 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 $-18.6 \leq c \leq 1.2$, and $\hat{\beta}_{OLS}$ dominates $\hat{\beta}_{FD}$ for c outside this range. When $\kappa = 1.0$, the range narrows to $-7.6 \leq c \leq 0.9$.

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 et al. (1996) in the context of unit root tests. On the other hand, when $\kappa > 0$, so that u_1 is $O_p(T^{1/2})$, 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 $\kappa \rightarrow \infty$, the minimum relative efficiency of $\hat{\beta}_{OLS}$ is 0.7538 and occurs at

$c = -3.076$, a result that was also derived by Chipman (1979) using methods different from those employed here.[7,8]

Finally, when the errors are $I(1)$, the variances of $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$, and $\hat{\beta}_{GLS}$ are monotonically increasing in 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 $\hat{\beta}_{GLS}$.

B. 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 (3). 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 et al. (1996)) and is referred to as the "conditional case." Thus, $\hat{\beta}_{CC}$ is the OLS estimator of β from equation (3) together with

$$y_1 = \alpha + \beta + u_1. \tag{5}$$

Finally, let $\hat{\beta}_{PW}$ denote the Prais–Winsten (1954) estimator; that is, the OLS estimator of β from equation (3) together with

$$(1-\rho_T^2)^{1/2}y_1 = (1-\rho_T^2)^{1/2}\alpha + (1-\rho_T^2)^{1/2}\beta + (1-\rho_T^2)^{1/2}u_1. \tag{6}$$

The Prais–Winsten estimator is defined for $\rho_T \le 1$ (equivalently $c \le 0$), 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. 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 $\rho_T = (1 + c/T)$:

(a) $T^{1/2}(\hat{\beta}_{CO} - \beta) \xrightarrow{L} N(0, G_1)$, where

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

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

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

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

(c) For $c \le 0$ (so that $\hat{\beta}_{PW}$ is defined), $T^{1/2}(\hat{\beta}_{PW} - \beta) \xrightarrow{L} N(0, G_3)$, where

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

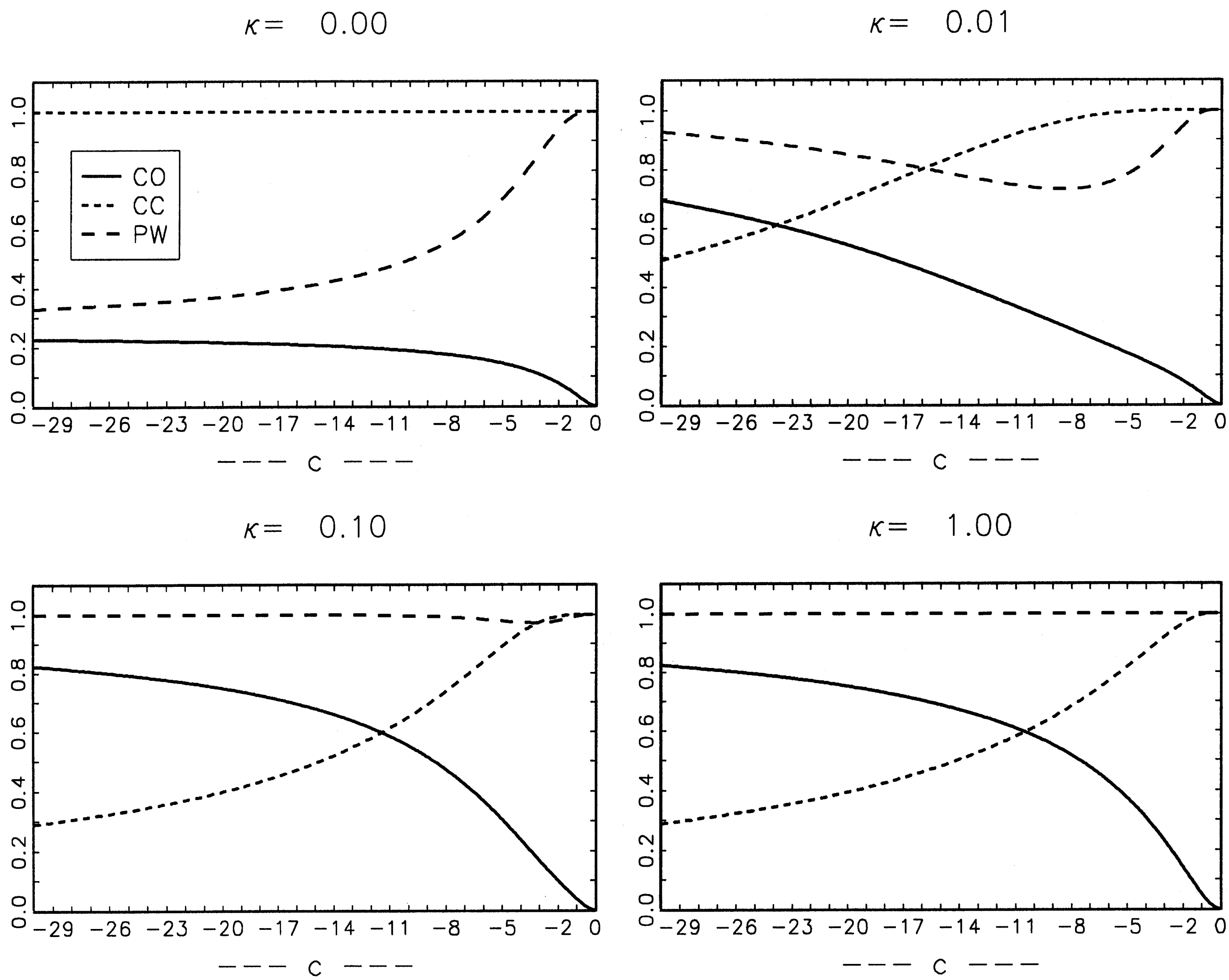
Proof: See the appendix.

Part (a) of the lemma implies that G_1, the limiting variance of $T^{1/2}(\hat{\beta}_{CO} - \beta)$, 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 (3) means that $\hat{\beta}_{CO}$ provides a poor estimate of β. When $c = 0$, α disappears from equation (3), and so this source of variance disappears from $\hat{\beta}_{CO}$. Figure 2 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

[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 $T \ge 2$. Because of a slight numerical error in Chipman's paper, his reported numerical results are different from those reported here. (Specifically, the value of c that we report ($c = -3.076$) is a more accurate estimate of the root to his polynomial (3.3) than the value reported in his paper ($c = -3.095$).)

[8] After we had finished this paper, Peter Phillips pointed out an interesting feature of our figure 1. The plots suggest that as $c \to -\infty$, the relative efficiency of the OLS estimator does not converge to 1, as one might expect from the asymptotic efficiency of OLS in the $I(0)$ model. This feature is discussed in detail in Phillips and Lee (1996), where it is shown that the result depends on how quickly c is allowed to approach $-\infty$ as T grows large.

FIGURE 2.—ASYMPTOTIC RELATIVE EFFICIENCIES OF $\hat{\beta}_{CO}$, $\hat{\beta}_{CC}$, AND $\hat{\beta}_{PW}$



the poor performance of the Cochrane–Orcutt estimator with trending regressors and ρ close to unity.[9]

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 2, $\hat{\beta}_{PW}$ is approximately efficient even when κ is very small. For example, for $\kappa = 0.01$ the relative efficiency of $\hat{\beta}_{PW}$ is larger than 0.73 for all values of c; for $\kappa = 0.05$ the relative efficiency is larger than 0.92; and for all values of $\kappa \geq 0.10$ $\hat{\beta}_{PW}$ 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.

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 (3) and (6). These estimators will be denoted by $\hat{\beta}_{FCO}$, $\hat{\beta}_{FCC}$, and $\hat{\beta}_{FPW}$. 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 $\hat{\beta}_{FCO}$ the nonlinear 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 (1) together with assumption 2 can be combined as

$$y_t = a + bt + \rho_T y_{t-1} + v_t, \qquad \text{for } t = 2, 3, \ldots, T \qquad (7)$$

where $a = \alpha(1 - \rho_T) + \beta\rho_T$ and $b = \beta(1 - \rho_T)$. Thus $\hat{\beta}_{FCO}$ can be formed from the OLS estimators from equation (7) as $\hat{\beta}_{FCO} = \hat{b}/(1 - \hat{\rho}_T)$ for $\hat{\rho}_T \neq 1$ and $\hat{\beta}_{FCO} = \hat{a}$ for $\hat{\rho}_T = 1$, where $\hat{a}$, $\hat{b}$, and $\hat{\rho}_T$ are the OLS estimators of the coefficients in equation (7). Equivalently, $\hat{\beta}_{FCO}$ can be constructed as the

[9] See Prais and Winsten (1954), Maeshiro (1976, 1979), Beach and MacKinnon (1978), Park and Mitchell (1980), Thornton (1987), and Davidson and MacKinnon (1993, sec. 10.6).

OLS estimator of β in equation (3), using $\hat{\rho}_T$ in place of ρ_T. Since the asymptotic distribution $T(1 - \hat{\rho}_T)$ is readily deduced when $\rho_T = (1 + c/T)$, (see Stock (1991), for example), the asymptotic distribution of $T^{1/2}(\hat{\beta}_{FCO} - \beta)$ can also be readily deduced.

The problem is more complicated when analyzing $\hat{\beta}_{FCC}$ and $\hat{\beta}_{FPW}$ because these estimators are generally based on iterative schemes for estimating ρ_T, α, and β. Since the limiting distribution of $\hat{\rho}_T$ depends in important ways on the precise way the data are "detrended" (for example, see Schmidt and Phillips (1992) and Elliott et al. (1996)), the limiting distribution of $\hat{\beta}_{FCC}$ and $\hat{\beta}_{FPW}$ will depend on the precise specification of the iterations. Rather than present results for specific versions of these estimators, we present limiting representations of $\hat{\beta}_{FCC}$ and $\hat{\beta}_{FPW}$ written as functions of $\hat{c}$, the limiting value of $T(\hat{\rho}_T - 1)$. Different estimators of ρ_T will lead to different limiting random variables $\hat{c}$ 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 $\hat{\beta}_{FCO}$ when $c = 0$ and $\hat{c}$ 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 equation (1) is $\hat{v}_t = u_t - \hat{\rho}_T u_{t-1}$, and the limiting values of the feasible GLS estimators can be written in terms of partial sums of $\hat{v}_t$ and the initial condition u_1. In the appendix we show that $T^{-1/2}u_1 \Rightarrow \tilde{W}_c(\kappa) \sim N(0, S_c(\kappa))$, where $S_c(\kappa)$ is defined in Theorem 2; we also show that $T^{-1/2}\sum_{t=1}^{[sT]} \hat{v}_t \Rightarrow \hat{W}(s)$, where $\hat{W}(s)$ is a functional of $W(s)$ and $\tilde{W}_c(\kappa)$.

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, $\rho_T = (1 + c/T)$, and $T(\hat{\rho}_T - 1) \Rightarrow \hat{c}$ jointly with $T^{-1/2}\sum_{t=1}^{[sT]} \epsilon_t \Rightarrow W(s)$. Then:

(a) $T^{1/2}(\hat{\beta}_{FCO} - \beta) \Rightarrow \hat{c}^{-1}12 \int_0^1(\frac{1}{2} - s)d\hat{W}(s)$

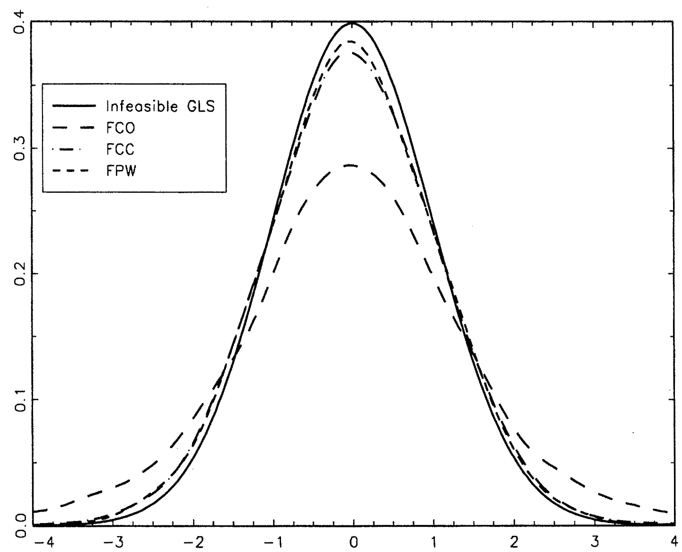
(b) $T^{1/2}(\hat{\beta}_{FCC} - \beta) \Rightarrow (1 - \hat{c} + \frac{1}{3}\hat{c}^2)^{-1}[(\hat{c} - \frac{1}{2}\hat{c}^2)\tilde{W}_c(\kappa) - \int_0^1 (\hat{c}s - 1)d\hat{W}(s)]$

(c) $T^{1/2}(\hat{\beta}_{FPW} - \beta) \Rightarrow [1 - \frac{1}{2}\hat{c} + \frac{1}{12}\hat{c}^2]^{-1} [\hat{c}\tilde{W}_c(\kappa) - \int_0^1 (1 + \frac{1}{2}\hat{c} - \hat{c}s)d\hat{W}(s)]$.

Proof: See the appendix.

This theorem allows us to offer practical advice about the choice of estimators. First, notice that $\hat{c}$ appears in the denominator of the limiting representation of $T^{1/2}(\hat{\beta}_{FCO} - \beta)$. 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

FIGURE 3.—DENSITIES OF FEASIBLE GLS ESTIMATORS



$(1 - \frac{1}{2}\hat{c} + \frac{1}{12}\hat{c}^2) > 0$ for all values of $\hat{c}$. This can be seen in figure 3, which plots the limiting probability densities of $T^{1/2}(\hat{\beta}_{FCO} - \beta)$, $T^{1/2}(\hat{\beta}_{FCC} - \beta)$, and $T^{1/2}(\hat{\beta}_{FPW} - \beta)$ for the case with $c = 0$, $\kappa = 1$, and $d(1) = 1$.[10] 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 $|T^{1/2}(\hat{\beta}_{FCO} - \beta)|$ exceeds 2 is approximately 20%, while the corresponding values for $\hat{\beta}_{FCC}$ and $\hat{\beta}_{FPW}$ are approximately 5%. Figure 3 suggests that little is lost using either $\hat{\beta}_{FCC}$ or $\hat{\beta}_{FPW}$ 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.[11] 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. (Results for $\hat{\beta}_{FCO}$ are not included because of the estimator's poor performance.) The last row of the table shows results for a "pretest" estimator ($\hat{\beta}_{PT}$) constructed from the OLS and FD estimators. Figure 1 provides the

[10] 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 equation (5). The Prais–Winsten estimator used $\min(1, \hat{\rho}_T)$.

[11] These MSEs were estimated using the simulations described in footnote 10.

TABLE 1.—AVERAGE MEAN-SQUARE ERROR OF ESTIMATORS

	κ					
Estimator	0.000	0.010	0.050	0.100	0.250	1.000
			$-30 \le c \le 0$			
$\hat{\beta}_{GLS}$	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
			$-2 \le c \le 0$			
$\hat{\beta}_{GLS}$	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
			$-10 \le c \le -2$			
$\hat{\beta}_{GLS}$	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
			$-30 \le c \le -10$			
$\hat{\beta}_{GLS}$	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

Note: Entries are mean-square error averaged over range of c indicated.

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 pretest 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 $\hat{c} \ge \bar{c}$, where $\bar{c}$ is the prespecified 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, $-2 \le c \le 0$), $\hat{\beta}_{FD}$ 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 0.75. The pretest estimator performs well, and is 1 to 2.5% less efficient than $\hat{\beta}_{FD}$, depending on the value of κ.

(iii) For values of c in the range $-10 \le c \le -2$, the relative performance of the estimators depends critically on the variance 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 $\kappa \ge 0.10$. When $\kappa = 0.05$ the feasible GLS estimators and $\hat{\beta}_{FD}$ are comparable.

(iv) For values of $-30 \le c \le -10$ and when $\kappa = 0$, $\hat{\beta}_{FCC}$ is the preferred estimator. When $\kappa \ge 0.05$, 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 $\kappa \ge 0.05$.

(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 when $c \le 0$, $\hat{\beta}_{FPW}$ is the most robust estimator, with a mean-squared error 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 is close to 0, but somewhat worse performance for large negative c.

C. *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. We now 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 $d(L) = d$, $\epsilon_t \sim NIID(0, 1)$, for various values of T, ρ, and for $\kappa = 1$.[12] (Results for $\kappa = 0$ are given in Canjels and Watson (1994).) 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 no difference 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 pretest 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 $\rho =$

[12] The mean-squared errors for $\hat{\beta}_{FCC}$, $\hat{\beta}_{FPW}$, and $\hat{\beta}_{PT}$ were estimated using 10,000 Monte Carlo draws, using $\hat{\rho} = \Sigma_{t=2}^{T} \hat{u}_t\hat{u}_{t-1}/\Sigma_{t=2}^{T-1} \hat{u}_t^2$, 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).

TABLE 2.—RELATIVE EFFICIENCIES OF ESTIMATORS: EXACT AND $I(1)$ APPROXIMATION
$\kappa = 1.0$

	$T = 30$		$T = 50$		$T = 100$	
	Exact	$I(1)$	Exact	$I(1)$	Exact	$I(1)$
			$\rho = 0.50$			
$\hat{\beta}_{OLS}$	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
			$\rho = 0.80$			
$\hat{\beta}_{OLS}$	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
			$\rho = 0.90$			
$\hat{\beta}_{OLS}$	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
			$\rho = 0.95$			
$\hat{\beta}_{OLS}$	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
			$\rho = 1.00$			
$\hat{\beta}_{OLS}$	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: Relative efficiency is the ratio of the variance of the infeasible GLS estimator to the variance of the estimator given in column 1. 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, respectively, for the estimators in column 1 and for all T and $|\rho| < 1$.

0.5. The predictions based on the $I(1)$ asymptotic relative efficiencies are off the mark. The $I(1)$ asymptotics suggest that $\hat{\beta}_{FCC}$ is strongly dominated by both $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FPW}$ when $\kappa = 1$ and strongly dominate the other estimators when $\kappa = 0$ (not shown). 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 considered (0.8, 0.9, 0.95, 1.0) the rankings implied by the $I(1)$ asymptotic relative efficiencies are more accurate than the $I(0)$ rankings. Indeed in all but one of the cases studied in the table, 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.

III. Confidence Intervals

When $\rho < 1$ (so that the errors are $I(0)$), confidence intervals for β 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 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 their nominal size than the $I(0)$ approximations. Unfortunately the method is also more complicated because the asymptotic distribution of the estimators depends on the nuisance parameters c and κ, and these parameters cannot be consistently estimated from the data. Hence 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 et al. (1995).

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 a prespecified constant. Then the Bonferoni confidence interval $\cup_{0 \leq \kappa \leq \bar{\kappa}} \cup_{c \in C_\kappa} B_\kappa(c)$ 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 by 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 $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$, $\hat{\beta}_{CC}$, or $\hat{\beta}_{PW}$, and their asymptotic variances given in Theorem 2 and Lemma 4. (These variances require $d(1)$, which can be consistently estimated using standard spectral estimators.) The set C_κ, the marginal confidence intervals for c, can be constructed using the methods developed in Stock (1991).[13]

In general this procedure is quite demanding. For each $\kappa \in [0, \bar{\kappa}]$, 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

[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 $\hat{\tau}^\tau$. Under the assumption that $\kappa = 0$, Stock shows that $\hat{\tau}^\tau \Rightarrow (\int_0^1 W_c^\tau(s)^2 ds)^{1/2}[c + \int_0^1 W_c^\tau(s) dW(s) / (\int_0^1 W_c^\tau(s)^2 ds)]$, 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 $T^{-1/2} u_{[sT]} \Rightarrow d(1) W_c(s)$ when $\kappa = 0$. As shown in the appendix, when $\kappa \neq 0$, $T^{-1/2} u_{[sT]} \Rightarrow d(1)[W_c(s) + e^{sc} \tilde{W}_c(\kappa)]$, where $\tilde{W}_c(\kappa) \sim N(0, S_c(\kappa))$ and is independent of $W_c(s)$. Using this result, it is straightforward to show that all of Stock's analysis continues to hold, with $W_c(s) + e^{sc} \tilde{W}_c(\kappa)$, replacing $W_c(s)$ in the above limiting representation for $\hat{\tau}^\tau$.

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 $\cup_{c\in C_\kappa} B_\kappa(c) = B_\kappa(\bar{c})$, 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 we performed suggest that $\cup_{c\in C_\kappa} B_\kappa(c) \approx B_\kappa(\bar{c})$ 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 $\cup_{0\le\kappa\le\bar{\kappa}}\cup_{c\in C_\kappa} B_\kappa(c) \approx B_\infty(\bar{c})$, where $\bar{c} = \sup_c\{c \in C_0\}$. Thus approximate $100(1 - \alpha_1 - \alpha_2)\%$ confidence intervals can be formed by (1) choosing the largest value of c in the $100(1 - \alpha_2)\%$ confidence interval constructed using the procedure from Stock (1991), and (2) 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 four 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 one-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 nonrandom 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 $\bar{c} \geq -7.6$. Third, since these confidence intervals may be conservative for all values of c and κ, a tighter $(1 - \alpha)\%$ confidence interval can be constructed by choosing α_1 and α_2 so that $\sup_{c,\kappa} \Pr(\beta \notin B_{\alpha_1,\alpha_2}) = \alpha$, where B_{α_1,α_2} is the confidence interval constructed using α_1 and α_2. (See Cavanagh et al. (1995).) Finally, since the confidence intervals proposed here are not based on an efficient procedure (like inverting an efficient test statistic), it may be possible to construct smaller confidence regions with the same coverage.

Table 3 shows estimated coverage rates for confidence intervals for different values of T and c, calculated as described above using the Prais–Winsten estimator. (Results for analogous experiments using the OLS and FD estimators are reported in Canjels and Watson (1994).) The design was much the same as in section IIC, i.e., $d(L) = d$ and $\epsilon_t \sim N(0, 1)$. Results are reported for conservative 90%, 95%, and 99% confidence intervals constructed with $\alpha_1 = \alpha_2$. Results for nonsymmetric 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 y_{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. 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. Finally, since the bounds are essentially tight for $c = 0$, we did not attempt to form tighter bounds by increasing α_1 and α_2 as suggested above.

While the coverage rates are close to their nominal levels for $c = 0$, they are conservative when $c < 0$. This occurs because of the sharp increase in the variance of $\hat{\beta}_{PW}$ for small c. For example, when the true value of $c = -5$, then $c = 0$ is often in the confidence set C_0, the variance of the

TABLE 3.—CONFIDENCE INTERVAL COVERAGE RATES (PERCENT)

		c				
Level	κ	0	−1	−5	−10	−20
			$T = 30$			
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
			$T = 50$			
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
			$T = 100$			
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: 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 were constructed from the Prais–Winsten estimator.

[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 κ = 0.0, 0.5, 1.0, and 10.0, respectively. The corresponding percentiles are −3.89, −3.84, and −3.84 for $c = -5.0$; −4.20, −4.20, −4.20, and −4.20 for $c = -10.0$; and −4.52, −4.54, −4.54, and −4.54 for $c = -20.0$. These percentiles are based on 5000 simulations with $T = 500$.

TABLE 4.—ANNUAL REAL PER-CAPITA GROWTH RATES

ID	Country	Sample	Period	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FD}$	$\hat{\beta}_{FCO}$	$\hat{\beta}_{FPW}$	$\hat{c}$	$\hat{\tau}^{T}$	β_{min}	β_{max}
1	Algeria	1960	1990	2.736	1.459	3.316	1.459	1.097	0.297	−1.139	4.057
2	Angola	1960	1989	−2.038	−1.004	−2.435	−1.605	−6.722	−2.361	−6.155	4.147
3	Benin	1959	1989	−0.403	−0.370	−0.547	−0.398	−16.035	−2.975	−2.087	1.346
4	Botswana	1960	1989	5.839	6.079	6.024	5.874	−14.886	−3.266	1.735	10.423
5	Burkina Faso	1959	1990	0.859	0.006	1.434	0.601	−9.511	−2.585	−1.905	1.917
6	Burundi	1960	1990	0.567	−0.425	0.963	0.464	−17.725	−3.998	−1.010	1.294
7	Cameroon	1960	1990	2.698	1.888	2.429	2.387	−7.469	−2.005	−1.076	4.852
8	Cape Verde Is.	1960	1989	3.880	3.338	4.402	3.634	−6.118	−2.090	−0.705	7.381
9	Central Afr. Rep.	1960	1990	−0.486	−0.589	−0.872	−0.542	−4.890	−1.514	−2.235	1.058
10	Chad	1960	1990	−2.584	−2.010	−2.579	−2.440	−10.915	−2.355	−5.521	1.502
11	Comoros	1960	1987	−0.044	0.520	−0.423	0.087	−10.927	−2.945	−2.721	3.762
12	Congo	1960	1990	3.314	2.788	3.366	3.244	−15.959	−4.247	2.017	4.323
14	Egypt	1950	1990	3.002	2.385	3.148	2.821	−10.774	−2.806	0.666	4.104
15	Ethiopia	1951	1986	0.831	0.669	0.675	0.756	−6.363	−1.562	−0.224	1.562
16	Gabon	1960	1990	2.298	2.620	−2.668	2.544	−2.499	−1.283	−1.142	6.381
17	Gambia	1960	1990	1.150	0.935	0.302	1.048	−5.896	−2.039	−2.654	4.524
18	Ghana	1955	1989	−0.266	−0.089	−0.408	−0.212	−9.673	−2.324	−2.359	2.181
19	Guinea	1959	1989	−0.304	−0.215	−0.566	−0.263	−6.022	−1.731	−1.737	1.307
20	Guinea-Biss	1960	1990	0.377	1.036	−0.017	0.622	−7.747	−2.281	−2.744	4.817
21	Ivory Coast	1960	1990	1.073	0.633	3.638	0.633	4.003	1.635	−1.904	3.170
22	Kenya	1950	1990	1.179	1.166	1.191	1.177	−16.247	−3.124	−1.386	3.718
23	Lesotho	1960	1990	4.402	4.053	3.103	4.187	−3.965	−1.524	0.162	7.944
24	Liberia	1960	1986	0.682	0.310	−2.698	0.401	−2.492	−0.965	−2.129	2.749
25	Madagascar	1960	1990	−1.962	−1.818	−2.297	−1.898	−6.419	−1.902	−3.338	−0.298
26	Malawi	1954	1990	1.171	1.226	0.796	1.193	−7.884	−1.964	−0.544	2.996
27	Mali	1960	1990	0.877	0.150	1.304	0.668	−9.779	−3.132	−1.827	2.127
28	Mauritania	1960	1990	−0.164	−0.207	−0.551	−0.179	−8.263	−2.121	−2.822	2.407
29	Mauritius	1950	1990	1.340	1.399	2.649	1.374	−4.772	−1.515	−1.128	3.925
30	Morocco	1950	1990	2.814	2.355	2.824	2.682	−10.958	−2.839	0.017	4.692
31	Mozambique	1960	1990	−2.309	−1.426	−3.556	−1.824	−4.800	−2.009	−4.098	1.245
32	Namibia	1960	1989	0.384	0.509	−1.741	0.459	−4.186	−1.554	−2.862	3.880
33	Niger	1960	1989	−0.415	−0.256	−1.742	−0.345	−6.305	−1.823	−3.383	2.871
34	Nigeria	1950	1990	1.989	1.337	0.894	1.607	−4.640	−1.776	−3.432	6.106
35	Reunion	1960	1988	3.764	3.799	3.322	3.774	−9.077	−2.183	1.985	5.612
36	Rwanda	1960	1990	1.974	0.791	1.999	1.601	−9.008	−2.243	−3.726	5.308
37	Senegal	1960	1990	0.138	0.204	0.105	0.138	−27.906	−4.943	−0.038	0.323
38	Seychelles	1960	1989	3.896	3.449	4.064	3.791	−11.245	−2.535	1.123	5.775
39	Sierra Leone	1961	1990	0.049	0.593	−1.013	0.278	−6.686	−1.972	−1.942	3.129
40	Somalia	1960	1989	−0.448	−0.551	−0.393	−0.458	−18.231	−3.696	−4.145	3.061
41	South Africa	1950	1990	1.792	1.343	29.503	1.343	0.127	0.053	0.160	2.525
42	Sudan	1971	1990	−0.286	−0.886	−0.377	−0.514	−6.173	−2.047	−4.333	2.561
43	Swaziland	1960	1989	1.626	1.985	−1.587	1.869	−3.303	−1.185	−1.848	5.818
44	Tanzania	1960	1988	1.532	1.686	1.030	1.600	−6.272	−2.060	−0.698	4.070
45	Togo	1960	1990	1.777	1.797	0.642	1.787	−5.198	−1.980	−0.687	4.281
46	Tunisia	1960	1990	3.761	3.222	2.500	3.262	−1.076	−0.455	1.849	4.594
47	Uganda	1950	1989	−0.188	0.946	−0.016	0.198	−9.253	−1.813	−2.644	4.536
48	Zaire	1950	1989	0.339	0.648	−2.382	0.556	−3.239	−1.364	−1.645	2.941
49	Zambia	1955	1990	-0.613	−0.597	−2.191	−0.603	−3.702	−1.213	−3.212	2.017
50	Zimbabwe	1954	1990	0.904	1.018	0.795	0.950	−7.632	−2.095	−0.735	2.771
52	Barbados	1960	1989	3.471	3.619	2.616	3.570	−3.433	−1.312	1.985	5.254
54	Canada	1950	1990	2.762	2.503	2.823	2.666	−8.576	−2.347	1.538	3.468
55	Costa Rica	1950	1990	2.310	2.363	1.377	2.343	−4.074	−1.459	0.958	3.767
57	Dominican Rep.	1950	1990	2.425	1.979	1.982	2.236	−7.323	−1.495	−0.178	4.135
58	El Salvador	1950	1990	1.063	1.009	−0.048	1.027	−3.504	−1.830	−1.445	3.463
60	Guatemala	1950	1990	1.239	0.790	−0.051	0.851	−1.676	−0.934	−0.974	2.555
61	Haiti	1960	1989	0.147	−0.331	0.077	−0.090	−5.497	−1.428	−1.807	1.144
62	Honduras	1950	1990	1.101	0.788	0.841	0.936	−5.456	−1.640	−0.727	2.303
63	Jamaica	1953	1989	1.420	2.020	−1.155	1.852	−3.005	−1.724	0.083	3.958
64	Mexico	1950	1990	2.536	2.259	2.111	2.388	−5.349	−1.756	0.388	4.130
65	Nicaragua	1950	1987	1.021	0.943	−2.580	0.963	−2.771	−0.962	−2.114	3.999
66	Panama	1950	1990	2.821	2.181	1.081	2.278	−1.832	−0.596	0.264	4.099
67	Puerto Rico	1955	1989	3.649	3.930	2.829	3.819	−4.212	−2.179	2.472	5.388
70	Trinidad & Tobago	1950	1990	2.870	2.596	0.340	2.673	−3.061	−0.991	0.118	5.074
71	U.S.A.	1950	1990	1.940	1.894	1.958	1.927	−11.035	−2.471	0.963	2.825
72	Argentina	1950	1990	0.922	0.366	−1.986	0.420	−1.343	−0.427	−1.301	2.032
73	Bolivia	1950	1990	1.317	0.632	0.848	0.783	−2.480	−0.994	−0.928	2.191
74	Brazil	1950	1990	3.469	2.858	−64.844	2.859	−0.052	−0.020	1.388	4.329
75	Chile	1950	1990	0.925	1.234	0.837	0.999	−12.784	−3.004	−1.578	4.045
76	Colombia	1950	1990	2.146	1.927	2.199	2.023	−4.968	−1.540	1.020	2.834
77	Ecuador	1950	1990	2.751	2.165	1.961	2.318	−2.885	−1.165	−0.016	4.345

TABLE 4.—ANNUAL REAL PER-CAPITA GROWTH RATES (CONTINUED)

ID	Country	Sample	Period	$\hat{\beta}_{OLS}$	$\hat{\beta}_{FD}$	$\hat{\beta}_{FCO}$	$\hat{\beta}_{FPW}$	$\hat{c}$	$\hat{\tau}^{\tau}$	β_{min}	β_{max}
78	Guyana	1950	1990	−0.218	−0.998	−0.806	−0.719	−3.924	−1.191	−4.007	2.012
79	Paraguay	1950	1990	2.068	1.407	2.376	1.743	−6.037	−1.816	−0.297	3.112
80	Peru	1950	1990	1.406	0.886	10.972	0.886	0.908	0.323	−1.131	2.903
81	Suriname	1960	1989	1.398	0.418	−25.354	0.432	−0.411	−0.129	−3.034	3.869
82	Uruguay	1950	1990	0.372	0.579	0.272	0.425	−12.139	−3.119	−1.755	2.912
83	Venezuela	1950	1990	0.439	0.549	−1.673	0.517	−3.255	−1.352	−1.078	2.177
85	Bangladesh	1959	1990	1.208	1.392	1.183	1.260	−10.046	−2.332	−1.169	3.953
87	China	1968	1990	5.752	5.984	5.574	5.875	−4.545	−1.979	3.852	8.117
88	Hong Kong	1960	1990	6.264	6.250	6.185	6.264	−21.335	−4.239	5.817	6.703
89	India	1950	1990	1.437	1.794	1.674	1.617	−5.870	−1.417	0.410	3.179
90	Indonesia	1960	1990	4.471	3.779	5.042	4.223	−8.018	−3.603	2.198	5.360
91	Iran	1955	1989	1.913	1.528	−5.335	1.579	−1.633	−0.733	−1.893	4.948
92	Iraq	1953	1987	1.767	0.476	−4.640	0.695	−1.950	−0.507	−4.467	5.420
93	Israel	1953	1990	3.620	3.637	−0.773	3.636	−1.314	−0.634	2.295	4.980
94	Japan	1950	1990	5.781	5.742	0.500	5.745	−1.037	−0.816	3.693	7.792
95	Jordan	1954	1990	3.589	3.110	1.431	3.259	−3.336	−1.029	−0.296	6.516
96	Korea, Rep.	1953	1989	5.944	5.692	6.970	5.813	−5.430	−2.458	4.084	7.300
99	Malaysia	1955	1990	4.264	3.871	4.439	4.127	−8.698	−2.575	1.287	6.456
101	Myanmar	1950	1989	2.411	2.563	2.321	2.447	−12.829	−2.674	−0.045	5.171
102	Nepal	1960	1986	1.924	1.547	2.110	1.814	−9.024	−2.239	−1.348	4.441
104	Pakistan	1950	1990	2.355	2.126	2.581	2.245	−6.257	−1.792	0.701	3.552
105	Philippines	1950	1990	2.001	2.073	1.091	2.045	−4.447	−2.200	−0.286	4.432
108	Singapore	1960	1990	6.724	6.190	6.537	6.370	−3.485	−1.533	3.625	8.755
109	Sri Lanka	1950	1989	1.854	1.838	2.257	1.846	−5.516	−1.867	0.234	3.442
110	Syria	1960	1990	3.702	3.222	1.042	3.310	−2.030	−0.521	0.487	5.956
111	Taiwan	1951	1990	5.653	5.603	6.042	5.627	−5.451	−2.157	4.608	6.598
112	Thailand	1950	1990	3.922	3.570	4.202	3.871	−17.954	−3.902	3.083	4.382
114	Yemen	1969	1989	4.727	5.676	4.270	5.001	−7.905	−2.353	3.094	8.258
115	Austria	1950	1990	3.640	3.664	2.523	3.658	−3.092	−1.774	2.748	4.581
116	Belgium	1950	1990	2.908	2.767	2.783	2.801	−2.724	−1.086	1.973	3.561
118	Cyprus	1950	1990	3.970	4.098	3.994	3.995	−15.073	−2.933	1.322	6.873
119	Czechoslovakia	1960	1990	3.315	3.041	0.058	3.066	−1.208	−0.528	1.567	4.516
120	Denmark	1950	1990	2.644	2.412	2.383	2.489	−3.629	−1.319	1.360	3.463
121	Finland	1950	1990	3.434	3.452	3.198	3.442	−6.720	−1.503	2.595	4.308
122	France	1950	1990	3.080	3.008	−1.350	3.010	−0.634	−0.427	2.332	3.683
123	Germany, West	1950	1990	3.199	3.576	2.264	3.403	−5.268	−2.918	2.581	4.571
124	Greece	1950	1990	4.328	3.887	46.239	3.887	0.104	0.066	2.661	5.113
125	Hungary	1970	1990	2.234	2.322	0.634	2.285	−3.910	−1.744	0.692	3.952
126	Iceland	1950	1990	3.422	2.969	3.376	3.337	−15.386	−3.097	0.817	5.121
127	Ireland	1950	1990	3.207	3.102	3.353	3.168	−8.567	−2.509	1.665	4.540
128	Italy	1950	1990	3.752	3.749	2.120	3.750	−2.177	−1.023	2.841	4.657
129	Luxembourg	1950	1990	2.185	2.246	2.275	2.196	−15.614	−3.174	0.603	3.889
130	Malta	1954	1989	5.496	5.024	5.856	5.304	−7.376	−3.413	2.490	7.558
131	Netherlands	1950	1990	2.763	2.588	2.103	2.633	−2.851	−1.443	1.602	3.574
132	Norway	1950	1990	3.346	3.051	3.293	3.248	−9.574	−2.370	2.083	4.019
133	Poland	1970	1990	0.694	1.242	−1.421	0.953	−5.086	−2.846	−14.136	16.621
134	Portugal	1950	1990	4.320	4.213	2.940	4.229	−1.809	−0.810	2.948	5.477
136	Spain	1950	1990	3.786	3.998	1.073	3.969	−1.697	−0.941	2.584	5.411
137	Sweden	1950	1990	2.375	2.312	1.502	2.322	−1.980	−0.994	1.711	2.913
138	Switzerland	1950	1990	2.083	2.219	1.636	2.160	−4.931	−1.904	1.157	3.281
139	Turkey	1950	1990	2.746	3.144	2.492	2.860	−11.021	−2.439	1.264	5.024
140	U.K.	1950	1990	2.241	2.306	2.243	2.247	−24.016	−4.627	2.076	2.431
141	U.S.S.R.	1970	1989	3.272	3.377	2.988	3.305	−7.166	−4.378	3.071	3.503
142	Yugoslavia	1960	1990	3.630	2.812	17.556	2.812	0.715	0.306	0.309	5.315
143	Australia	1950	1990	2.184	1.870	2.150	2.072	−8.909	−2.256	0.664	3.076
144	Fiji	1960	1990	2.043	2.006	1.856	2.021	−4.244	−1.454	−0.196	4.209
145	New Zealand	1950	1990	1.674	1.388	1.601	1.561	−7.900	−2.073	0.046	2.730
146	Papua New Guinea	1960	1990	0.215	0.643	−1.454	0.445	−4.967	−2.671	−1.000	2.286

Notes: Column labeled ID shows the country ID from the Penn World tables. 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 parameter, constructed as $T(\hat{\rho}-1)$; $\hat{\tau}^{\tau}$ is the augmented Dickey–Fuller t-statistic; β_{min} and β_{max} are the end points of the 95% confidence interval for β constructed using the Prais–Winsten estimator, as described in the text.

estimators is much larger when $c = 0$ than when $c = -5$, and this leads to a wide confidence interval for β.

IV. Economic Growth Rates for 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. Since the logarithm of per-capita GDP for many of the countries is reasonably modeled by equations (1) and (2), this data set seems well suited for the methods developed above. (Of course, caution should be exercised in using the estimates for any specific country, since some countries experienced growth far different from that assumed in equations (1) and (2).)

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 4–7 present four estimates of average trend growth ($\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$, $\hat{\beta}_{FCO}$, and $\hat{\beta}_{FPW}$, respectively); column 8 shows the estimate of c used to construct the feasible GLS estimates ($\hat{c}$); column 9 shows the Dickey–Fuller unit-root test statistic ($\hat{\tau}^{\tau}$) used to construct a confidence interval for c, and columns 10 and 11 present lower and upper limits of the approximated 95% confidence interval for β constructed from the $\hat{\beta}_{PW}$ (β_{min} and β_{max}, respectively). The Prais–Winsten estimator used min(1, $\hat{\rho}_T$) as an estimator of ρ. The estimate $\hat{c}$ and the $\hat{\tau}^{\tau}$ statistic were calculated from the regression of Δy_t onto y_{t-1}, p lags of Δy_t and (1, t), where p was chosen by BIC. 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 Brazil (country 74) constructed from $\hat{\beta}_{OLS}$, $\hat{\beta}_{FD}$, and $\hat{\beta}_{FPW}$ range from 2.8% to 3.5%, while the estimate constructed from $\hat{\beta}_{FCO}$ is −65%. Indeed for 9 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, $\hat{\beta}_{OLS}$ and $\hat{\beta}_{FD}$ 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 $-1.14 \leq \beta \leq 4.06$. For Algeria, the Dickey–Fuller t-statistic is 0.297, 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 1.46% ($=\hat{\beta}_{FD}$), and this is the center of the confidence interval. The standard deviation of the annual growth rates is 7.2%; thus if the annual growth rates were serially uncorrelated, the standard deviation of the sample mean ($=\hat{\beta}_{FD} = \hat{\beta}_{PW}$) would be 1.31% ($=7.2\%/\sqrt{30}$). For Algeria, the growth rates are slightly negatively correlated and the estimated standard deviation of $\hat{\beta}_{PW}$ used to construct the confidence interval is 1.18%.

Finally, a few of the confidence intervals are quite narrow. For example, the estimated confidence interval for the United Kingdom (series 140) is $2.08 \leq \beta \leq 2.43$. This series is less persistent than most of the others, and the Dickey–Fuller t-statistic is −4.63. This leads to a confidence interval for c with an upper limit of $c = -16.8$ (corresponding to $\rho = 0.59$). Estimates of β are much more precise when $c = -16.8$ than when $c = 0$. Indeed the ratio of the asymptotic standard deviation for $\hat{\beta}_{PW}$ for $c = -16.8$ and $c = 0$ is 0.17, which approximately corresponds to the difference between the widths of the confidence intervals for β for the United Kingdom and the United States (country 71).

V. Concluding Remarks

In this paper we study the problems of estimation and inference in the linear trend 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: (1) the local-to-unity parameter that measures the persistence in the errors and (2) 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, leading 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.

REFERENCES

Beach, C. M., and J. G. MacKinnon, "A Maximum Likelihood Procedure for Regression with Autocorrelated Errors," *Econometrica* 46:1 (1978), 51–58.
Bobkosky, M. J., "Hypothesis Testing in Nonstationary Time Series," Department of Statistics, University of Wisconsin, Ph.D. Thesis (1983).
Canjels, E., and M. W. Watson, "Estimating Deterministic Trends in the Presence of Serially Correlated Errors," NBER Technical Working Paper 165 (1994).
Cavanagh, C. L., "Roots Local to Unity," Department of Economics, Harvard University, Manuscript (1985).
Cavanagh, C., G. Elliott, and J. Stock, "Inference in Models with Nearly Nonstationary Regressors," *Econometric Theory* 11:5 (1995), 1131–1147.
Chan, N. H., "On Parameter Inference for Nearly Nonstationary Time Series," *Journal of the American Statistical Association* 83 (1988), 857–862.
Chan, N. H., and C. Z. Wei, "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *Annals of Statistics* 15 (1987), 1050–1063.

Chipman, John S., "Efficiency of Least-Squares Estimation of Linear Trend when Residuals Are Autocorrelated," *Econometrica* 47 (1979), 115–128.

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

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

Dickey, D. A., and W. A. Fuller, "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74 (1979), 427–431.

Dufour, Jean-Marie, "Exact Test and Confidence Sets in Linear Regressions with Autocorrelated Errors," *Econometrica* 58:2 (1990), 475–494.

Durlauf, S. N., and P. C. B. Phillips, "Trends versus Random Walks in Time Series Analysis," *Econometrica* 56 (1988), 1333–1354.

Elliott, Graham, "Efficient Tests for a Unit Root when the Initial Observation Is Drawn from Its Unconditional Distribution," Harvard University, Manuscript (1993).

Elliott, Graham, Thomas J. Rothenberg, and James H. Stock, "Efficient Tests of an Autoregressive Unit Root," *Econometrica* 64:4 (1996), 813–836.

Grenander, U., "On the Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance," *Annals of Mathematical Statistics* 25 (1954), 252–272.

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

Lee, C. C., and P. C. B. Phillips, "Efficiency Gains Using GLS over OLS under Nonstationarity," Yale University, Manuscript (1994).

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

Maeshiro, Asatoshi, "Autoregressive Transformation, Trended Independent Variables and Autocorrelated Disturbances," this REVIEW 58 (1976), 497–500.

——— "On the Retention of the First Observations in Serial Correlation Adjustment of Regression Models," *International Economic Review* 20:1 (1979), 259–265.

Magee, L., "A Note on Cochrane–Orcutt Estimation," *Journal of Econometrics* 35 (1987), 211–218.

Nagaraj, N. K., and W. A. Fuller, "Estimation of the Parameters of Linear Time Series Models Subject to Nonlinear Restrictions," *Annals of Statistics* 19:3 (1991), 1143–1154.

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

Phillips, P. C. B., "Toward a Unified Asymptotic Theory for Autoregression," *Biometrika* 74 (1987), 535–547.

Phillips, P. C. B., and C. C. Lee, "Efficiency Gains from Quasi-Differencing under Nonstationarity," in P. M. Robinson and M. Rosenblatt (eds.), *Conference on Applied Probability and Time Series,* vol. II, *Time Series Analysis in Memory of E. J. Hannan* (New York: Springer, 1996).

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

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

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

Rao, P., and Z. Griliches, "Small-Sample Properties of Several Two-Stage Regression Methods in the Context of Auto-Correlated Errors," *Journal of the American Statistical Association* 64 (1969), 253–272.

Sampson, M., "The Effect of Parameter Uncertainty on Forecast Variances and Confidence Intervals for Unit Root and Trend Stationary Time-Series Models," *Journal of Applied Econometrics* 6:1 (1991), 67–76.

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

Schmidt, P., and P. C. B. Phillips, "LM Tests for a Unit Root in the Presence of Deterministic Trends," *Oxford Bulletin of Economics and Statistics* 54 (1992), 257–287.

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

Stock, James H., "Confidence Intervals of the Largest Autoregressive Root in U.S. Macroeconomic Time Series," *Journal of Monetary Economics* 28 (1991), 435–460.

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

Thornton, D. L., "A Note on the Effect of the Cochrane–Orcutt Estimator in the AR(1) Regression Model," *Journal of Econometrics* 36 (1987), 369–376.

APPENDIX

Theorem Proofs

A.1. Preliminaries

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

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

A.2 Proof of Theorem 2

Proof of (a): By direct calculation,

$$T^{1/2}(\hat{\beta}_{\text{OLS}} - \beta) = \frac{T^{-3/2}\sum_{t=1}^{T}(t/T)u_t - \left[T^{-1}\sum_{t=1}^{T}(t/T)\right]\left[T^{-3/2}\sum_{t=1}^{T}u_t\right]}{T^{-1}\sum_{t=1}^{T}(t/T)^2 - \left[T^{-1}\sum_{t=1}^{T}(t/T)\right]^2}.$$

Thus

$$T^{1/2}(\hat{\beta}_{\text{OLS}} - \beta) = 12T^{-3/2}\sum_{t=1}^{T}u_t\left(\frac{t}{T} - \frac{1}{2}\right) + o_p(1)$$
$$\Rightarrow d(1)12\int_0^1\left(s - \frac{1}{2}\right)[W_c(s) + e^{sc}\tilde{W}_c(\kappa)]\,ds \sim N(0, R_1)$$

where $R_1 = A_1 + A_2$, with

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

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

To calculate A_1, note that

$$\int_0^1\left(s - \frac{1}{2}\right)W_c(c)\,ds = \int_0^1\left(s - \frac{1}{2}\right)\int_0^s e^{c(s-\tau)}dW(\tau)\,ds$$
$$= \int_0^1\left[\int_\tau^1\left(s - \frac{1}{2}\right)e^{cs}\,ds\right]e^{-c\tau}dW(\tau)$$
$$= \int_0^1 b(\tau)dW(\tau)$$

with

$$b(\tau) = \left[\int_\tau^1 \left(s - \frac{1}{2}\right) e^{cs}\, ds\right] e^{-c\tau}.$$

Thus

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

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

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

Proof of (b):

$$T^{1/2}(\hat\beta_{FD} - \beta) = T^{-1/2}u_T - T^{-1/2}u_1 = T^{-1/2}\tilde u_T - T^{-1/2}u_1(1 - \rho_T^{T-1})$$
$$\Rightarrow d(1)[W_c(1) - (1 - e^c)\tilde W_c(\kappa)]$$
$$\sim N(0, d(1)^2[S_c(1) + (1 - e^c)^2 S_c(\kappa)]).$$

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 $\delta = (\alpha \quad \beta)'$, $x_1 = (\sigma_{u_1}^{-1} \quad \sigma_{u_1}^{-1})'$, $x_t = [(1 - \rho_T) \quad t - \rho_T(t - 1)]'$ for $t = 2, \ldots, T$. Let $Q = \Sigma x_t x_t'$ and $r = \Sigma x_t e_t$, 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,

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

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

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

$$r_1 = \sigma_{u_1}^{-2}u_1 + (1 - \rho_T) \sum_{t=2}^T v_t$$

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

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

$\kappa = 0$: By direct calculation,

$$T^{-1}(q_{11}q_{22} - q_{12}^2) \to \left(1 - c + \frac{1}{3}c^2\right)$$

$$T^{-1/2}q_{11}r_{12} = T^{-1/2} \sum_{t=2}^T v_t\left(1 - c\frac{t}{T}\right) + o_p(1)$$

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

Thus

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

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

$$\sim N\left(0, d(1)^2\left(1 - c + \frac{1}{3}c^2\right)^{-1}\right).$$

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,

$$(q_{11}q_{22} - q_{12}^2) \to [S_c(\kappa)^{-1} + c^2]\left(1 - c + \frac{1}{3}c^2\right) - \left(\frac{1}{2}c^2 - c\right)^2$$

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

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

Thus

$$T^{1/2}(\hat\beta_{GLS} - \beta) = [(S_c(\kappa)^{-1} + c^2)(1 - c + \tfrac{1}{3}c^2) - (\tfrac{1}{2}c^2 - c)^2]^{-1}$$
$$\times \Bigg\{-(\tfrac{1}{2}c^2 - c)S_c(\kappa)^{-1}T^{-1/2}u_1$$
$$+ T^{-1/2}\sum_{t=2}^T v_t\left[\left(1 - c\frac{t}{T}\right)(S_c(\kappa)^{-1} + c^2) + (\tfrac{1}{2}c^3 - c^2)\right]\Bigg\} + o_p(1)$$
$$\Rightarrow [(c^2 + S_c(\kappa)^{-1})(1 - c + \tfrac{1}{3}c^2) - (\tfrac{1}{2}c^2 - c)^2]^{-1}$$
$$\times \{d(1)[-(\tfrac{1}{2}c^2 - c)\tilde W_c(\kappa)S_c(\kappa)^{-1}$$
$$+ \int_0^1 [((S_c(\kappa)^{-1} + c^2](1 - cs) + (\tfrac{1}{2}c^2 - c)c]dW(s)]\}$$
$$\sim N(0, R_3)$$

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 of Theorem 2 (c), 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 $Q = \Sigma x_t x_t'$ and $r = \Sigma x_t e_t$, 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 for the proof we evaluate these expressions for each estimator.

Proof of (a): When $c = 0$, $T^{1/2}(\hat\beta_{CO} - \beta) = T^{1/2}\Sigma_{t=2}^T v_t$, and the result follows directly. For $c \neq 0$,

$$q_{11} = (T - 1)(1 - \rho_T)^2$$

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

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

$$r_1 = (1 - \rho_T) \sum_{t=2}^T v_t$$

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

Thus

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

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

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

so that

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

$$= \left(\frac{12}{c}\right)T^{-1/2}\sum v_t\left(\frac{1}{2} - \frac{t}{T}\right) + o_p(1) \qquad \text{(A.1)}$$

$$\Rightarrow \left(\frac{12}{c}\right)d(1)\int_0^1\left(\frac{1}{2} - s\right)dW(s). \qquad \text{(A.2)}$$

The results follows by noting that

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

where

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

Proof of (b): For $\hat\beta_{CC}$,

$$q_{11} = 1 + (T-1)(1-\rho_T)^2$$

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

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

$$r_1 = u_1 + (1-\rho_T)\sum_{t=2}^T v_t$$

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

Thus

$$T^{-1}(q_{11}q_{22} - q_{12}^2) \to \left(1 - c + \frac{1}{3}c^2\right)$$

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

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

so that

$$T^{1/2}(\hat\beta_{CC} - \beta) = \frac{c(1-\frac{1}{2}c)T^{-1/2}u_1 - T^{-1/2}\sum v_t\left(c\frac{t}{T} - 1\right)}{1 - c + \frac{1}{3}c^2} + o_p(1) \qquad \text{(A.3)}$$

$$\Rightarrow \frac{d(1)}{1 - c + \frac{1}{3}c^2} \times \left[c\left(1 - \frac{1}{2}c\right)\tilde W_c(\kappa) - \int_0^1 (cs-1)dW(s)\right]. \qquad \text{(A.4)}$$

The result follows by noting that

$$d(1)\left(1 - c + \frac{1}{3}c^2\right)^{-1}\left[c\left(1 - \frac{1}{2}c\right)\tilde W_c(\kappa) - \int_0^1 (cs-1)\,\mathrm{dW}(s)\right] \sim N(0, G_2)$$

where

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

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

Proof of (c): For $\hat\beta_{PW}$,

$$q_{11} = (1-\rho_T^2) + (T-1)(1-\rho_T)^2$$

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

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

$$r_1 = (1-\rho_T^2)u_1 + (1-\rho_T)\sum_{t=2}^T v_t$$

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

Thus

$$q_{11}q_{22} - q_{12}^2 \to (c^2 - 2c)\left(1 - c + \frac{1}{3}c^2\right) - \left(\frac{1}{2}c^2 - c\right)^2$$

$$= (c^2 - 2c)\left(1 - \frac{1}{2}c + \frac{1}{12}c^2\right)$$

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

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

so that

$$T^{1/2}(\hat{\beta}_{PW} - \beta) = \frac{cT^{-1/2}u_1 - T^{-1/2}\sum_{t=2}^{T} v_t\left(c\frac{t}{T} - \frac{1}{2}c - 1\right)}{1 - \frac{1}{2}c + \frac{1}{12}c^2} + o_p(1) \tag{A.5}$$

$$\Rightarrow \frac{d(1)}{1 - \frac{1}{2}c + \frac{1}{12}c^2} \cdot \left[c\tilde{W}_c(\kappa) - \int_0^1 \left(cs - \frac{1}{2}c - 1\right) dW(s)\right] \tag{A.6}$$

The result follows by noting that

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

where

$$G_3 = \frac{d(1)^2}{(1 - \frac{1}{2}c + \frac{1}{12}c^2)^2}\left[c^2 S_c(\kappa) + \int_0^1 \left(cs - \frac{1}{2}c - 1\right)^2 ds\right]$$

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

A.4. *Proof of Theorem 5*

It is straightforward to verify that the analogues of equations (A.1), (A.3), and (A.5) 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 equations (A.2), (A.4), and (A.6) using $T^{-1/2}\sum_{t=1}^{[sT]} \hat{v}_t \Rightarrow \hat{W}(c)$. To see this, and to derive an expression for $\hat{W}(c)$, write

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

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

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

$$T^{-1/2}\sum_{t=1}^{[sT]} \hat{v}_t = T^{-1/2}\sum_{t=1}^{[sT]} v_t - (\hat{c}_T - c)T^{-1}\sum_{t=1}^{[sT]} T^{-1/2} \times \left(\sum_{j=0}^{t-2} \rho_T^j v_{t-1-j}\right) - (\hat{c}_T - c)(T^{-1/2}u_1)T^{-1}\sum_{t=1}^{[sT]} \rho_T^{t-1}$$

$$\Rightarrow d(1)\hat{W}(s)$$

where

$$\hat{W}(s) = W(s) - (\hat{c} - c)\int_0^s W_c(\tau)d\tau - (\hat{c} - c)\tilde{W}_c(\kappa)\frac{1 - e^{sc}}{-c}$$

and the last line follows from $\hat{c}_T \Rightarrow \hat{c}$,

$$T^{-1/2}\sum_{j=0}^{[\tau T]} \rho_T^j v_{[\tau T]-j} \Rightarrow W_c(\tau) \quad \text{and} \quad T^{-1}\sum_{t=1}^{[sT]} \rho_T^{t-1} \rightarrow \frac{1 - e^{sc}}{-c}.$$