Cowles Foundation Paper 936
Reprinted from P.M. Robinson and M. Rosenblatt (eds.), Athens Conference on Applied Probability and Time Series, Vol. II, 1996

EFFICIENCY GAINS FROM QUASI-DIFFERENCING UNDER NONSTATIONARITY*

Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

and

Chin Chin Lee
Department of Economics
London School of Economics and Political Science

16 March 1996

Abstract

A famous theorem on trend removal by OLS regression (usually attributed to Grenander and Rosenblatt (1957)) gave conditions for the asymptotic equivalence of GLS and OLS in deterministic trend extraction. When a time series has trend components that are stochastically nonstationary, this asymptotic equivalence no longer holds. We consider models with integrated and near-integrated error processes where this asymptotic equivalence breaks down. In such models, the advantages of GLS can be achieved through quasi-differencing and we give an asymptotic theory of the relative gains that occur in deterministic trend extraction in such cases. Some differences between models with and without intercepts are explored.

1. Introduction

Grenander and Rosenblatt (1957) analysed asymptotic efficiency conditions in time series regressions with stationary errors. They considered

*A first draft containing some of the results reported here was written in 1993. The present paper is an abridged version of Phillips and Lee (1996). Our thanks go to the NSF for research support under Grant Numbers SES 9122142 and SBR 94-22922.

univariate regression models with trends such as $y_t = \beta' z_t + u_t$, where $z_t = (1, t, ..., t^k)$ and u_t is stationary with spectral density $f_u(\lambda) > 0$, and demonstrated the asymptotic equivalence of GLS and OLS trend extraction techniques. Hannan (1970, chapter VII) extended the Grenander-Rosenblatt theory to the case of multivariate time series regressions and provided a general treatment of the subject. The Grenander-Rosenblatt result relies on the continuity of the spectrum of u_t at the origin (where the spectral mass of z_t in the above model is concentrated) and it is satisfied in most models that involve stationary time series. But, the condition is violated when there is a unit root in the data generating process of u_t. In fact, the condition fails whenever u_t is strongly dependent or integrated of order d with $d > 0$ (denoted as I(d)). For in that case, the spectral density of u_t behaves like a multiple of λ^{-2d} as $\lambda \to 0$ and is unbounded at $\lambda = 0$. In such cases as these, the asymptotic equivalence of GLS and OLS breaks down and we can achieve efficiency gains in estimating the trend coefficients β by using GLS methods. When u_t is near-integrated in the sense that it has an autoregressive root that is local to unity, there is again a peak at the origin in its spectrum and we can still expect gains to accrue from the use of GLS estimation.

The present contribution calculates the efficiency gains in GLS trend extraction when u_t is an integrated or near-integrated process. These cases are the most commonly studied in the econometrics literature, they have bearing on the issue of unit root testing, and they lend themselves to simple quasi-differencing formulations that are convenient in practical work. In contrast to the integrated and near-integrated cases, the effects of strong dependence on the efficiency of OLS have received attention in the literature. In particular, Yajima (1988), Beran (1994, ch. 9) and Samarov and Taqqu (1988) study GLS efficiency gains in models with stationary long-memory errors where $0 \leq d < 1/2$. The case where there are nonstationary strongly dependent errors with $1/2 < d < 1$ was analysed in Lee and Phillips (1994).

2. Efficiency Gains in Models with Near Integrated Errors

Suppose a time series y_t is generated by

$$y_t = \beta_k' z_{kt} + u_t \ , \quad t = 1, ..., T \ , \qquad (1)$$
$$u_t = \alpha u_{t-1} + \varepsilon_t \ , \quad \alpha = 1 + c/T \ ,$$

where $z_{kt} = (t, ..., t^k)'$, and c is a constant that represents local departures from unity. The parameter setting $\alpha = 1 + c/T$ facilitates efficiency calculations using local-to-unity asymptotics (see Phillips, 1987a, and Chan and

Wei, 1987). The Grenander-Rosenblatt theory applies when $|\alpha| < 1$, and our interest is in the unit root and intermediate cases. Hence, attention here focuses on the domain $c \in (-\infty, 0]$.

Initial conditions for u_t are set at $t = 0$ and u_0 may be any random variable with finite variance σ_0^2. Cases where $\sigma_0^2 \to \infty$ are sometimes of interest and these can correspond to situations where the initial conditions are in the increasingly distant past, although observations on the process y_t are available only from $t = 1$. The effect of such alternative initializations on our results are considered later.

The primary requirement on the shocks ε_t is that normalized partial sums $S_t = \sum_{s=1}^{t} \varepsilon_s$ of ε_t satisfy an invariance principle and this will be so under a wide variety of differing conditions on ε_t. The following conditions on ε_t are sufficient for the limit theory here.

2.1 Assumption EC *(Error Conditions)*

(i) $E\varepsilon_t = 0 \ \forall t$; (ii) $\sup_t E|\varepsilon_t|^{b+\delta} < \infty$ *for some* $b > 2$ *and* $\delta > 0$;

(iii) $\sigma^2 = \lim E(S_T^2/T)$ *exists, and* $\sigma^2 > 0$; (iv) ε_t *is strong mixing with coefficients* α_m *that satisfy* $\sum_{m=1}^{\infty} \alpha_m^{1-2/b} < \infty$.

In the following, we use $W(r)$ to denote standard Brownian motion and $J_c(r) = \int_0^r e^{(r-s)c} dW(s)$ to denote a linear diffusion process. Note that $J_c(r)$ satisfies the linear stochastic differential equation $dJ_c(r) = cJ_c(r)dr + dW(r)$. Under Assumption EC we have:

2.2 Lemma

(i) $T^{-1/2} S_{[Tr]} \Rightarrow \sigma W(r)$; (ii) $D_{kT}^{-1/2} \sum_{t=2}^{T} z_{kt}\varepsilon_t \Rightarrow \sigma \int_0^1 g_k(r) dW(r)$;

(iii) $T^{-1/2} u_{[Tr]} \Rightarrow \sigma J_c(r)$; (iv) $T^{-1} D_{kT}^{-1/2} \sum_{t=1}^{T} z_{kt} u_t \Rightarrow \sigma \int_0^1 g_k(r) J_c(r) dr$;

where $D_{kT} = \mathrm{diag}(T^3, T^5, ..., T^{2k+1})$, $g_k(r)' = (r, ..., r^k)$ *and* $\Rightarrow$ *signifies weak convergence.*

Simple least squares regression on (1) leads to the trend coefficient estimator $\widehat{\beta}_{kc} = \left(\Sigma_1^T z_{kt} z_{kt}'\right)^{-1} (\Sigma_1^T z_{kt} y_t)$. GLS regression requires use of the full covariance structure of the error process u_t. The Grenander-Rosenblatt theory can be expected to cover contributions to the covariance structure that come from the stationary or weakly dependent components ε_t, but not those that come from the autoregressive root $\alpha = 1 + c/T$ since it is the latter that produces a peak in the spectrum of u_t. Hence, as an alternative to OLS, we consider a partial GLS detrending procedure that is based on the quasi-differenced data $\widetilde{z}_{kt} = z_{kt} - \alpha z_{kt-1}$ and $\widetilde{y}_t = y_t - \alpha y_{t-1}$ for $t = 2, ..., T$, combined with the initial observations $\widetilde{z}_{k1} = z_{k1}$, $\widetilde{y}_1 = y_1$ for $t = 1$. This leads to the estimator $\widetilde{\beta}_{kc} = \left(\Sigma_1^T \widetilde{z}_{kt}\widetilde{z}_{kt}'\right)^{-1} \Sigma_1^T \widetilde{z}_{kt}\widetilde{y}_t$.

We show that the partial GLS estimator $\widetilde{\beta}_{kc}$ of β_k in (1) is asymptotically more efficient than $\widehat{\beta}_{kc}$ under both a unit root ($c = 0$) and a near unit root ($c < 0$). The following results give the limit distributions of these estimators.

2.3 Theorem

$$F_{kT}^{1/2}(\widehat{\beta}_{kc} - \beta_k) \Rightarrow \sigma Q_k^{-1} \int_0^1 g_k(r) J_c(r) dr \equiv N(0, V_{kc}^{ols})$$

where $F_{kT}^{1/2} = T^{-1} D_{kT}^{1/2} = \text{diag}(T^{1/2}, T^{3/2}, ..., T^{k-1/2})$, $Q_k = \int_0^1 g_k(r) g_k(r)' dr$ *is a* $k \times k$ *matrix with elements* $q_{ij} = 1/(i+j+1)$ *and*

$$V_{kc}^{ols} = \sigma^2 Q_k^{-1} \int_0^1 \int_0^1 g_k(r) e^{(r+s)c}(1/2c)(1 - e^{-2c(r \wedge s)}) g_k(s)' dr ds Q_k^{-1\prime}.$$

2.4 Theorem

$$F_{kT}^{1/2}(\widetilde{\beta}_{kc} - \beta_k) \Rightarrow \sigma \left[\int_0^1 f_{ck}(r) f_{ck}(r)' dr\right]^{-1} \int_0^1 f_{ck}(r) dW(r) \equiv N(0, V_{kc}^{gls}),$$

with

$$V_{kc}^{gls} = \sigma^2 \left[\int_0^1 f_{ck}(r) f_{ck}(r)' dr\right]^{-1} = \sigma^2 \left[\overline{Q}_k + c^2 Q_k - c(\widetilde{Q}_k + \widetilde{Q}_k')\right]^{-1}.$$

Here, $f_{ck}(r) = g_k^{(1)}(r) - c g_k(r)$, $g_k^{(1)}(r) = (1, 2r, ..., kr^{k-1})'$, *and* $\overline{Q}_k$ *and* $\widetilde{Q}_k$ *are* $k \times k$ *matrices with elements* $\overline{q}_{ij} = ij/(i+j-1)$ *and* $\widetilde{q}_{ij} = i/(i+j)$, *respectively.*

Define the relative efficiency of $\widehat{\beta}_{kc}$ to $\widetilde{\beta}_{kc}$ by $R_{kc} \equiv \det(V_{kc}^{ols}) / \det(V_{kc}^{gls})$. To provide some illustrative comparisons, take the case of the linear trend model where $k = 1$ in (1). Then, when $c = 0$, $T^{1/2}(\widehat{\beta}_{10} - \beta_1) \Rightarrow N(0, 6\sigma^2/5)$ — a result obtained earlier in Durlauf and Phillips (1988). On the other hand, $T^{1/2}(\widetilde{\beta}_{10} - \beta_1) \Rightarrow \sigma W(1) = N(0, \sigma^2)$. Hence, for linear trend extraction there is an asymptotic efficiency gain of 20% from the use of the partial GLS estimator $\widetilde{\beta}_{10}$ when u_t is integrated of order 1. When $c \neq 0$, the variances of the limit variates are

$$V_{1c}^{ols} = 9\sigma^2[3e^{2c}(c-1)^2 + 2c^3 + 3c^2 - 3]/6c^5, \quad \text{and } V_{1c}^{gls} = 3\sigma^2/(3 - 3c + c^2).$$

In this case, the relative efficiency $R_{1c} = V_{1c}^{ols}/V_{1c}^{gls}$, is graphed against negative values of c in Figure 1. As $c \to -\infty$, $R_{1c} \to 1$, so there are no gains from the use of GLS–detrending in the limiting case. This is to be expected since $c \to -\infty$ is the limit of the domain of definition of c that corresponds to the stationary case, for which the Grenander–Rosenblatt asymptotic equivalence result holds. Figure 1 also shows that the maximum gains in efficiency from GLS occur for finite $c < 0$, rather than at zero.

In the general case, write the limit variates from theorems 2.3 and 2.4 as $\widehat{Z}_c = \sigma Q_k^{-1} \int_0^1 g_k(r) J_c(r) dr$, and $\widetilde{Z}_c = \sigma \left[\int_0^1 f_{ck}(r) f_{ck}(r)' dr\right]^{-1} \int_0^1 f_{ck}(r) dW(r)$. Then, as $c \to -\infty$ the asymptotic equivalence of $\widehat{Z}_c$ and $\widetilde{Z}_c$ is given by

2.5 Theorem $\sqrt{-c}\left(\widehat{Z}_c - \widetilde{Z}_c\right) \xrightarrow{p} 0$ and $R_{kc} \to 1$ as $c \to -\infty$.

3. The Effects of a Fitted Intercept

A constant term is not included in (1) because the intercept is not consistently estimable. Nevertheless, it is usual in empirical work for regression detrending procedures to involve fitted intercepts. So it is of some interest to consider the asymptotic behaviour of the estimators $\widetilde{\beta}_{kc}$ and $\widehat{\beta}_{kc}$ in this case. In related work, Canjels and Watson (1995) studied the case of linear trend extraction with a fitted intercept and near integrated errors. The treatment that follows considers the case of general polynomial trends with fitted intercepts and near integrated errors, and indicates some subtleties in the limit theory that arise as $c \to -\infty$ due to the doubly-infinite triangular array structure of y_t.

Consider the following model in place of (1)

$$y_t = \beta_0 + \beta_k' z_{kt} + u_t = \beta' z_t + u_t \ , \quad t = 1, ..., T \ . \tag{2}$$

It turns out that when the localizing parameter c is fixed, the presence of the constant term β_0 in the regression (2) does not influence the asymptotic distribution of the partial GLS estimator $\widetilde{\beta}_{kc}$. To see this, note that

$$\Sigma_1^T \widetilde{z}_t \widetilde{z}_t' = \begin{bmatrix} 1 + \frac{c^2(T-1)}{T^2} & z_{k1}' - \frac{c}{T}\Sigma_2^T \widetilde{z}_{kt}' \\ z_{k1} - \frac{c}{T}\Sigma_2^T \widetilde{z}_{kt} & \Sigma_1^T \widetilde{z}_{kt} \widetilde{z}_{kt}' \end{bmatrix} \tag{3}$$

where $\tilde{z}_t = z_t - \alpha z_{t-1} = [-c/T, \tilde{z}_{kt}]' = [-c/T, \Delta z'_{kt} - (c/T)z'_{kt-1}]'$ for $t =$ 2, ..., T,and $\tilde{z}_1 = z_1 = [1, \tilde{z}_{k1}]' = [1, z'_{k1}]'$. Setting $D_T = diag(1, F_{kT})$, we have

$$D_T^{-1/2}\left(\Sigma_1^T \tilde{z}_t \tilde{z}_t'\right) D_T^{-1/2} \to \begin{bmatrix} 1 & 0 \\ 0 & \int_0^1 f_{ck}(r) f_{ck}(r)' dr \end{bmatrix}. \qquad (4)$$

Since this matrix is block diagonal, it follows that the inclusion of a fitted intercept in a partial GLS regression on (2) does not alter the asymptotic distribution of the estimates of the trend coefficient vector β_k that is given in theorem 2.4. Thus, the GLS estimates of β_k have the same limit distribution whether or not an intercept is included in the regression. This result depends critically on the assumption that the localizing parameter is fixed.

Unlike $\tilde{\beta}_{kc}$, the limit distribution of the OLS trend estimator $\hat{\beta}_{kc}$ is affected by a fitted intercept. In this case, the limit distribution is found to be

$$\begin{aligned} F_{kT}^{1/2}(\hat{\beta}_{kc} - \beta_k) &\Rightarrow \sigma H_k^{-1} \int_0^1 (g_k(r) - h_k) J_c(r) dr = \hat{Z}_{mc}, \text{ say}, \qquad (5) \\ &\equiv N(0, V_{kc}^{fmols}), \end{aligned}$$

where H_k is $k \times k$ with elements $h_{ij} = 1/(i+j+1) - 1/(i+1)(j+1)$ and h_k is k-vector column with i'th element $1/(i+1)$.

In the linear trend case ($k = 1$), the asymptotic relative efficiency of $\hat{\beta}_{1c}$ to $\tilde{\beta}_{1c}$ is

$$R_{1c}^g = \frac{6c^{-5}(3\exp(2c)(c^2 - 4c + 4) + 12c\exp(c)(c-2) + 2c^3 + 9c^2 + 12c - 12)}{3/(3 - 3c + c^2)}, \qquad (6)$$

which is plotted in Figure 2. Note that the limit of R_{1c}^g as $c \to -\infty$, does not appear from this figure to be unity, as would be expected from the Grenander-Rosenblatt theory, a point that is unnoticed in the work of Canjels and Watson (1995). In fact, a simple calculation shows that $R_{1c}^g \to 4$ as $c \to -\infty$.

We now show why GLS appears to be more efficient than OLS at the boundary of the domain of definition of c. First, observe that the asymptotic theory for $\widehat{\beta}_{kc}$ and $\widetilde{\beta}_{kc}$ essentially involves a triangular array limit theory when both $|c|$ and T are large. Limits taken along different diagonals of this array are not necessarily equivalent. To fix ideas, the partial GLS estimator of β_k in (2) has the explicit form

$$\widetilde{\beta}_{kc} = \left[\Sigma_1^T \widetilde{z}_{kt}\widetilde{z}_{kt}' - \left(z_{k1} - \frac{c}{T}\Sigma_2^T \widetilde{z}_{kt}\right)\left(1+\frac{c^2}{T}\right)^{-1}\left(z_{k1} - \frac{c}{T}\Sigma_2^T \widetilde{z}_{kt}\right)'\right]^{-1} \times \left[\Sigma_1^T \widetilde{z}_{kt}\widetilde{y}_t - \left(z_{k1} - \frac{c}{T}\Sigma_2^T \widetilde{z}_{kt}\right)\left(1+\frac{c^2}{T}\right)^{-1}\left(\widetilde{y}_1 - \frac{c}{T}\Sigma_2^T \widetilde{y}_t\right)\right], \quad (7)$$

and for c^2/T large this is approximately

$$\begin{aligned} \breve{\beta}_k &= \left[\Sigma_1^T \widetilde{z}_{kt}\widetilde{z}_{kt}' - T^{-1}\left(\Sigma_1^T \widetilde{z}_{kt}\right)\left(\Sigma_1^T \widetilde{z}_{kt}\right)'\right]^{-1}\left[\Sigma_1^T \widetilde{z}_{kt}\widetilde{y}_t - T^{-1}\left(\Sigma_1^T \widetilde{z}_{kt}\right)\left(\Sigma_1^T \widetilde{y}_t\right)\right] \\ &= \left[\Sigma_1^T\left(\widetilde{z}_{kt} - \overline{\widetilde{z}}_k\right)\left(\widetilde{z}_{kt} - \overline{\widetilde{z}}_k\right)'\right]^{-1}\left[\Sigma_1^T\left(\widetilde{z}_{kt} - \overline{\widetilde{z}}_k\right)\widetilde{y}_t\right], \end{aligned} \quad (8)$$

which is the OLS regression coefficient of $\widetilde{z}_{kt}$ in a regression of $\widetilde{y}_t$ on $\widetilde{z}_{kt}$ with a fitted intercept. Call this estimator the *fitted mean* GLS estimator of β_k in (2). Now, in place of theorem 2.4, we get the following asymptotic distribution theory.

3.1 Theorem *The limit distribution of the fitted mean GLS estimator* $\breve{\beta}_k$ *in* (2) *is given by*

$$F_{kT}^{1/2}(\breve{\beta}_k - \beta_k) \Rightarrow \sigma\left[\int_0^1 \overline{f}_{ck}(r)\overline{f}_{ck}(r)'dr\right]^{-1}\left[\int_0^1 \overline{f}_{ck}(r)dW(r)\right] \equiv N(0, V_{kc}^{fmgls}),$$

where $\overline{f}_{ck}(r) = f_{ck}(r) - \int_0^1 f_{ck}(r)dr$, *and* $V_{kc}^{fmgls} = \sigma^2\left[\int_0^1 \overline{f}_{ck}(r)\overline{f}_{ck}(r)'dr\right]^{-1}$.

Let $\breve{Z}_{mc} = \sigma\left[\int_0^1 \overline{f}_{ck}(r)\overline{f}_{ck}(r)'dr\right]^{-1}\left[\int_0^1 \overline{f}_{ck}(r)dW(r)\right]$ be the variate representing the limit distribution of the fitted mean GLS estimator given in theorem 3.1. Define the efficiency ratio of this estimator against OLS as $R_{kc}^m \equiv \det(V_{kc}^{fmols})/\det(V_{kc}^{fmgls})$. Then, we have accordance with the Grenander-Rosenblatt theory at the limit of the domain of definition of c as follows.

3.2 Theorem: $\sqrt{-c}\left(\widehat{Z}_{mc} - \breve{Z}_{mc}\right) \xrightarrow{p} 0$ *and* $R_{kc}^m \to 1$ *as* $c \to -\infty$.

An intercept in the regression also affects the asymptotics of the partial GLS estimator when $c^2/T \to c_1$, or equivalently, $c/\sqrt{T} \to c_0$, where c_0 is some finite negative constant and $c_1 = c_0^2$. In this case when $c/\sqrt{T} \sim c_0$, the partial GLS estimator given in (7) is approximately

$$\begin{aligned} \widehat{\beta}_k &= \beta_k + \left[\Sigma_1^T \widetilde{z}_{kt}\widetilde{z}_{kt}' - \frac{c_1}{1+c_1}\frac{1}{T}\left(\Sigma_2^T \widetilde{z}_{kt}\right)\left(\Sigma_2^T \widetilde{z}_{kt}\right)'\right]^{-1} \\ &\times \left[\Sigma_1^T \widetilde{z}_{kt}\widetilde{u}_t + \frac{c_0}{1+c_1}\frac{1}{\sqrt{T}}\left(\Sigma_2^T \widetilde{z}_{kt}\right)\left(u_1 - c_0\frac{1}{\sqrt{T}}\Sigma_2^T \widetilde{u}_t\right)\right]. \end{aligned}$$

Then, in place of theorem 3.1, we get the following limit theory as $T \to \infty$ for $\widehat{\beta}_k$

$$F_{kT}^{1/2}(\widehat{\beta}_k - \beta_k) \Rightarrow V^{-1}\left[\sigma \int_0^1 f_{ck}^{c_1}(r)dW(r) + \frac{c_0}{1+c_1}\left(\int_0^1 f_{ck}(r)dr\right)u_1\right],$$

where $V = \int_0^1 f_{ck}(r)f_{ck}(r)'dr - (c_1/(1+c_1))\left(\int_0^1 f_{ck}(r)dr\right)\left(\int_0^1 f_{ck}(r)dr\right)'$, and $f_{ck}^{c_1}(r) = f_{ck}(r) - (c_1/(1+c_1))\int_0^1 f_{ck}(s)ds$. This limit distribution is somewhat unusual because the first period error term u_1 plays a role in the asymptotics. This is explained by the fact that the normalized second moment matrix (3) is not block diagonal in the limit as $T \to \infty$ when $c/\sqrt{T} \sim c_0$, the intercept in (2) is not consistently estimated, and consequently u_1 has an effect on the limit distribution of $\widehat{\beta}_k$.

Setting $\sigma_1^2 = var(u_1)$ and $h_{ck} = \int_0^1 f_{ck}(r)dr$, the variance of the limiting distribution of $\widehat{\beta}_k$ is

$$V_{kc}^g = V^{-1}\left\{\sigma^2 \int_0^1 f_{ck}^{c_1}(r)f_{ck}^{c_1}(r)'dr + \sigma_1^2 \frac{c_1}{(1+c_1)^2}h_{ck}h_{ck}'\right\}V^{-1}.$$

To illustrate, take the case $k = 1$ and suppose $\sigma_1^2 = 0$. The relative asymptotic efficiency $R_{1c}^0 \equiv V_{1c}^{fmols}/V_{1c}^g$ of the partial GLS estimator $\widehat{\beta}_1$ and the OLS estimator is plotted in Figure 3 against c_0. As is apparent from the figure, the efficiency curve tends to 4 as $c_0 \to 0$ and tends to 1 as $c_0 \to -\infty$.

Finally, we go back to the direct comparison of OLS and GLS in the model (2). In the general case, define the efficiency ratio of the fitted mean OLS estimator $\widehat{\beta}_{kc}$ and the GLS trend coefficient estimator $\widetilde{\beta}_{kc}$ by $R_{kc}^g \equiv$

$\det(V_{kc}^{fmols})/\det(V_{kc}^{gls})$. The limiting behaviour of this ratio as $c \to -\infty$ is given in the next result.

3.3 Theorem $R_{kc}^{g} \to (k+1)^2$ *as* $c \to -\infty$.

For $k = 1$ this reduces to the earlier result discussed above, where $R_{1c}^{g} \to 4$. The factor $(k+1)^2$ measures the additional variance in the limit of the OLS procedure that is due to estimating an intercept in the regression.

4. Alternative Initializations

The results given above rely on the initial observation u_0 having constant variance σ_0^2, so that $u_0 = O_{a.s.}(1)$ as $T \to \infty$. There is some merit to making assumptions about u_0 which give it properties that are analogous to those of u_t itself. This can be done by putting the initial conditions that determine u_0 into the increasingly distant past as $T \to \infty$. One way of doing this (*e.g.* Uhlig, 1995, or Canjels and Watson, 1995) is to define $u_0 = u + \sum_{j=0}^{[T\tau]} \alpha^j \varepsilon_{-j}$, for some $\tau \geq 0$, and $u = O_{a.s.}(1)$ and with ε_{-j} satisfying assumption 2.1. Then $T^{-1/2}u_0 \Rightarrow \sigma J_{c,0}(\tau)$, where $J_{c,0}$ is a diffusion process generated by $dJ_{c,0}(r) = cJ_{c,0}(r)dr + dW_0(r)$, in which W_0 is a standard Brownian motion independent of W. All of the above theory can be developed for this initialization of u_0, with no changes of substance in the limit theory. For example, in place of lemma 2.2 (iii) and (iv) we have:

(iii′) $T^{-1/2}u_{[T\tau]} \Rightarrow \sigma J_{c\tau}(r)$; (iv′) $T^{-1}D_{kT}^{-1/2}\sum_{t=1}^{T} z_{kt}u_t \Rightarrow \sigma \int_0^1 g_k(r)J_{c\tau}(r)dr$; where $J_{c\tau}(r) = J_c(r) + e^{cr}J_{c,0}(\tau)$. For fixed τ, $J_{c,0}(\tau) \equiv N(0, S_c(\tau))$, where $S_c(\tau) = (e^{2\tau c} - 1)/(2c)$ - *e.g.* Phillips (1987a)- and $J_{c,0}(\tau)$ is independent of $J_c(r)$. Then, the limit distribution of the OLS estimator $\widehat{\beta}_{kc}$ is found to be

$$F_{kT}^{1/2}(\widehat{\beta}_{kc} - \beta_k) \Rightarrow \sigma Q_k^{-1}\int_0^1 g_k(r)J_{c\tau}(r)dr$$

$$= \sigma Q_k^{-1}\left\{\int_0^1 g_k(r)J_c(r)dr + \int_0^1 g_k(r)e^{cr}dr J_{c,0}(\tau)\right\} \equiv N\left(0, V_{\tau kc}^{ols}\right),$$

with

$$V_{\tau kc}^{ols} = V_{kc}^{ols} + \sigma^2 S_c(\tau)Q_k^{-1}\left(\int_0^1 g_k(r)e^{cr}dr\right)\left(\int_0^1 g_k(r)e^{cr}dr\right)' Q_k^{-1},$$

and $Q_k = \int_0^1 g_k(r)g_k(r)'dr$, as before. Similarly, for the partial GLS estimator we get

$$\begin{aligned} F_{kT}^{1/2}(\widetilde{\beta}_{kc} - \beta_k) &\Rightarrow \sigma\left[\int_0^1 f_{ck}(r)f_{ck}(r)'dr\right]^{-1}\left\{\left[\int_0^1 f_{ck}(r)dW(r)\right] + J_{c,0}(\tau)e\right\} \\ &\equiv N(0, V_{\tau kc}^{gls}), \end{aligned}$$

with $V^{gls}_{\tau kc} = V^{gls}_{kc} + \sigma^2 S_c(\tau)\left[\int_0^1 f_{ck}(r)f_{ck}(r)'dr\right]^{-1} ee' \left[\int_0^1 f_{ck}(r)f_{ck}(r)'dr\right]^{-1}$ and $e' = (1,0,...,0)$.

The asymptotic relative efficiency of $\widehat{\beta}_{kc}$ to $\widetilde{\beta}_{kc}$ is now given by the ratio $R_{\tau kc} \equiv \det(V^{ols}_{\tau kc})/\det(V^{gls}_{\tau kc})$, and this depends on the initialization parameter τ. Figure 4 plots the efficiency curves against negative values of c for various τ. Apparently, τ has little effect on the relative efficiency of $\widehat{\beta}_{kc}$ to $\widetilde{\beta}_{kc}$ for values of $c \leq -4$. However, when $c \in (-4,0]$, the effect of more distant initial conditions is seen to be substantial. A simple calculation shows that, as $\tau \to \infty$, $S_c(\tau) \to 1/(-2c)$ and $J_{c,0}(\tau) \Rightarrow N(0,1/(-2c))$. Then, the initial conditions dominate the limit theory of the estimators for $c \sim 0$. In fact, for large τ and as $c \to 0$ we find that

$$R_{\tau kc} \sim \frac{\det\left[\frac{1}{-2c}Q_k^{-1}\left(\int_0^1 g_k(r)e^{cr}dr\right)\left(\int_0^1 g_k(r)e^{cr}dr\right)' Q_k^{-1}\right]}{\det\left[\frac{1}{-2c}\left[\int_0^1 f_{ck}(r)f_{ck}(r)'dr\right]^{-1} ee' \left[\int_0^1 f_{ck}(r)f_{ck}(r)'dr\right]^{-1}\right]}$$
$$\to \begin{cases} 9/4 & \text{for } k=1 \\ \infty & \text{for } k>1 \end{cases}$$

On the other hand, as $c \to -\infty$, $\sqrt{-c}\,[J_{c,\infty}(r) - J_c(r)] \xrightarrow{p} 0$, $S_c(\infty) \to 0$ and $J_{c,0}(\infty) \xrightarrow{p} 0$. Hence, the limit theory for $c \to -\infty$ that is given in theorems 2.5 and 3.2 continues to apply even with the new initialization.

5. Conclusion

This paper shows that GLS methods are asymptotically more efficient than OLS in estimating deterministic trend coefficients when the error process is integrated or near-integrated. Maximal gains tend to occur when the localizing parameter c is less than zero in the near integrated case, unless the initial conditions are in the very distant past. If the trend extraction procedures involve a fitted intercept, some interesting subtleties in the limit theory arise as $c \to -\infty$, the lower limit of its domain of definition that corresponds to the case of stationary errors. In this case, y_t is generated by a doubly infinite triangular array, and the limit distribution of the GLS estimator depends on the relative approach to infinity of the two parameters $-c$ and T.

The gains in efficiency that accrue from the GLS trend extraction procedures studied here suggest that there are likely to be similar advantages in other models that involve nonstationary processes, such as multiple equation systems with stochastic and deterministic cointegration.

6. Proofs

We outline the proofs of the results given in the text. Further details are given in Phillips and Lee (1996).

6.1 Proof of Lemma 2.2 Parts (i) and (iii) are proved in Phillips (1987a, b). Parts (ii) and (iv) are proved in Phillips and Perron (1988) for $k = 1$. The extension to $k > 1$ is straightforward.

6.2 Proofs of Theorems 2.3 and 2.4 These follow in a simple way from the form of $F_{kT}^{1/2}(\widehat{\beta}_{kc}-\beta_k)$and $F_{kT}^{1/2}(\widetilde{\beta}_{kc}-\beta_k)$ and the results in lemma 2.2.

6.3 Proof of Theorem 2.5 Since $J_c(r)$ satisfies the differential equation $dJ_c(r) = cJ_c(r)dr + dW(r)$, we have

$$Q_k^{-1}\int_0^1 g_k(r)J_c(r)dr = \frac{1}{-c}Q_k^{-1}\int_0^1 g_k(r)dW(r) + \frac{1}{c}Q_k^{-1}\int_0^1 g_k(r)dJ_c(r) \quad (9)$$

Note that

$$\begin{aligned}\int_0^1 g_k(r)dJ_c(r) &= [g_k(r)J_c(r)]_0^1 - \int_0^1 g_k^{(1)}(r)J_c(r)dr \\ &= g_k(1)J_c(1) - \left\{\frac{1}{-c}\int_0^1 g_k^{(1)}(r)dW(r) + \frac{1}{c}\int_0^1 g_k^{(1)}(r)dJ_c(r)\right\} \\ &= g_k(1)J_c(1) - \frac{1}{c}\left\{\left[g_k^{(1)}(r)J_c(r)\right]_0^1 - \int_0^1 g_k^{(2)}(r)J_c(r)dr\right\} + O_p\left(\frac{1}{|c|}\right) \\ &= \frac{1}{c}\int_0^1 g_k^{(2)}(r)J_c(r)dr + O_p\left(\frac{1}{|c|}\right) \qquad (10)\end{aligned}$$

since $J_c(1) \equiv N\left(0, \frac{1}{2c}\left(e^{2c}-1\right)\right) = O_p\left(1/|c|\right)$ and $(1/c)\int_0^1 g_k^{(1)}(r)dW(r) = O_p\left(1/|c|\right)$ as $c \to -\infty$. Continuing the process leading to (10) until we get to $g_k^{(k+1)}(r) = 0$, we deduce that $\int_0^1 g_k(r)dJ_c(r) = O_p\left(1/|c|\right)$. Hence, from (9) we obtain

$$Q_k^{-1}\int_0^1 g_k(r)J_c(r)dr = \frac{1}{-c}Q_k^{-1}\int_0^1 g_k(r)dW(r) + O_p\left(1/|c|^2\right)$$

Thus,

$$\widehat{Z}_c = \sigma Q_k^{-1} \int_0^1 g_k(r) J_c(r) dr = \frac{\sigma}{-c} Q_k^{-1} \int_0^1 g_k(r) dW(r) + O_p\left(1/|c|^2\right). \quad (11)$$

But

$$\begin{aligned} \widetilde{Z}_c &= \sigma \left[\int_0^1 (g_k^1(r) - c g_k(r))(g_k^1(r) - c g_k(r))' dr \right]^{-1} \int_0^1 (g_k^1(r) - c g_k(r)) dW(r) \\ &= \frac{\sigma}{-c} \left[\int_0^1 g_k(r) g_k(r)' dr \right]^{-1} \int_0^1 g_k(r) dW(r) + O_p\left(1/|c|^2\right). \quad (12) \end{aligned}$$

The stated results now follow from (11), (12) and the fact that $Q_k = \int_0^1 g_k(r) g_k(r)' dr$.

6.4 Proof of Theorem 3.1 The result follows simply from (8) using lemma 2.2 by writing

$$\begin{aligned} F_{kT}^{1/2}\left(\breve{\beta}_k - \beta_k\right) &= \left[F_{kT}^{-1/2} \Sigma_1^T \left(\tilde{z}_{kt} - \bar{\tilde{z}}_k\right)\left(\tilde{z}_{kt} - \bar{\tilde{z}}_k\right)' F_{kT}^{-1/2}\right]^{-1} \left[F_{kT}^{-1/2} \Sigma_1^T \left(\tilde{z}_{kt} - \bar{\tilde{z}}_k\right) \varepsilon_t\right] \\ &\Rightarrow \sigma \left[\int_0^1 \overline{f}_{ck}(r) \overline{f}_{ck}(r)' dr\right]^{-1} \left[\int_0^1 \overline{f}_{ck}(r) dW(r)\right]. \end{aligned}$$

6.5 Proof of Theorem 3.2 Observe that as $c \to -\infty$ the function $f_{ck}(r) = g_k^{(1)}(r) - c g_k(r)$ behaves like $-c g_k(r)$. Similarly, $\overline{f}_{ck}(r) = f_{ck}(r) - \int_0^1 f_{ck}(s) ds$ behaves like $-c\left(g_k(r) - \int_0^1 g_k(r) dr\right) = -c(g_k(r) - h_k) = -c\overline{g}_k(r)$, say. It follows that as $c \to -\infty$

$$\begin{aligned} \breve{Z}_{mc} &= \sigma \left[\int_0^1 \overline{f}_{ck}(r) \overline{f}_{ck}(r)' dr\right]^{-1} \left[\int_0^1 \overline{f}_{ck}(r) dW(r)\right] \\ &\sim \frac{\sigma}{-c} \left[\int_0^1 \overline{g}_k(r) \overline{g}_k(r)' dr\right]^{-1} \left[\int_0^1 \overline{g}_k(r) dW(r)\right]. \quad (13) \end{aligned}$$

Now the limit distribution of the OLS trend coefficient estimator with a fitted intercept is given in (5), which in the above notation is represented by the variate

$$\widehat{Z}_{mc} = \sigma \left[\int_0^1 \overline{g}_k(r) \overline{g}_k(r)' dr\right]^{-1} \left[\int_0^1 \overline{g}_k(r) J_c(r) dr\right].$$

Just as in the proof of theorem 2.5 above, we find that as $c \to -\infty$

$$\widehat{Z}_{mc} = \frac{1}{-c} \left[\int_0^1 \overline{g}_k(r) \overline{g}_k(r)' dr\right]^{-1} \int_0^1 \overline{g}_k(r) dW(r) + O_p\left(1/|c|^2\right). \quad (14)$$

Then, $\sqrt{-c}\left(\widehat{Z}_{mc}-\breve{Z}_{mc}\right)\xrightarrow{p}0$, and $R^m_{kc}\rightarrow 1$, as required.

6.6 Proof of Theorem 3.3 The efficiency ratio in this case is $R^g_{kc}\equiv \det(V^{fmols}_{kc})/\det(V^{gls}_{kc})$. Using (12), the limit variate $\widetilde{Z}_c$ can be written as

$$\widetilde{Z}_c=\frac{\sigma}{-c}\left[\int_0^1 g_k(r)g_k(r)'dr\right]^{-1}\int_0^1 g_k(r)dW(r)+O_p\left(1/|c|^2\right).$$

From this expression and (14) it follows that as $c\rightarrow-\infty$ R^g_{kc} has the limit

$$R^g_{kc}=\det\left\{\left(V^{gls}_{kc}\right)^{-1}V^{fmols}_{kc}\right\}\rightarrow\det\left[\int_0^1 g_k(r)g_k(r)'dr\right]/\det\left[\int_0^1 \overline{g}_k(r)\overline{g}_k(r)'dr\right].$$

Now

$$\left[\int_0^1 \overline{g}_k(r)\overline{g}_k(r)'dr\right]=\left[\int_0^1 g_k(r)g_k(r)'dr\right]-\left[\left(\int_0^1 g_k(r)dr\right)\left(\int_0^1 g_k(r)dr\right)'\right]=Q_k-h_kh_k'$$

and $\det(Q_k-h_kh_k')=\det(Q_k)\left(1-h_k'Q_k^{-1}h_k\right)=\left\{\det\left[\int_0^1 g_k(r)g_k(r)'dr\right]\right\}\left(1-h_k'Q_k^{-1}h_k\right)$. Hence, $R^g_{kc}\rightarrow 1/\left(1-h_k'Q_k^{-1}h_k\right)>1$. Induction shows that $1-h_k'Q_k^{-1}h_k=1/(k+1)^2$, giving the stated result.

7. References

Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman and Hall.

Canjels, E. and M. Watson (1995). "Estimating deterministic trends in the presence of serially correlated errors," forthcoming in *Review of Economics and Statistics.*

Chan, N. H. and C. Z. Wei (1987). "Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics*, 15, 1050-1063.

Durlauf, S. N. and P. C. B. Phillips (1988) "Trends versus random walks in time series analysis," *Econometrica*, 56, 1333-1354.

Elliott, G., T. J. Rothenberg and J. H. Stock (1995), "Efficient tests of an autoregressive unit root", *Econometrica.*

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

Hannan E. J. (1970). *Multiple Time Series.* New York: Wiley.

Lee C. C. and P. C. B. Phillips (1994). "Efficiency gains using GLS over OLS under nonstationarity," mimeographed, Yale University.

Phillips, P. C. B. (1987a). "Towards a unified asymptotic theory for autoregression," *Biometrika*, 74, 535–547.

Phillips, P. C. B. (1987b). "Time series regression with a unit root," *Econometrica*, 55, 277–301.

Phillips, P. C. B. and C. C. Lee (1996). "Efficiency gains from quasi-differencing under nonstationarity," mimeographed, Yale University.

Phillips, P. C. B. and P. Perron (1988). "Testing for a unit root in time series regression," *Biometrika*, 74, 535–547.

Samarov, A. and M. S. Taqqu (1988). "On the efficiency of the sample mean in long memory noise," *Journal of Time Series Analysis*, 9, 191–200.

Uhlig, H. (1994). "On Jeffreys' prior when using the exact likelihood function," *Econometric Theory*, 10, 633-644.

Yajima, Y. (1988). "On estimation of a regression model with long-memory stationary errors," *Annals of Statistics*, 16, 791–807.

Figure 1: Relative efficiency in linear trend model with near-integrated errors

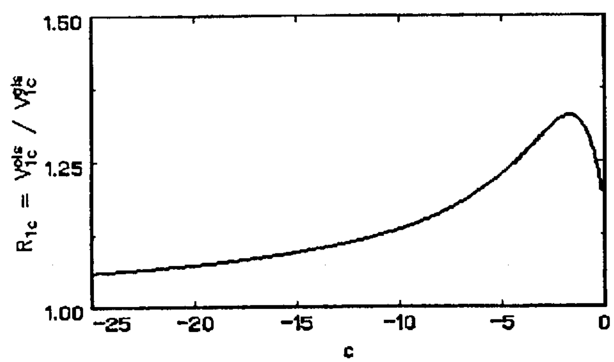

Figure 2: Relative efficiency in linear trend plus intercept model

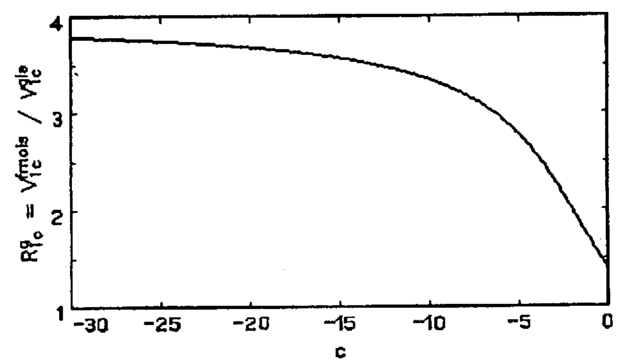

Figure 3: Relative efficiency in linear trend model with intercept and with $c \sim c_0 T^{1/2}$

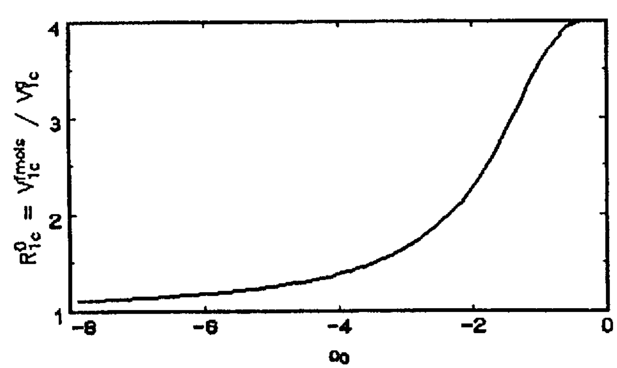

Figure 4: Efficiency curves for various τ