COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1040



# HYPER-CONSISTENT ESTIMATION OF A UNIT ROOT IN TIME SERIES REGRESSION

Peter C. B. Phillips

December 1992

# HYPER-CONSISTENT ESTIMATION OF A UNIT ROOT IN TIME SERIES REGRESSION

by

Peter C. B. Phillips*

*Cowles Foundation for Research in Economics*
*Yale University*

## 0. ABSTRACT

It is shown that the fully modified ordinary least squares (FM-OLS) estimator of a unit root in time series regression is $T^{3/2}$-consistent. Relative to FM-OLS, therefore, the least squares and maximum likelihood estimators are infinitely deficient asymptotically. Simulations show that this dominance of FM-OLS persists even in small samples.

*Keywords:* Fully modified least squares; hyper-consistent; unit root.

November, 1992

*All of the computations and graphics reported in this paper were performed by the author on a 486-33PC. My thanks go to the NSF for support under Grant No. SES 9122142 and to Glena Ames for her skill and effort in keyboarding the manuscript of the paper.

# 1. INTRODUCTION

It has been known for a long time (at least since White, 1958) that the least squares (OLS) or maximum likelihood estimator (MLE) is $T$-consistent in a Gaussian AR(1) with a unit coefficient. This result was extended by the author (1987) in earlier work to a general class of time series models with an autoregressive unit root. The limit distribution of the OLS estimator is asymmetric with a long left hand tail that accords with the well known downward bias of the estimator in finite samples. This limit theory has caused many difficulties in statistical inference and has been an obstacle to the development of a theory of optimal estimation that accommodates nonstationary regression.

The possibility that one can improve on the maximum likelihood estimator in nonstationary regression by means of a general estimation method, rather than by simple bias corrections, seems not to have been contemplated in earlier research. The present paper shows that this possibility is a real one and that a substantial gain not just of efficiency but also in the actual rate of convergence to the unit root can be achieved by the use of the fully modified (FM) least squares principle. The FM-OLS estimator was introduced in Phillips and Hansen (1990) to provide optimal estimates of single equation or subsystem cointegrating regressions in models of nonstationary economic time series. The method uses an augmented regression model that involves differences as well as levels of the regressors, but is semiparametric in its treatment of the coefficients of the differences in this expanded model.

The present paper demonstrates that when the FM-OLS method is applied to a time series regression model with a unit root the estimator is hyper-consistent in the sense that its rate of convergence exceeds the rate of $O(1/T)$ that applies for the MLE. Under general conditions we show that the FM-OLS estimator is $T^{3/2}$-consistent for a unit root and, thus, OLS and the MLE are infinitely deficient relative to this estimator. Some simulations reveal that FM-OLS produces very substantial gains in concentration probability in finite samples.

The paper also derives the limit distribution of the FM-OLS estimator in time series regression with a unit root. Unlike the limit distribution of OLS, this distribution is symmetric, so that the FM-OLS estimator is asymptotically median unbiased. The new limit distribution is characterized in terms of Brownian motion and a random variable whose distribution depends on that of the data.

The paper is organized as follows. Section 2 gives the model, the assumptions that we need and describes the FM-OLS estimator in the present context. Our main result is given in Section 3. Some simulation exercises are reported briefly in Section 4. Section 5 gives some further discussion of the results and proofs are given in Section 6.

## 2. THE MODEL, ASSUMPTIONS AND THE FM-OLS ESTIMATOR

Let $\{y_t\}$ be a scalar time series generated by

(1) $$y_t = \alpha y_{t-1} + u_t\,, \quad t = 1, 2, \ldots$$

with

(2) $$\alpha = 1\,,$$

and an initialization $y_0$ at $t = 0$ that can be any random variable including, of course, a constant. It will be convenient, but not essential, in what follows to assume that the error process $\{u_t\}$ in (1) is a linear process that satisfies

ASSUMPTION EC (*Error Condition*)

(a) $$u_t = C(L)\varepsilon_t = \Sigma_{j=0}^{\infty} c_j \varepsilon_{t-j}\,, \quad \Sigma_0^{\infty} j^2 c_j^2 < \infty\,, \quad C(1) \neq 0$$

(b) $\{\varepsilon_t\}$ *is* iid *with zero mean, variance* $\sigma_\varepsilon^2$ *and finite fourth cumulant.*

Under Assumption EC the fourth cumulants of $u_t$ are finite and absolutely summable. That is, if $\kappa_4(0, q, r, s)$ is the fourth cumulant of $(u_t, u_{t+q}, u_{t+r}, u_{t+s})$ then

(3) $$\Sigma_{q,r,s} |\kappa_4(0, q, r, s)| < \infty$$

(e.g., see Hannan, 1970, p. 211). Further, Assumption EC validates a functional central limit theorem for partial sums of $u_t$ and $u_t^2$ (e.g., see Phillips and Solo, 1992, Theorem 3.3).

Alternative conditions which allow for some heterogeneity in the process $\varepsilon_t$ and relax the independence requirement of EC(b) are also possible. A variety of alternatives of this kind that still accommodate a functional limit law for partial sums of $u_t$ are given in Phillips and Solo (1992). Yet another alternative is to replace EC(a) with mixing or near epoch dependence conditions directly on the process $u_t$, as in Phillips (1987). As indicated above, the configuration of error conditions under which our main result holds is wide and the specific assumptions used are not of great material significance in the following development

We do need to be explicit about the smoothness of the spectrum of the error process $u_t$ in the neighborhood of the origin. This is determined by the size of the parameter $q$ ($\geq 1$) for which

$$\Sigma_{j=-\infty}^{\infty}|j|^{q-1}E(u_{t+j}u_t) < \infty \ .$$

As shown in Phillips and Solo (1992, Lemma 5.10) this holds for $q = 2$ when $\Sigma_0^{\infty}jc_j^2 < \infty$ and this in turn is assured by the summability condition in EC(a).

Under EC the long run variance of $u_t$ is

$$(4) \qquad \omega^2 = \Sigma_{j=-\infty}^{\infty}E(u_0u_j) = \sigma_\varepsilon^2C(1)^2 > 0 \ .$$

Applying least squares to (1) we have, in the usual regression notation, $\hat{\alpha} = y'y_{-1}/y_{-1}'y_{-1}$. Then, as in Phillips (1987), we have the limit theory

$$(5) \qquad T(\hat{\alpha}-1) \to_d \left(\int_0^1 BdB + \lambda\right)/\int_0^1 B^2 \ ,$$

where $B \equiv BM(\omega^2)$ and $\lambda = \Sigma_0^{\infty}E(u_0u_j) = (\omega^2 - \sigma_\varepsilon^2)/2$. The estimator $\hat{\alpha}$ is $T$-consistent for $\alpha = 1$.

The FM-OLS estimator of $\alpha$ in (1) is based on the OLS estimator $\hat{\alpha}$ and has the following form

$$(6) \qquad \hat{\alpha}^+ = (y_{-1}'y_{-1})^{-1}(y_{-1}'y^+ - T\hat{\delta}^+)$$

where

$$(7) \qquad y^+ = y - \Delta y_{-1}\hat{\Omega}_{11}^{-1}\hat{\Omega}_{10} \ , \quad \text{and}$$

(8) $\qquad \hat{\delta}^{+} = \hat{\Delta}_{01} - \hat{\Omega}_{01}\hat{\Omega}_{11}^{-1}\hat{\Delta}_{11}$ .

In these formulae $\hat{\Omega}$ is an estimator of the long run variance matrix of $(u_t, \Delta y_{t-1})$ and is partitioned as

$$\hat{\Omega} = \begin{bmatrix} \hat{\Omega}_{00} & \hat{\Omega}_{01} \\ \hat{\Omega}_{10} & \hat{\Omega}_{11} \end{bmatrix} .$$

The vector $\hat{\Delta}' = (\hat{\Delta}_{01}, \hat{\Delta}_{11})$ is an estimator of the one-sided long run covariance of $u_t$ with $(u_t, \Delta y_{t-1})$, i.e.

(9) $\qquad \Delta' = (\Delta_{01}, \Delta_{11}) = \Sigma_{j=0}^{\infty} E(u_{t+j}(u_t, \Delta y_{t-1}))$ .

The FM estimator was originally developed by Phillips and Hansen (1990) to estimate cointegrating relations between I(1) variables. Essentially, the new dependent variable $y^{+}$ in (7) incorporates an endogeneity correction (to allow for the fact that the original dependent variable $y_t$ and the regressor $y_{t-1}$ in (1) are balanced in the long run since they differ by only a stationary process); and (8) embodies a serial correlation correction (to accommodate the dependence of $y_{t-1}$ on the past history of $u_t$). The motivation for this estimator is discussed at greater length in the earlier paper. Note in the present case that $y_t$ and $y_{t-1}$ are trivially cointegrated when $\alpha = 1$. (They are also trivially cointegrated when $|\alpha| < 1$ in (2) because both are stationary.)

The estimators $\hat{\Omega}$ and $\hat{\Delta}$ in (7) and (8) are designed to be consistent for $\Omega$ and $\Delta$. Given the generality of the process $u_t$, this requires the use of nonparametric methods. We also estimate $u_t$ itself by a preliminary OLS regression giving $\hat{u}_t = y_t - \hat{\alpha} y_{t-1}$. Kernel estimates of the long run covariance matrices $\Omega$ and $\Delta$ take the form

(10) $\qquad \hat{\Omega}_{ab} = \Sigma_{j=-T+1}^{T-1} w(j/K)\hat{\gamma}_{ab}(j)$ , and $\hat{\Delta}_{ab} = \Sigma_{j=0}^{T-1} w(j/K)\hat{\gamma}_{ab}(j)$ ,

where $w(\cdot)$ is a kernel function, $\hat{\gamma}_{ab}(j) = T^{-1}\Sigma' \hat{v}_{at+j}\, \hat{v}_{bt}$ (with $\hat{v}_{at} = \hat{u}_t$, $\hat{v}_{bt} = \Delta y_{t-1}$) and $K$ is a lag truncation or bandwidth parameter. The symbol "$\Sigma'$" in the expression for $\hat{\gamma}_{ab}(j)$ signifies summation over $1 \le t, t+j \le T$.

We will need to be more specific about the kernel estimates in (10). In particular, we prescribe the following class of admissible kernels.

ASSUMPTION KL (*Kernel Condition*): *The kernel function* $w(\cdot) : \mathbb{R} \to [-1, 1]$ *is a continuous even function with* $w(0) = 1$, *is continuously differentiable at all but a finite number of points of* $\mathbb{R}$ *and* $\int_{\mathbb{R}} |w(x)| dx < \infty$.

Parzen's (1957) characteristic exponent $r$ ($> 0$) of the kernel $w(\ )$ is the largest integer such that

$$\lim_{u \to 0} \frac{1 - w(u)}{|u|^r} < \infty .$$

This implies that

$$\lim_{u \to 0} \frac{dw(u)/du}{|u|^{r-1}} = w_{(r)} < \infty .$$

We will be using kernels for which $r = 2$ and this includes the Parzen, Tukey-Hanning and quadratic spectral (QS) kernels (e.g. see Priestley, 1981, p. 463). We make this explicit in the following.

ASSUMPTION LR (*Long Run Covariance Matrix Estimation*):

(a) *Any kernel function satisfying Assumption* KL *and with characteristic exponent* $r = 2$ *is used in the estimation of the long run covariance matrices* $\Omega$ *and* $\Delta$ *in* (10).

(b) *The bandwidth parameter* $K$ *in the kernel estimates* (10) *has growth rate given by*

$$K = O(T^{1/4+\kappa}) \quad \text{for some} \quad \kappa \in (0, 1/4) ,$$

*i.e.* $K^2/T \to 0$ *and* $K^4/T \to \infty$ *as* $T \to \infty$.

Part (a) of this assumption restricts our use of kernels to those whose behavior near the origin is essentially quadratic ($r = 2$). In practice, this is unlikely to be very restrictive because most of the commonly used kernels are of this type. Part (b) of the assumption controls the growth rate of the bandwidth parameter to lie within the zone $O(T^{1/4}) \ll K$

$\ll O(T^{1/2})$. This is somewhat more restrictive, since it explicitly rules out the "optimal" growth rate $O(T^{1/5})$ that applies for many of the kernels satisfying Assumptions KL and LR(a) -- e.g., see Hannan (1970, p. 286) or Andrews (1991). We can, in fact, accommodate growth rates for $K$ that include the optimal rate of $O(T^{1/5})$, but as the proof of our main result in the next section makes clear we will not achieve a $T^{3/2}$-consistency rate for the estimator $\hat{\alpha}^+$ in that event, although a $T^{1+\delta}$ for some $\delta > 0$ is still possible.

## 3. THE MAIN RESULT

The following lemma is useful in helping to establish our main theorem.

3.1 LEMMA: *Under Assumption* EC, KL *and* LR *we have* $\hat{\Omega}_{11}^{-1}\hat{\Omega}_{10} - 1 = o_p(T^{-1/2})$, as $T \to \infty$. □

As shown prior to the proof of this lemma in Section 6, both $\hat{\Omega}_{11}$ and $\hat{\Omega}_{10}$ tend in probability as $T \to \infty$ to the long run variance $\omega^2$ of $\{u_t\}$. The difference in these estimates, $\hat{\Omega}_{11} - \hat{\Omega}_{10}$, is then $o_p(T^{-1/2})$. The main requirements in achieving this order of smallness in the difference are the use of a kernel with exponent $r = 2$ and the bandwidth condition that $K = O(T^{1/4+\kappa})$ for some $\kappa \in (0, 1/4)$. When these requirements are not met, we do not achieve the $T^{3/2}$-consistency rate for the FM estimator $\hat{\alpha}^+$ that is given in the main theorem below. Instead we get a $T^{1+\delta}$-consistency rate wherein $\delta$ ($> 0$) depends on the actual growth rate of $K$ and the properties of the kernel. Since this weaker result is less interesting we give only the main theorem here.

3.2 THEOREM: *Under Assumptions* EC, KL *and* LR *the* FM-OLS *estimator* $\hat{\alpha}^+$ *is* $T^{3/2}$*-consistent for* $\alpha = 1$. *The limit distribution of* $\hat{\alpha}^+$ *is given by*

$$(11) \qquad T^{3/2}(\hat{\alpha}^+ - 1) \longrightarrow_d B(1)u_\infty / \int_0^1 B(r)^2 dr \ ,$$

*where* $B(r) \equiv BM(\omega^2)$ *and is independent of the random variable* $u_\infty$ *which has the same distribution as the error* $u_t$.

### 3.3. DISCUSSION

(i) Theorem 3.2 shows that the FM estimation procedure accelerates the convergence of the least squares regression estimator $\hat{\alpha}$ at $\alpha = 1$. Relative to $\hat{\alpha}$, the FM-OLS estimator $\hat{\alpha}^+$ is infinitely more efficient asymptotically. In effect, as $T \to \infty$ least squares requires increasingly more data to match the efficiency of $\hat{\alpha}^+$ based on $T$ observations. In the terminology of Hodges and Lehmann (1970), OLS is asymptotically deficient relative to FM-OLS at $\alpha = 1$.

(ii) When $u_t \equiv$ iid $N(0, \sigma_u^2)$ the estimator $\hat{\alpha}$ is the maximum likelihood estimator (MLE) of $\alpha$ in (1). Since the results of Theorem 3.2 apply in this case, the estimator $\hat{\alpha}^+$ dominates the MLE asymptotically. There are many examples in statistics, primarily involving non-regular estimation cases, where maximum likelihood is not asymptotically efficient. The present example is, in fact, a regular estimation problem and the log likelihood ratio of the model is locally asymptotically quadratic (LAQ) in the sense of Le Cam (see Le Cam and Yang, 1990, Section 5, for discussion of this property). However, in the neighborhood of the unit root $\alpha = 1$, the LAQ expression gives neither a normal approximation but instead involves a Gaussian functional approximation, as discussed in Phillips (1989) and Jeganathan (1989). In this case a general theory of optimality for ML estimation has not been developed. What is known is that under the random time change

$$(12) \qquad \tau_n = \inf\left\{T : \Sigma_1^T y_{t-1}^2 \geq n\right\},$$

which effectively fixes the (random) information of the data, the new estimator sequence $\hat{\alpha}_{(n)}$ that is based on $\tau_n$ observations is uniformly asymptotically normal (see Lai and Seigmund, 1983) and is asymptotically minimax in the sense of Hájek (1972) as $n \to \infty$. The latter property has recently been shown by Greenwood and Shiryaev (1991) -- see Greenwood and Wefelmeyer (1991) for some further results on this topic. Under (12) the information content of the data is rendered nonrandom by the (random) sampling scheme. It is therefore perhaps not too surprising that classical asymptotic results apply in this event. Theorem 3.2 shows that if the data is observed according to a conventional time clock then substantial efficiency gains over the MLE are possible at $\alpha = 1$. Relative to the MLE $\hat{\alpha}$, we call the estimator $\hat{\alpha}^+$ hyper-consistent at $\alpha = 1$.

(iii) Examples of super efficient estimation have been well known since the Hodges example given in Le Cam (1953). As shown by Le Cam, the set of points of super efficiency has Lebesgue measure zero in conventional $\sqrt{T}$-consistent and asymptotically normal estimation problems. In the present case, these latter conditions do not apply, and $\alpha = 1$ is already well known to be a critical point in the parameter space where $T$-consistent estimation rather than $\sqrt{T}$-consistent estimation is possible. Theorem 3.2 shows that an even higher rate of consistency is possible at $\alpha = 1$. We remark that the FM-OLS estimator $\hat{\alpha}^+$ has the same limit distribution as the MLE $\hat{\alpha}$ when $|\alpha| < 1$ and $\hat{\alpha}$ is $\sqrt{T}$-consistent for $\alpha$ (i.e. when $\{u_t\}$ $\equiv$ iid $N(0, \sigma_u^2)$). This latter property follows from a treatment in other work by the author (1991b) of the FM estimator's asymptotics in models with cointegrated and stationary regressors.

(iv) The outcome of the theorem indicates that there is information in the data about $\alpha = 1$ that is not fully exploited by the MLE. The MLE relies on a simple regression of $y_t$ on $y_{t-1}$ in (1). The FM-OLS estimator $\hat{\alpha}^+$ uses an augmented regression model that involves the difference $\Delta y_{t-1}$ as a regressor but is semi-parametric in its treatment of the coefficient of $\Delta y_{t-1}$ in this regression. The fact that $\hat{\alpha}^+$ is hyper-consistent relative to $\hat{\alpha}$ suggests that there is information in the differences $\Delta y_{t-1}$ and their long run covariance with the regression errors that is not fully exploited by maximum likelihood in the levels regression of $y_t$ on $y_{t-1}$.

(v) The limit distribution given in (11) is symmetric about the origin. This is easily seen from the fact that $B(r) =_d -B(r)$ and thus

$$B(1)u_{\infty}/\int_0^1 B(r)^2 dr =_d -B(1)u_{\infty}/\int_0^1(-B(r))^2 dr = -B(1)u_{\infty}/\int_0^1 B(r)^2 dr \ .$$

Note that in the case where $u_t \equiv$ iid $N(0, \sigma_u^2)$, the limit variate (11) has the same distribution as $W(1)\xi/\int_0^1 W(r)^2 dr$, where $W(r) \equiv BM(1)$ and $\xi \equiv N(0, 1)$. This distribution will be plotted in the next section of the paper. Given the symmetry of the limit distribution (11), we deduce that $\hat{\alpha}^+$ is an asymptotically median unbiased estimator of $\alpha = 1$.

## 4. SIMULATIONS

We report briefly some simulation evidence on the sampling properties of the FM-OLS estimator $\hat{\alpha}^+$ in comparison with the OLS estimator $\hat{\alpha}$ in the model (1) when $\alpha = 1$. Using 10,000 replications we computed $\hat{\alpha}^+$ for sample sizes $T = 75, 150, 300$ (using the corresponding fixed bandwidths $K = 3, 6, 9$ and a Parzen kernel in the computation of the long run covariance matrices) and graphed the sampling distributions in Figure 1. This figure shows kernel estimates (using a normal kernel) of the densities of $\hat{\alpha}^+$ for these sample sizes against that of $\hat{\alpha}$ for $T = 150$. The results are quite dramatic. The FM estimator $\hat{\alpha}^+$ is more concentrated about $\alpha = 1$ for the sample size $T = 75$ than the OLS estimator $\hat{\alpha}$ is when $T = 150$. As $T$ increases, the distribution of $\hat{\alpha}^+$ concentrates about $\alpha = 1$ and also becomes increasingly symmetric about unity.

- - - - - - - - - - - - - - -

Figures 1 and 2 about here

- - - - - - - - - - - - - - -

Figure 2 graphs kernel estimates of the densities of the centered and scaled quantities $T(\hat{\beta}-1)$ for $\hat{\beta} = \hat{\alpha}^+, \hat{\alpha}$. Notice that we use the $O(T)$ scaling that is appropriate for the OLS estimator $\hat{\alpha}$ in all of the graphs in this figure. As is apparent from the figure the densities of $T(\hat{\alpha}^+-1)$ continue to concentrate about the origin as $T$ increases, reflecting the higher rate of convergence of the estimator $\hat{\alpha}^+$. The density of $T(\hat{\alpha}-1)$ displays the long left hand tail that characterizes both the finite sample and limit distributions of $\hat{\alpha}$. As noted in Figure 1, the distribution of $T(\hat{\alpha}^+-1)$ is more concentrated and is more symmetric than that of $T(\hat{\alpha}-1)$.

- - - - - - - - - - -

Figure 3 about here

- - - - - - - - - - -

Figure 3 graphs kernel estimates of the densities of the FM estimator centered and scaled as $T^{3/2}(\hat{\alpha}^+-1)$. The figure also graphs the limit density given in formula (11) of Theorem 3.2 by using 10,000 replications of 500 observations generated from a Gaussian random walk to approximate the right side of (11) by the sample quantity $T^{-1}y_{T-1}u_T/T^{-2}\Sigma y_{t-1}^2$. As is apparent from the figure, the densities of $T^{3/2}(\hat{\alpha}^+-1)$ approach the limit density slowly. Even for $T = 300$ the sampling distribution of $T^{3/2}(\hat{\alpha}^+-1)$ is some way from the limit density. In spite of this slow convergence to the limit density, Figures 1 and 2 show that $\hat{\alpha}^+$ substantially dominates $\hat{\alpha}$ in terms of concentration about $\alpha = 1$ in finite samples.

## 5. IMPLICATIONS

The results of this paper have implications that seem intriguing for both theory and empirical practice in nonstationary time series regression. At the theoretical level, the dominating property of the FM-OLS estimator over the OLS and MLE estimators indicates that the maximum likelihood principle of estimation does not use information in the data fully efficiently when that data has nonstationary characteristics. To the extent that the FM estimator uses both levels and differences of the series it would appear that there is information content in the differences that is untapped by maximum likelihood. The explanation lies in the fact that the long run covariance matrix between the regression errors and differences of the regressors is zero when the regressors are stationary but non zero (and, in fact, *equal* to the long run covariance matrix of the errors) when the regressors have a unit root. The FM estimator uses this information (in the construction of a semiparametric estimate of the coefficient of the differences), whereas the MLE does not. Obviously, this is a very interesting matter of statistical theory that deserves further investigation.

At the practical empirical level, improved estimation of a unit root in time series regression seems likely to be quite important. First, there is the opportunity to develop unit root tests which exploit the new estimator and have more power than conventional tests. The distribution theory in Theorem 3.2 provides an obvious starting point here. Note that tests based on the FM estimator $\hat{\alpha}^+$ will automatically have an asymptotic power of unity in local-to-unity regions of the form $\alpha = 1+c/T$ for $c \neq 0$. This means that, at least in large samples, tests based on $\hat{\alpha}^+$ should dominate all unit root tests that are based on $T$-consistent estimators of $\alpha$. Furthermore, since the FM estimate of a unit root is hyper-consistent while the FM estimates of cointegrating coefficients are $T$-consistent and optimal (Phillips-Hansen, 1990) and those of stationary coefficients are $\sqrt{T}$-consistent (Phillips, 1991b), it would seem that these estimates can be used to construct asymptotic chi-squared tests of hypotheses such as noncausality and neutrality in vector autoregressions, which is generally not possible with OLS estimates (see: Phillips and Durlauf, 1986; Sims, Stock and Watson, 1991; and Toda and Phillips, 1991). These and other possibilities will be explored by the author in future research.

## 6. PROOFS

Throughout this section it will be convenient to assume that we are working with long run variance matrix estimates that are of the same general form, say

$$\text{(P1)} \quad \hat{\Omega} = \Sigma_{j=-K+1}^{K-1} w(j/K)\hat{\gamma}(j) \ , \quad \text{and} \quad \hat{\Delta} = \Sigma_{j=0}^{K-1} w(j/K)\hat{\gamma}(j) \ ,$$

where $\hat{\gamma}(j) = T^{1}\Sigma'\hat{v}_{t+j}\hat{v}_t'$. (P1) corresponds to (10) when the kernel is truncated, as in the case of the Parzen and Tukey-Hanning kernels, where $w(x) = 0$ for $|x| > 1$. The results given also hold for untruncated kernel estimates like QS kernel estimates, but the proofs are a little more complicated because we need to deal separately with sums over $|j| < K$ and $|j| \geq K$ in expressions like (10). To keep matters simple we will work with estimates of the form given by (P1).

In the autocovariance $\hat{\gamma}(j)$ that appears in (P1), $\hat{v}_t$ is defined by

$$\text{(P2)} \quad \hat{v}_t = \begin{bmatrix} \hat{u}_t \\ \Delta y_{t-1} \end{bmatrix} = \begin{bmatrix} u_t \\ u_{t-1} \end{bmatrix} + \begin{bmatrix} (\alpha-\hat{\alpha})y_{t-1} \\ u_{t-1} \end{bmatrix} .$$

Now $\hat{\alpha}$ is $T$-consistent for $\alpha$ and $T^{-1}\Sigma y_{t-1}u_t = O_p(1)$. Moreover, as shown in the proof of Theorem 3.1 of Phillips (1991c, pp. 432-433)

$$\text{(P3)} \quad K^{-1} \sum_{j=-K+1}^{K-1} w(j/K)(T^{-1}\Sigma' u_{t+j}y_{t-1}) = O_p(1) \ .$$

Since $K/T = o_p(T^{-1/2})$ under Assumption LR, we can therefore replace $\hat{\gamma}(j)$ in (10) by $\gamma(j) = T^{-1}\Sigma' v_{t+j}v_t'$, where $v_t = (u_t, u_{t-1})'$, up to an error of $o_p(T^{-1/2})$. That is, (P4) $\hat{\Omega} = \Sigma_{j=-K+1}^{K-1} w(j/K)\gamma(j) + o_p(T^{-1/2})$ , $\hat{\Delta} = \Sigma_{j=0}^{K-1} w(j/K)\gamma(j) + o_p(T^{-1/2})$ .

Now, using (10) we deduce that

$$\hat{\Omega} \rightarrow_p \Omega = \Sigma_{j=-\infty}^{\infty} E(v_{t+j}v_t') = \begin{bmatrix} \omega^2 & \omega^2 \\ \omega^2 & \omega^2 \end{bmatrix},$$

and

$$\hat{\Delta} \rightarrow_p \Delta = \Sigma_{j=0}^{\infty} E(v_{t+j}v_t') = \begin{bmatrix} (\omega^2 + \sigma_u^2)/2 & (\omega^2 - \sigma_u^2)/2 \\ (\omega^2 - \sigma_u^2)/2 & (\omega^2 + \sigma_u^2)/2 \end{bmatrix}.$$

### 6.1. Proof of the Lemma

Using (P1) we have

(P5) $\hat{\Omega}_{11}^{-1}\hat{\Omega}_{10} - 1 = \Sigma_{j=-T+1}^{T-1} w(j/K)[\hat{\gamma}_{10}(j) - \hat{\gamma}_{11}(j)]/\Sigma_{j=-T+1}^{T-1} w(j/\kappa)\hat{\gamma}_{11}(j)$ .

We can write the numerator of this expression as

$$\Sigma_{j=-T+1}^{T-1} w(j/K)T^{-1}(\Sigma' u_{t-1+j}\hat{u}_t - \Sigma' u_{t-1+j}u_{t-1})$$

(P6) $= \Sigma_{j=-T+1}^{T-1} w(j/K)[T^{-1}\Sigma' u_{t-1+j}u_t - T^{-1}\Sigma' u_{t-1+j}u_{t-1}] + o_p(T^{-1/2})$ ,

where we use the same argument to justify the order of magnitude of the error as we did in reaching (P4) earlier.

The sum in (P6)can be rearranged as

$$\Sigma_{j=-K+1}^{K-1} w(j/K)T^{-1}[\Sigma' u_{t-1+j}u_t - \Sigma' u_{t-1+j}u_{t-1}]$$

$$= \Sigma_{j=-K+2}^{K-1}\{w(j/K) - w((j-1)/K)\}T^{-1}\Sigma' u_{t-1+j}u_t + w((K-1)/K)T^{-1}\Sigma' u_{t-2+K}u_t$$

$$- w(-(K-1)/K)T^{-1}\Sigma' u_{t-K}u_t + o_p(T^{-1/2})$$

(P7) $= \Sigma_{j=-K+2}^{K-1}\{w(j/K) - w((j-1)/K\}T^{-1}\Sigma' u_{t-1+j}u_t + o_p(T^{-1/2})$ ,

since $E(u_{t-K}u_t) = O(K^{-2}) = o(T^{-1/2})$ under assumption LR, and because $w((K-1)/K) \rightarrow 0$ as $T$ (and hence $K$) $\rightarrow \infty$.

Over intervals where $w(\cdot)$ is continuously differentiable we have

$$w(j/K) - w((j-1)/K) = K^{-1}w'(j^*/K)$$

where $j^* \in (j-1, j)$ and is defined for each $j$. Then (P7) can be written in the form

$$\text{(P8)} \quad K^{-1}\sum_{j=-K+2}^{K-1}{}^{*} w'(j^*/K)\gamma_{11}(j-1) + \sum_{j}{}_{*} \{w(j/K) - w((j-1)/K)\}\gamma_{11}(j-1) + o_p(T^{-1/2}) ,$$

where "$\Sigma^*$" signifies summation over subintervals where $w(\cdot)$ is continuously differentiable and "$\Sigma_*$" signifies summation over the finite and bounded above (as $T, K \to \infty$) number of intervals which contain a point where $w(\cdot)$ is not continuously differentiable. Let $r_* > 0$ be such a value and suppose $j/K \to r_*$ as $T \to \infty$. Then if $M > 0$ is such that $|w'(x)| \le M$ for all points $x$ where $w(\ )$ is differentiable we have by the mean value theorem (e.g. Dieudonné (1969, Theorem 8.5.2, p. 160)

$$|(w(j/K) - w((j-1)/K))\gamma_{11}(j-1)| \le MK^{-1}|\gamma_{11}(j-1)| = o_p(T^{-1/2}) ,$$

The last order of magnitude follows because $\gamma_{11}(j-1) = \gamma_{11}(j-1) - E(u_{t+j-1}u_t) + E(u_{t+j-1}u_t) = E(u_{t+j-1}u_t) + O_p(T^{-1/2})$ and $E(u_{t+j-1}u_t) = O(K^{-2}) = o(T^{-1/2})$ because $j = [r_*K] + 1 \to \infty$ as $K \to \infty$ with $T$. This shows that the second sum involving $\Sigma_*$ in (P8) is negligible asymptotically. We can now write (P8) as

$$\text{(P9)} \quad K^{-r}\sum_{j=-K+2}^{K-1}{}^{*} \frac{w'(j^*/K)}{|j^*/K|^{r-1}}|j^*/j|^{r-1}|j|^{r-1}\gamma_{11}(j-1) + o_p(T^{-1/2}) = F_T + o_p(T^{-1/2}) , \quad \text{say.}$$

By assumptions EC, KL and LR, the expectation of the first term, $F_T$, of (P9) is easily seen to be of order $O(K^{-r}) = o(T^{-1/2})$ since $r = 2$ for the kernels considered. Moreover, as in Hannan (1970, Theorem 9, p. 280) we have

$$\lim_{T\to\infty} KT\, var(F_T) = \lim_{T\to\infty} KT\, var\left\{K^{-1}\sum_{j=-K+2}^{K-1}{}^{*} w'(j^*/K)\gamma_{11}(j-1)\right\}$$

$$\text{(P10)} \quad = \lim_{T\to\infty}\left[\frac{T}{K}\, var\left\{\sum_{j=-K+2}^{K-1}{}^{*} w'(j^*/K)\gamma_{11}(j-1)\right\}\right] = \text{constant} .$$

Combining (P10) with the order of magnitude of the mean of $F_T$ we have

(P11) $F_T = O(K^{-\tau}) + O_p(1/KT) = o_p(T^{-1/2})$ .

This shows that

$$\hat{\Omega}_{11}^{-1}\hat{\Omega}_{10} - 1 = o_p(T^{-1/2}) ,$$

as required.

### 6.2. Proof of the Theorem

Using (6)-(8) the FM-OLS estimator is

$$\hat{\alpha}^+ = (y_{-1}'y_{-1})^{-1}\left\{y_{-1}'y - y_{-1}'\Delta y_{-1}\hat{\Omega}_{11}^{-1}\hat{\Omega}_{10} - T\hat{\Delta}_{01} + T\hat{\Delta}_{11}\hat{\Omega}_{11}^{-1}\hat{\Omega}_{10}\right\} .$$

So, under (2), we have

$$T(\hat{\alpha}^+ - 1) = (T^{-2}y_{-1}'y_{-1})^{-1}\left\{[T^{-1}y_{-1}'u - (T^{-1}y_{-1}'\Delta y_{-1})\hat{\Omega}_{11}^{-1}\hat{\Omega}_{10} + [\hat{\Delta}_{11}\hat{\Omega}_{11}^{-1}\hat{\Omega}_{10} - \hat{\Delta}_{01}]\right\}$$

(P12) $$= (T^{-2}y_{-1}'y_{-1})^{-1}\{[A] + [B]\} , \text{ say} .$$

As in Phillips (1987), $T^{-2}y_{-1}'y_{-1} \xrightarrow{d} \int_0^1 B^2$ as $T \to \infty$ where $B = BM(\omega^2)$. We therefore concentrate on the numerator terms $[A]$ and $[B]$ of (P12).

We start with $[A]$. Using the lemma we obtain

$$[A] = T^{-1}y_{-1}'u - T^{-1}y_{-1}'u_{-1}[1 + o_p(T^{-1/2})]$$

$$= T^{-1}y_{-1}'u - (T^{-1}y_{-2}'u_{-1} + T^{-1}u_{-1}'u_{-1}) + o_p(T^{-1/2})$$

(P13) $$= T^{-1}y_{T-1}u_T - T^{-1}u_{-1}'u_{-1} + o_p(T^{-1/2}) .$$

Next, turn to $[B]$. We have

$$[B] = \hat{\Delta}_{11}\hat{\Omega}_{11}^{-1}\hat{\Omega}_{10} - \hat{\Delta}_{01} = \left\{\Sigma_{j=0}^{K-1}w(j/K)\hat{\gamma}_{11}(j)\right\}\left\{1 + o_p(T^{-1/2})\right\} - \Sigma_{j=0}^{K-1}w(j/K)\hat{\gamma}_{01}(j)$$

$$= \Sigma_{j=0}^{K}w(j/K)\left\{\gamma_{11}(j) - \gamma_{01}(j)\right\} + o_p(T^{-1/2}) ,$$

in view of (P4) and the lemma. This last expression can be written directly as

$$\Sigma_{j=0}^{K-1} w(j/K)\left\{T^{-1}\Sigma' u_{t-1+j}u_{t-1} - T^{-1}\Sigma' u_{t+j}u_{t-1}\right\} + o_p(T^{-1/2})$$

$$= w(-1/K)T^{-1}\Sigma u_{t-1}^2 - w((K-1)/K)T^{-1}\Sigma' u_{t+K-1}u_{t-1} + o_p(T^{-1/2})$$

$$\text{(P14)} \quad = T^{-1}\Sigma u_{t-1}^2(1 + O(K^{-2})) + o_p(T^{-1/2}) \ ,$$

since $1 - w(-1/K) = O(K^{-2})$ and $E(u_{t+K-1}u_{t-1}) = O(K^{-2})$, both of which are $o(T^{-1/2})$ under assumption KL, and $w((K-1)/K) \to 0$ as $K \to \infty$.

Combining (P13) and (P14) we have

$$\text{(P15)} \quad [A] + [B] = T^{-1}y_{T-1}u_T + o_p(T^{-1/2}) \ .$$

Now

$$T^{-1/2}y_{T-1} \rightarrow_d B(1) \ , \quad \text{and} \quad u_T \rightarrow_d u_\infty \ ,$$

so that (P12), (P14) and (P15) lead to

$$T^{3/2}(\hat{\alpha}^+ - 1) \rightarrow_d B(1)u_\infty/\int_0^1 B^2 \ ,$$

which establishes the theorem.

## 7. REFERENCES


Andrews, D. W. K. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, 59, 817-858.

Dieudonné, J. (1969). *Foundations of Modern Analysis*. New York: Academic Press.

Greenwood, P. E. and A. N. Shiryaev (1991). "Asymptotic minimaxity of a sequential estimator for a first order autoregressive model," *Stochastics* (forthcoming).

Greenwood, P. E. and W. Wefelmeyer (1991). "Asymptotic minimax results for stochastic process families with critical points," University of British Columbia, mimeographed.

Hájek, J. (1972). "Local asymptotic minimax and admissibility in estimation," *Proceedings of Sixth Berkeley Symposium in Mathematical Statistics and Probability*, 1, 175-194.

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

Hodges, J. L. and E. L. Lehmann (1970). "Deficiency," *Annals of Mathematical Statistics*, 41, 783-801.

Jeganathan, P. (1989). "Some aspects of asymptotic theory with applications to time series models," mimeographed, University of Michigan.

Lai, T. L. and D. Seigmund (1983). "Fixed accuracy estimation of an autoregressive parameter," *Annals of Statistics*, 11, 478-485.

Le Cam, L. (1953). "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates," *University of California Publications in Statistics*. 1. 277-330.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. New York: Springer Verlag.

Le Cam, L. and G. L. Yang (1990). *Asymptotics in Statistics: Some Basic Concepts*. New York: Springer Verlag.

Parzen, E. (1957). "On consistent estimates of the spectrum of a stationary time series," *Annals of Mathematical Statistics*, 28, 329-348.

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

Phillips, P. C. B. (1989). "Partially identified econometric models," *Econometric Theory*, 5, 181-240.

Phillips, P. C. B. (1991a). "Optimal inference in cointegrated systems," *Econometrica*, 59, 283-306.

Phillips, P. C. B. (1991b). "Time series regression with cointegrated regressors," Yale University, mimeographed.

Phillips, P. C. B. (1991c). "Spectral regression for cointegrated time series," in W. Barnett, J. Powell and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Economics and Statistics*. New York: Cambridge University Press.

Phillips, P. C. B. and S. N. Durlauf (1986). "Multiple time series with integrated variables," *Review of Economic Studies*, 53, 473-496.

Phillips, P. C. B. and V. Solo (1992). "Asymptotics for linear processes," *Annals of Statistics*, 20, 971-1001.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Volumes I and II. New York: Academic Press.

Sims, C. A., J. H. Stock and M. W. Watson (1990). "Inference in linear time series models with some unit roots," *Econometrica*, 58, 113-144.

Toda, H. and P. C. B. Phillips (1991). "Vector autoregressions and causality," Cowles Foundation Discussion Paper No. 977.

White, J. S. (1958). "The limiting distribution of the serial correlation coefficient in the explosive case," *Annals of Mathematical Statistics*, 29, 1188-1197.

Figure 1: Densities of FM_OLS & OLS

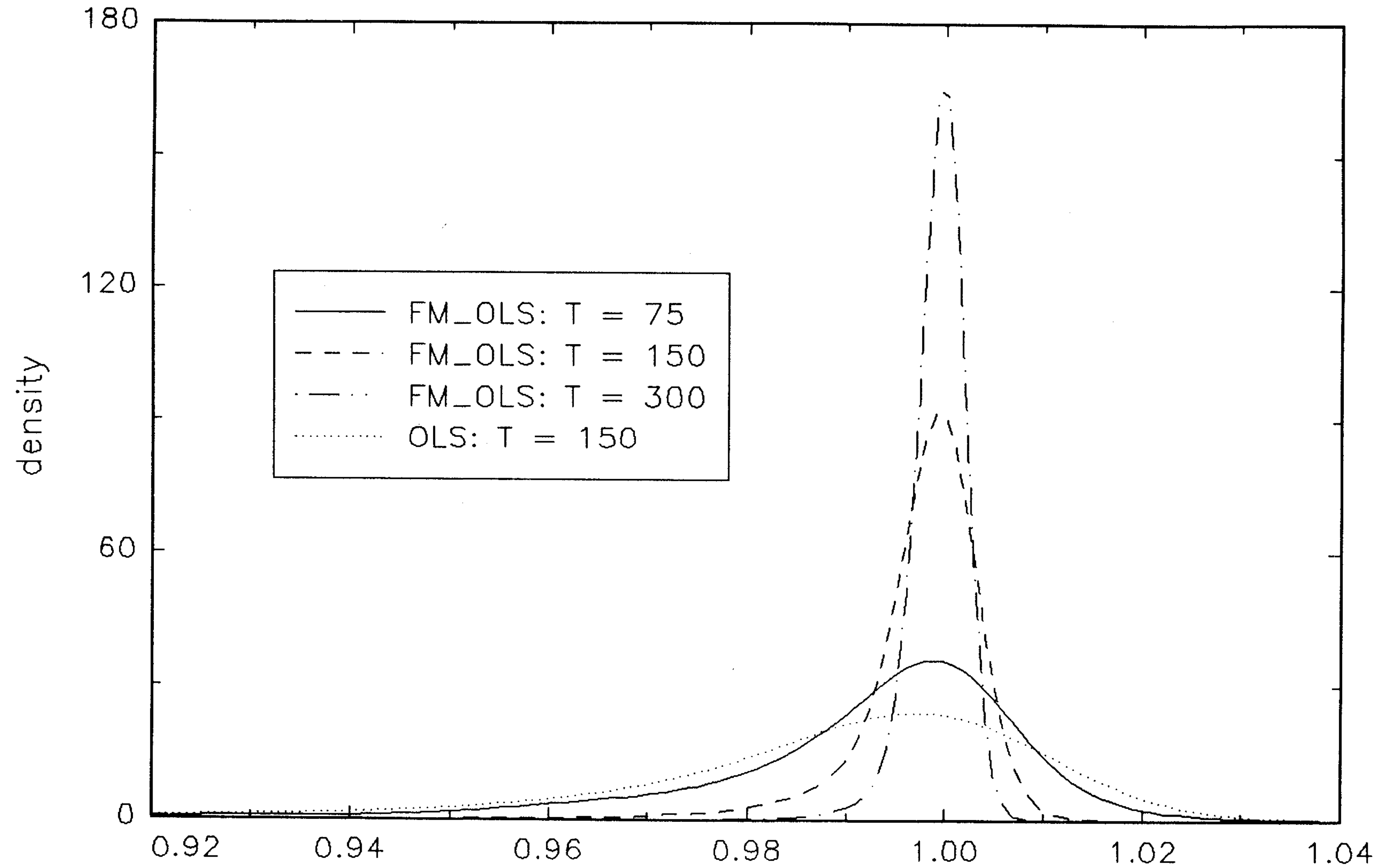

Figure 2: Densities of FM_OLS & OLS
all scaled as: T*(ahat − 1)

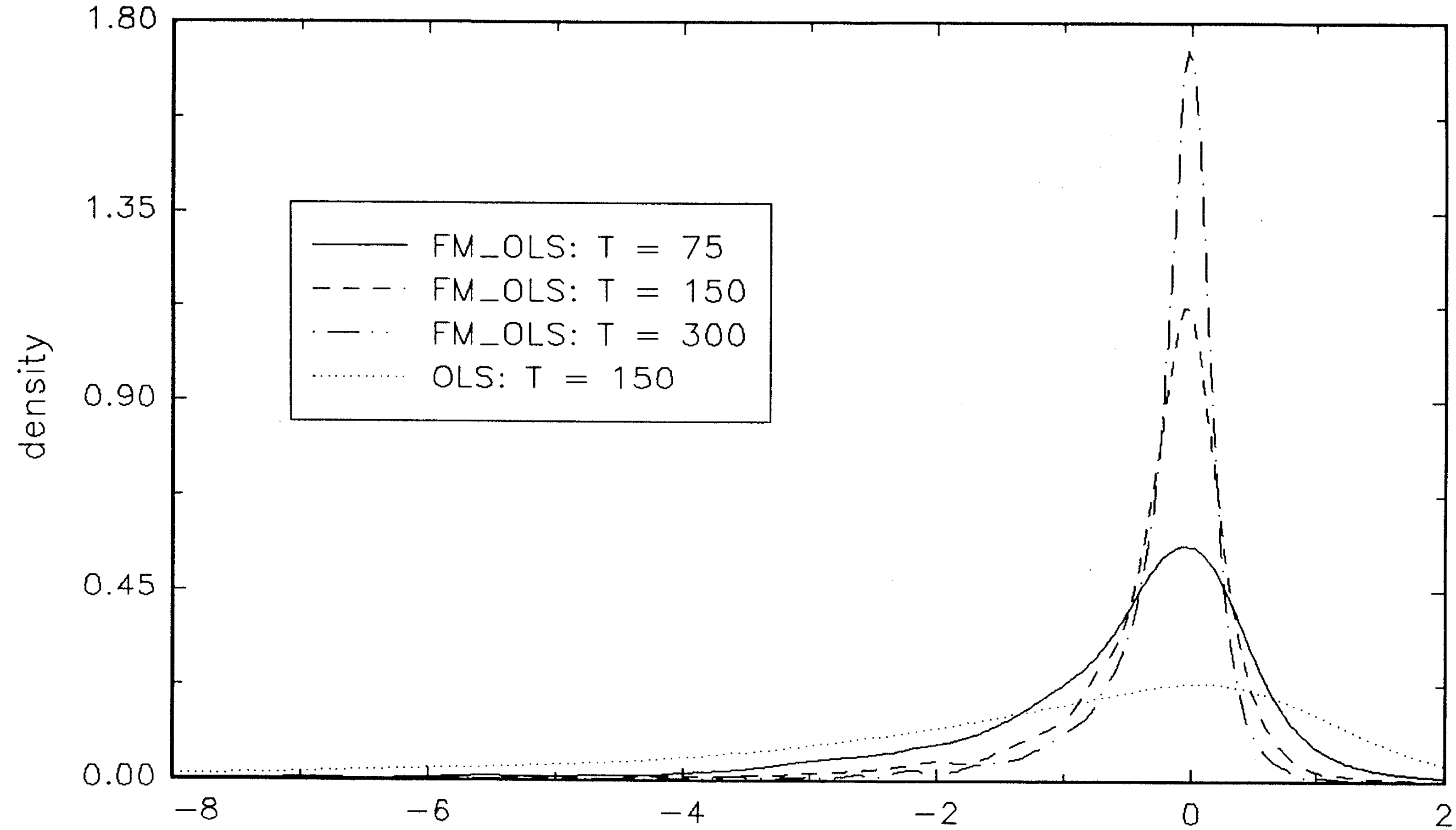

Figure 3: Sample and Limit densities of FM_OLS: T^1.5*(ahat$^{+}$ − 1)