Estimators for Persistent and Possibly Non-Stationary Data with Classical Properties

Yuriy Gorodnichenko [*] Anna Mikusheva[†] Serena Ng [‡]

First Draft: June 2007
This draft: July 2009

Abstract

This paper finds three estimators of the largest autoregressive root that are $\sqrt{T}$ consistent and asymptotically normal even in the local-to-unity framework. The point of departure is that the estimators are based on moments of data that are stationary when evaluated at the true parameter vector. Critical values from the normal distribution can be used for hypothesis testing, irrespective of the treatment of the deterministic terms. Simulations show that the estimates are approximately mean unbiased and the t test has good size in the parameter region when the least squares estimates usually yield distorted inference.

JEL Classification: C1

Keywords: Persistence, unit root, cointegration, predictability tests, local-to-unity

[*]Department of Economics, University of California at Berkeley, Berkeley, CA 94720 Email: ygorodni@econ.berkeley.edu.

[†]Department of Economics, M.I.T., 50 Memorial Drive, Building E52, Cambridge, MA, 02142. Email: amikushe@mit.edu

[‡]Department of Economics, Columbia University, 420 West 118 St., New York, NY 10027 Email: Serena.Ng@Columbia.edu. Ng acknowledges financial support from the National Science Foundation (SES-0549978).

1 Introduction

Estimators of the autoregressive parameter of the AR(1) model are generally thought to be super-consistent when there is a unit root, but that the asymptotic distribution is non-standard. This paper finds an estimator that has a convergence rate of $T^{3/4}$ and has a non-standard distribution, and three estimators with a convergence rate of $\sqrt{T}$ with t-statistics that are asymptotically normal.

The primary reason for the classical properties of our estimators is that the moments evaluated at the true parameter vector are stationary, and a central limit theorem applies. The studentized estimators are asymptotically standard normal, allowing one to conduct inference without deciding *a priori* whether or not the regressors are non-stationary. The same set of critical values can also be used irrespective of how the deterministic trend function is specified. Because our estimators are not super-consistent when the regressors are truly non-stationary, not surprisingly, this slower rate of convergence translates into power loss when the unit root hypothesis is being tested. This is the price we pay for practical simplicity and robustness. However, these properties can be highly useful in applied work because the answers to many macroeconomic questions are sensitive to assumptions about the nature of the trend and to whether the corresponding regressions are run in levels or in first-differences (see Han, Phillips, and Sul (2009a,b)). A general method that would work uniformly well whether the regressors are integrated (with a unit root)), stationary or nearly-integrated may still be attractive.

Our discussion focuses on the AR(1) model $y_t = \alpha_0 y_{t-1} + e_t$, $e_t \sim i.i.d.(0,1)$. Most linear estimators of α are $\sqrt{T}$ consistent and asymptotically normal when $\alpha_0 < 1$ but are super-consistent and non-normal when $\alpha_0 = 1$. Our main estimator is denoted 'QD' and it adopts a method of moments setup. Let w_t be the data and let θ_0 denote the true parameter vector of dimension K. Consider a vector of $M \times 1$ moments $g(w_t;\theta) = g_t(\theta)$ and assume that $Eg(w_t;\theta_0) = 0$. Let $\bar{g}(\theta) = \frac{1}{T}\sum_{t=1}^{T} g_t(\theta)$ be the vector of sample moments evaluated at an arbitrary θ. The generalized methods of moments estimator using a $M \times M$ positive definite weighting matrix W_T is

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmin}}\, Q_T(\theta), \qquad Q_T(\theta) = \bar{g}(\theta)' W_T \bar{g}(\theta). \tag{1}$$

Whereas the standard theory assumes w_t is stationary ergodic, we allow w_t to be possibly non-stationary. The classical OLS estimator uses the moment condition $E\left[(y_t - \alpha y_{t-1})y_{t-1}\right] = E\left[e_t(\alpha)y_{t-1}\right] = 0$. The corresponding $Q_T(\alpha)$ is asymptotically bounded at the true value α_0, but asymptotically explodes at $\alpha \neq \alpha_0$ when α_0 equals unity or is in the $1/T$ neighborhood of one. What we need for a $\sqrt{T}$ consistent and normal estimator is that $Q_T(\theta)$ converges to $Q(\theta)$ uniformly in θ and the cental limit theorem holds for $\widehat{\theta}$ at the true value θ_0.

Our main idea is to use a stationary instrument in place of y_{t-1} in the moment condition. For example, our FD estimator uses the moment condition $E\left[e_t(\alpha)\Delta y_{t-1}\right]=0$, while three other estimators use $E\left[e_t(\alpha)e_{t-1}\right]=0$. The latter moment condition arguably has more information than the former since Δy_{t-1} may be over-differenced in the stationary case. As e_{t-1} is not observed, we suggest two ways of implementing this. The first is a two-step HD estimator whereby one first proxies e_{t-1} by some first stage residuals $\widehat{e}_{t-1}$ (such as the OLS) so that the moment condition is linear in α. The second and more direct implementation is to perform non-linear estimation and estimate the error term e_{t-1} simultaneously with α. But because the objective function is quadratic in α, it explodes in the neighborhood of the unit root process. We show that such an estimator converges at rate $T^{3/4}$ and has a non-standard distribution. However, we observe that the higher lag auto-covariances have the same exploding summand. By subtracting the variance from all autocovariances, we are able to obtain a $\sqrt{T}$ consistent QD estimator that is asymptotically normal.

The key to the proposed estimators is to find moments such that central limit theorem holds, at least when the sample moments are evaluated at the true parameter vector. We begin in Section 2 by first considering an exactly identified model, abstracting from deterministic terms and serial correlation in the errors. The general AR(p) case with deterministic terms will then be discussed. Extensions to predictive regressions will also be considered.

2 The AR(1) Model

Consider the simple AR(1) without deterministic terms, so that

$$y_t=\alpha_0 y_{t-1}+e_t, \quad e_t\sim i.i.d.\ (0,\sigma^2). \tag{2}$$

When $\alpha_0<1$, standard asymptotic theory holds, and $T^{-1}\sum_{t=1}^T y_{t-1}^2 \xrightarrow{p} E(y_t^2)=\frac{\sigma^2}{1-\alpha_0^2}$. In the local-to-unity framework with $\alpha_0=1+c/T$, the functional central limit theorem holds that $\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]}e_t \Rightarrow \sigma W(r)$ and thus $T^{-2}\sum_{t=1}^T y_{t-1}^2 \Rightarrow \sigma^2\int_0^1 J_c(r)^2dr$, where $\Rightarrow$ denotes weak convergence in distribution, $J_c(r)$ is an Ornstein-Uhlenbeck process generated by the standard Brownian motion $W(r)$ defined on $C[0,1]$, the space of continuous functions on $[0,1]$. The fact that the sample moments have different properties when $\alpha_0<1$ and when $\alpha_0=1$ have been the basis of many unit root tests. The distribution of OLS estimator has a discontinuity at $\alpha_0=1$ which makes inference difficult.

2.1 The QD Estimator

Our main estimator is based on matching the sample moments of some transformation of the data with the moments of the model under the same transformation. For the AR(1) model, define $e_t(\alpha) = y_t - \alpha y_{t-1}$ as the quasi-difference of y_t. Let

$$\gamma_j(\alpha) = E(e_t(\alpha)e_{t-j}(\alpha))$$

be the autocovariance of e_t at lag j. The sample analog in terms of observed variables is

$$\widehat{\gamma}_j(\alpha) = \frac{1}{T}\sum_{t=1}^{T}(y_t - \alpha y_{t-1})(y_{t-j} - \alpha y_{t-j-1}).$$

Consider the estimator

$$\widehat{\alpha}_J^{QD_0} = \arg\min_{\alpha}\sum_{j=1}^{J}\left(\widehat{\gamma}_j - \gamma_j\right)^2.$$

Also define

$$\widehat{\alpha}_J^{QD} = \underset{\alpha}{\operatorname{argmin}}\sum_{j=1}^{J}\bar{g}_j^{QD}(\alpha)'\bar{g}_j^{QD}(\alpha)$$

where

$$\bar{g}_j^{QD}(\alpha) = \widehat{\gamma}_j(\alpha) - \widehat{\gamma}_0(\alpha) = \frac{1}{T}\sum_{t=1}^{T}(y_t - \alpha y_{t-1})\Big[(y_{t-j} - \alpha y_{t-j-1}) - (y_t - \alpha y_{t-1})\Big].$$

The estimators are labelled 'QD' since they are based on the quasi-differences of y_t. The primary difference between QD_0 and QD is that the former does not subtract $\widehat{\gamma}_0(\alpha)$ from $\widehat{\gamma}_j(\alpha)$.

Proposition 1 *Let y_t be generated as in (2). Then*

i. $\sqrt{T}(\widehat{\alpha}_J^{QD} - \alpha_0) \Rightarrow N(0, \sigma_J^2)$ *uniformly over* $|\alpha_0| \leq 1$*, where* $\sigma_J^2 = \frac{\left(\sum_{j=1}^J \alpha_0^{(j-1)}\right)^2 + \sum_{j=1}^J \alpha_0^{2(j-1)}}{\left(\sum_{j=1}^J \alpha_0^{2(j-1)}\right)^2}$.

ii. $\widehat{\alpha}_J^{QD_0}$ *is super-consistent in the local-to-unity framework with* $\alpha_0 = 1 + c/T$*. In particular,*

$$\begin{cases} T^{3/2}(\widehat{\alpha}_J^{QD_0} - \alpha_0)^2 \Rightarrow \frac{-\xi}{\int_0^1 J_c^2(s)ds} & \text{if } \xi < 0 \\ T(\widehat{\alpha}_J^{QD_0} - \alpha_0) \Rightarrow \frac{1 + 2\int_0^1 J_c(s)dW(s)}{2\int_0^1 J_c^2(s)ds} & \text{if } \xi > 0 \end{cases} \tag{3}$$

where J_c is an Ornstein-Uhlenbeck process generated by the Brownian motion W that is independent of $\xi \sim N(0, 1/J)$.

Proposition 1 shows that it is possible to have estimators of α that converges at a rate slower than T and also not normally distributed, but an estimator of α with classical properties is also possible. The proposition implies that when $J = 1$, then uniformly over $|\alpha_0| \leq 1$,

$$\sqrt{T}(\widehat{\alpha}_1^{QD} - \alpha_0) \Rightarrow N(0, 2)$$

The two QD estimators are non-standard covariance structure estimators. Usually, the unknown parameters enter only the analytical covariances of the model, which in our case, is $\gamma_1(\alpha)$. With our estimator, α enters the sample covariance $\widehat{\gamma}_1(\alpha)$ as well. If $|\alpha_0| << 1$, our estimator is easily shown to have classical properties under regularity conditions. But in the local-to-unity framework, $\bar{g}(\alpha)$ is not well behaved for all values of α. Thus while $\widehat{\alpha}^{QD_0}$ is consistent, the asymptotic distribution is non-standard. However, the problematic term that frustrates a quadratic expansion of $\widehat{\gamma}_j(\alpha)$ around α_0 is asymptotically collinear with the corresponding term in $\widehat{\gamma}_0(\alpha)$. Note that $\widehat{\gamma}_j(\alpha) - \widehat{\gamma}_0(\alpha)$ is the sample analog of $E(e_t(\alpha)[e_{t-j}(\alpha) - e_t(\alpha)])$, and the sample mean of $e_{t-j}(\alpha) - e_t(\alpha)$ is bounded for all values of α. This leaves us with a $\bar{g}^{QD}(\alpha)$ that, when evaluated at α_0, obeys a central limit theorem. Proposition 1 follows.

The QD estimator has an asymptotic distribution that is normal and continuous at $\alpha_0 = 1$. Few estimators of α are $\sqrt{T}$ consistent and asymptotically normal when $\alpha_0 = 1$. There are two exceptions. So and Shin (1999) considered $\widehat{\alpha} = \frac{\sum_{t=2}^T x_t y_t}{\sum_{t=1}^T x_t y_{t-1}}$ while Phillips and Han (2008) considered $\widehat{\alpha} = \frac{\sum_{t=2}^T \Delta y_{t-1}(2\Delta y_t + \Delta y_{t-1})}{\sum_{t=2}^T (\Delta y_{t-1})^2}$. Both are linear estimators and can also be seen as using stationary instruments.[1] Our QD estimator is non-linear; it is motivated by the intuitive fact that $E(e_t e_{t-1}) = 0$ when evaluated at the true value of α. Our key to asymptotic normality is that the sample analog of $E(e_t e_{t-1}) = 0$ obeys a central limit theorem. We now consider linear estimators based on this same moment.

2.2 A Hybrid Estimator

Recall that the least squares estimator uses the moment $g_t(\alpha) = y_{t-1}e_t$. Although y_{t-1} is orthogonal to e_t, the sample moment has a random limit in the local-to-unity framework. Suppose e_{t-1} was observed and we replace y_{t-1} by e_{t-1} in the moment condition. It is uncorrelated with $e_t(\alpha)$ and is hence a valid instrument. The only problem is that e_{t-1} is not observed. To resolve this problem, we use the fact that the least squares estimator $\widehat{\alpha}^{OLS}$ is consistent for all $|\alpha_0| \leq 1$. Thus, let $\widetilde{e}_{t-1} = y_t - \widehat{\alpha}^{OLS} y_{t-1}$. Conveniently, generated instruments do not require a correction for the

[1] Choi (1993) showed that the least squares estimates of the autoregressive coefficients are $\sqrt{T}$ consistent and asymptotically normal, even though the sum of the autoregressive coefficients has a Dickey-Fuller type distribution when the true sum is unity. For cointegration regressions, Laroque and Salanie (1997) used two OLS regressions in stationary variables to obtain a $\sqrt{T}$ consistent estimate of the cointegrating vector.

standard errors like generated regressors. Let $\widehat{\alpha}^{HD}$ be a hybrid quasi-difference estimator satisfying

$$\bar{g}^{HD}(\widehat{\alpha}^{HD}) = \frac{1}{T}\sum_{t=k}^{T}\widetilde{e}_{t-k}(y_t - \widehat{\alpha}^{HD}y_{t-1}) = 0.$$

It is easy to see that

$$\widehat{\alpha}^{HD} = \frac{\sum_{t=k}^{T} y_t\widetilde{e}_{t-k}}{\sum_{t=k}^{T} y_{t-1}\widetilde{e}_{t-k}} = \alpha_0 + \frac{\sum_{t=k}^{T} e_t(\alpha_0)\widetilde{e}_{t-k}}{\sum_{t=k}^{T} y_{t-1}\widetilde{e}_{t-k}}.$$

Consistency of $\widehat{\alpha}^{HD}$ follows from the fact that $\widehat{\gamma}_k(\alpha_0)\xrightarrow{p}\gamma_k(\alpha_0)$. We refer to $\widehat{\alpha}^{HD}$ as a hybrid estimator because it is based on the covariance between the quasi-difference of y_t and a stationary random variable. The objective function is now linear in α. This is unlike QD_0 which is based on the product of two quasi-differenced variables. It is the quadratic term in this product that makes the distribution of QD_0 non-standard. It is then straightforward to show that HD behaves in the classical way even in local-to-unity region:

Proposition 2 *Let y_t be generated as in (2) with $\alpha_0 = 1 + c/T$. Then*

$$\sqrt{T}(\widehat{\alpha}^{HD} - \alpha_0) \Rightarrow 2(1 + J_c(1)^2)^{-1}N(0,1). \tag{4}$$

To construct the t-statistic one can use

$$\widehat{Avar}(\widehat{\alpha}^{HD}) = \widehat{\sigma}^2\left(T^{-1}\sum_{t=1}^{T} y_{t-1}\widetilde{e}_{t-1}\right)^2,$$

where $\widehat{\sigma}^2$ is a consistent estimate of σ^2. Then $t^{HD} \Rightarrow N(0,1)$.

2.3 The FD Estimator

The HD is a two-step IV estimator. Consider a one-step estimator using Δy_{t-1} as instrument:

$$\widehat{\alpha}^{FD} = \frac{\sum_{t=2}^{T} y_t\Delta y_{t-1}}{\sum_{t=2}^{T} y_{t-1}\Delta y_{t-1}}. \tag{5}$$

Under the AR(1) model,

$$\widehat{\alpha}^{FD} = \alpha_0 + \frac{T^{-1}\sum_{t=2}^{T} e_t\Delta y_{t-1}}{T^{-1}\sum_{t=2}^{T} y_{t-1}\Delta y_{t-1}}.$$

Consistency follows from the fact that $T^{-1}\sum_{t=1}^{T} e_t\Delta y_{t-1}\xrightarrow{p}E(e_t\Delta y_{t-1})$, which is zero. In the local to unity framework,

$$\sqrt{T}(\widehat{\alpha}^{FD} - \alpha_0) \Rightarrow 2(1 + J_c(1)^2)^{-1}N(0,1).$$

The estimator again has an asymptotic variance that is random, but can be estimated by

$$\widehat{Avar}(\widehat{\alpha}^{FD}) = \widehat{\sigma}^2 \left(T^{-1} \sum_{t=1}^{T} y_{t-1} \Delta y_{t-1} \right)^2,$$

$\widehat{\sigma}^2$ is a consistent estimate of σ^2. The standardized estimator is again approximately normal.

We have suggested three estimators that are $\sqrt{T}$ consistent and asymptotically normal in the local to unity framework. These estimators also have classical properties in the standard asymptotic framework when α_0 is strictly bounded from the unit circle. In particular, $\sqrt{T}(\widehat{\alpha}^{QD_0} - \alpha_0) \xrightarrow{d} N(0,1)$, $\sqrt{T}(\widehat{\alpha}^{QD} - \alpha_0) \xrightarrow{d} N(0,2)$, $\sqrt{T}(\widehat{\alpha}^{HD} - \alpha_0) \xrightarrow{d} N(0,1)$, and $\sqrt{T}(\widehat{\alpha}^{FD} - \alpha_0) \xrightarrow{d} N(0, 2(1+\alpha_0))$. Notably, these are inefficient estimators relative to OLS, since $\sqrt{T}(\widehat{\alpha}^{OLS} - \alpha_0) \xrightarrow{d} N(0, 1-\alpha_0^2)$. However, the QD, HD, and FD estimators all have stanardized distributions that are continuous in α_0. Testing the hypothesis that $\alpha_0 = 0.95$ is as simple as testing the hypothesis that $\alpha_0 = 0.5$. The practical appeal is that asymptotic normality permits standard inference. The usual critical values of -1.64 and -2.32 can be used when the significance level of the test is 5 and 1 percent, respectively. We will see in simulations that their size and power properties are stable throughout the parameter space of α.

3 Deterministic Terms and Correlated Errors

We now consider a more general data generating process

$$\begin{aligned}
y_t &= d_t + x_t, && (6)\\
x_t &= \alpha_0 x_{t-1} + u_t, \quad \beta(L) u_t = e_t, \\
\alpha_0 &= 1 + \frac{c}{T}, && (7)\\
\beta(L) &= 1 - \beta_1 L - \ldots - \beta_p L^p, \quad \sum_{j=0}^{\infty} |\beta_j| < \infty, \\
e_t &\sim iid(0, \sigma^2) \quad \omega^2 = \sigma^2 (1 - \beta(1)^2)^{-1}
\end{aligned}$$

where $e_0 = 0$, $E(x_0^2) < \infty$, $\beta(L) = \psi(L)^{-1}$ with $\sum_{j=0}^{\infty} |\psi_j| < \infty$, $e_t \sim iid(0, \sigma^2)$. The non-normalized spectral density at frequency zero of u_t is given by $\omega^2 = \sigma^2(1 - \beta(1))^{-2}$. The deterministic terms are captured by $d_t = \sum_{j=0}^{r} \delta_j t^j$ where r is the order of the deterministic trend function. We focus on the intercept only case with $d_t = \delta_0$ and the linear trend case with $d_t = \delta_0 + \delta_1 \; t$. Hereafter, we let $\theta = (\alpha, \sigma^2, \beta_1, \ldots, \beta_p)$ be the $K \times 1$ vector of parameters of the model. The true parameter vector is denoted θ_0 and the correct lag length is denoted p_0.

Let $\widehat{x}_t = y_t - \widehat{d}_t$, the residuals from a projection of y_t on the deterministic terms. Now when an intercept is included, the Brownian motion in G_0 will be replaced by a demeaned Brownian motion. For example, $\bar{G}_T^{HD}(\alpha_0) \Rightarrow G_0^{HD} = -\frac{\sigma^2}{2}(\bar{J}_c(1)^2+1)$ where $\bar{J}_c(r) = J_c(r) - \int_0^1 J_c(r)dr$. The extension to the linear trend case is similar, with the obvious replacement of $\bar{J}(r)$ by $\widetilde{J}_c(r)$, where $\widetilde{J}_c(r) = J_c(r) - \int_0^1 (4-6s)J_c(s)ds - r\int_0^1(12-6s)W(s)ds$ is a detrended standard Brownian motion. GLS and recursive detrending can easily be accommodated. The population quantity of the detrended variables

$$\gamma_k(\alpha) = E(x_t - \alpha x_{t-1})(x_{t-k} - \alpha x_{t-k-1})$$

for $k \geq 0$ can be replaced by the sample analog,

$$\widehat{\gamma}_k(\alpha) = \frac{1}{T}\sum_{t=1}^{T}(\widehat{x}_t - \alpha\widehat{x}_{t-1})(\widehat{x}_{t-k} - \alpha\widehat{x}_{t-k-1}).$$

When u_t is serially correlated and $\beta(L)$ is a finite p-th order polynomial in L, we simply match up to p autocovariances $\gamma_k(\alpha)$, $k = 1, \ldots, p$. As with any estimator of the autoregressive model, the lag length p is important. To see how p affects inference, suppose $p_0 = 2$, so the DGP is $y_t = \rho_0 y_{t-1} - b_{0,1}\Delta y_{t-1} + e_t$. If the researcher (wrongly) assumes $p = 1$, then $\widehat{\gamma}_k - \gamma_k$ will not be zero for any $k > 1$. The J test of overidentifying restrictions provides a natural guide to the selection of p. In this sense, p is no longer a nuisance parameter but is chosen to satisfy the moment conditions. Estimates corresponding to a J test that rejects the moment conditions should be disregarded.

4 Simulations

We now use simulations to illustrate the properties of our estimators. For D=QD_0, QD, HD or FD, let

$$\bar{g}^D(\theta, p) = (\bar{g}_1^D(\theta, p), \ldots, \bar{g}_M^D(\theta, p))'$$

where M is the number of lagged autocovariances. The estimators are constructed as

$$\widehat{\theta}^D = argmin_{\theta\in\theta} \;\; \bar{g}^D(\theta, p)' W_T \bar{g}^D(\theta, p).$$

We simulate data as follows:

$$\begin{aligned} y_t &= d_t + x_t \\ (1-\alpha L)x_t &= e_t, \quad e_t \sim N(0,1) \end{aligned}$$

The parameter of interest is α. We bound the parameter space for α to $[-1.5, 1.5]$. If the converged estimates are outside of this range, we change the starting value up to 3 times. Very rarely does an estimate falls outside of this range. The simulations are based on 2,000 replications.

As a point of reference, OLS estimates for ρ are reported. These are based on the regression

$$y_t = \delta_0 + \delta_1 t + \rho y_{t-1} + \sum_{j=1}^{p-1} b_j \Delta y_{t-j} + error.$$

For QD and QD_0 estimators we use two-step GMM (optimal weighting matrix), while for HD and FD estimators we use an identity weighting matrix.

Table 1 reports the mean estimates for the intercept and the linear trend model when $T = 200$ and 500, along with the J test for overidentifying restrictions. The downward bias in OLS estimates when the data are highly persistent is well known. The HD and FD estimators are also downward biased, but less so than OLS. The QD is the most accurate of the estimators considered. Note that the QD_0 is generally precise, even though its distribution is non-normal. The last panel of Table 1 shows that the J statistic has the correct size.

Table 2 reports the finite sample power for one sided tests at alternatives evaluated at $\alpha = \alpha_0 - .05$ and $\alpha = \alpha_0 - .10$. The high power reported for OLS when α exceeds .9 mainly reflects size distortions. The result that stands out is that the size of the proposed tests is quite uniform over the entire parameter space. This permits testing if $H_0 : \alpha = .95$, for example, a parameter region when least squares based tests have highly distorted size.

Ultimately, the proposed estimators are useful only if the potential for a more accurate size when the data are highly persistent does not come at the cost of power loss outside of the persistent range. This turns out to be the case. For example, when α_0 is .5, the power of OLS, QD, and HD are quite similar, even though the power of FD is somewhat lower. The reason is that when α_0 is far from the unit circle, OLS is $\sqrt{T}$ consistent, just like the QD, FD, and HD. Power is also fairly similar for both the intercept only and the linear trend model.

Figure 1 plots the distribution of t-statistics for QD, QD_0, FD, HD and OLS estimators at T=200 and T=500. As one can see, the QD_0 is non-normal. However, the normal approximation to the finite sample distribution of the three estimators are good. This is quite remarkable, considering that a non-standard distribution is expected when one works with highly persistent data.

5 Predictive Regressions

The proposed estimators can be used in other problems when highly persistent data have caused problems for estimation and inference. In Gorodnichenko and Ng (2007), the QD is used to estimate dynamic stochastic general equilibrium (DSGE) models. The estimators can be also be used in predictive regressions. Suppose there is one regressor, and

$$\begin{aligned} y_t &= \beta x_{t-1} + u_{yt} \\ x_t &= \alpha_x x_{t-1} + e_{xt}, \\ u_{yt} &= \alpha_y u_{yt-1} + e_{yt}, |\alpha_y| < 1 \end{aligned} \tag{8}$$

where $e_{yt} \sim (0, \sigma_y^2), e_{xt} \sim (0, \sigma_x^2)$, $\text{cov}(e_{xt}, e_{yt}) = \sigma_{xy}, \text{cov}(e_{xt-j}, e_{yt-k}) = 0,\ \forall j, k \neq 0$. If $\alpha_x = 1$, $(1 - \beta)$ is a cointegrating vector, and least squares provide super-consistent estimates but inference is non-standard. As is well known, the finite sample distribution of $\widehat{\beta}^{OLS}$ is not well approximated by the normal distribution if x_t is highly persistent and possibly non-stationary. The challenge is how to conduct inference that is robust to the dynamic properties of the data.

Let $\theta = (\alpha_x, \alpha_y, \beta, \sigma_x^2, \sigma_y^2, \sigma_{xy})'$. Let γ_k be the model-implied autocovariance between y_t and x_{t-k}, both quasi-differenced at α_x, and let $\widehat{\gamma}_k$ is the sample analog. Our QD estimator is

$$\widehat{\theta} = \underset{\theta}{\text{argmin}}\ \bar{g}^{QD}(\theta)' W_T \bar{g}^{QD}(\theta)$$

where $\bar{g}^{QD}(\theta)$ is defined as in AR(1) case, but in addition to autocovariances for y_t and x_t, the cross-covariances are also considered. Let $Y_t = (y_t, x_t, u_{yt})'$ and $V_t = (e_{yt}, e_{xt})'$. The state-space representation of this model is:

$$\begin{bmatrix} \Delta^{\alpha_x} y_t \\ \Delta^{\alpha_x} x_t \\ \Delta^{\alpha_x} u_{yt} \\ e_{yt} \\ e_{xt} \end{bmatrix} = \begin{bmatrix} 0 & \beta & \alpha_y & -\alpha_y & 0 \\ 0 & \alpha_x & 0 & 0 & -\alpha_y \\ 0 & 0 & \alpha_y & -\alpha_y & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta^{\alpha_x} y_{t-1} \\ \Delta^{\alpha_x} x_{t-1} \\ \Delta^{\alpha_x} u_{y,t-1} \\ e_{y,t-1} \\ e_{x,t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e_{yt} \\ e_{xt} \end{bmatrix}$$

which can be simplified to

$$\begin{bmatrix} \Delta^{\alpha_x} y_t \\ \Delta^{\alpha_x} x_t \\ e_{xt} \end{bmatrix} = \begin{bmatrix} 0 & \beta & 0 \\ 0 & \alpha_x & -\alpha_y \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta^{\alpha_x} y_{t-1} \\ \Delta^{\alpha_x} x_{t-1} \\ e_{x,t-1} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} e_{yt} \\ e_{xt} \end{bmatrix} \tag{9}$$

Using (9), we can compute all the autocovariance of quasi-differenced variables as implied by the model. Matching the sample covariances of the quasi-differenced variables with those of the model yields estimates of the parameters with normal properties. Using the moments of the quasi-differenced variables in estimation can potentially be used in a broad range of applications.

6 Concluding Comments

In this paper, we suggest that moments based on quasi-differenced data can be used to derive estimators with classical properties. We also show that for the AR(1) model, it is possible to have an estimator of α that converges at a rate of $T^{3/4}$, which appears to be a new result. It is an open question whether estimators with a convergence rate faster than $\sqrt{T}$ can be asymptotically normal.

Quasi-differencing renders possibly non-stationary processes stationary so that classical limit theorems can be applied. But it is also because of this that the estimates are $\sqrt{T}$ consistent and not super-consistent in the local-to-unity framework. In exchange for this slower convergence, our estimators are asymptotically normal throughout the parameter space.

Quasi-differencing has a long tradition in econometrics and underlies GLS estimation. Canjels and Watson (1997) and Phillips and Lee (1996) found that quasi-differencing gives more precise estimates of the trend parameters when the errors are highly persistent. Pesavento and Rossi (2006) suggest that for such data, quasi-differencing can improve the coverage of impulse response functions. In both studies, the data are quasi-differenced at $\alpha = \bar{\alpha}$ which is fixed at the value as suggested by the local to unity framework. This parameter is being estimated in our QD and HD. Further work in this direction may prove to be useful.

Appendix

Proof of Proposition 1

Properties of QD First let us consider the problem of matching the $j-$th autocovariance. That is, $Q_j(\alpha) = (\widehat{\gamma}_j(\alpha) - \gamma_j)^2$ and $\widehat{\alpha} = \arg\min_\alpha Q_j(\alpha)$. Under our assumptions $\omega = \gamma_0 = \sigma^2$, and $\gamma_j = 0$ for all $j > 0$.

$$Q_j(\alpha) = \left(\left(\frac{1}{T}\sum e_t e_{t-j} - \gamma_j\right) - \frac{\alpha-\alpha_0}{T}\sum\left[e_t y_{t-j-1} + e_{t-j}y_{t-1}\right] + \right.$$
$$\left. + \frac{(\alpha-\alpha_0)^2}{T}\sum y_{t-1}y_{t-j-1}\right)^2.$$

We denote the $i-$th derivative of $Q_j(\alpha)$ by $Q_j^{(i)}$. The first order condition for the optimization is $Q_j^{(1)}(\widehat{\alpha}) = 0$. The usual GMM logic deduces the asymptotics of $\widehat{\alpha}$ from the Taylor expansion of the form

$$Q_j^{(1)}(\widehat{\alpha}) = Q_j^{(1)}(\alpha_0) + (\widehat{\alpha} - \alpha_0)Q_j^{(2)}(\alpha^*) = 0.$$

However, this logic does not work here since $\frac{Q_j^{(2)}(\alpha_0)}{Q_j^{(2)}(\alpha^*)} = 0$ for $\alpha^* \neq \alpha_0$. In fact, the next term in the Taylor expansion has the same problem. We expand the first order condition until the residual term keeps the same order of magnitude for α in the neighborhood of α_0:

$$Q_j^{(1)}(\alpha_0) + (\widehat{\alpha}-\alpha_0)Q_j^{(2)}(\alpha_0) + \frac{(\widehat{\alpha}-\alpha_0)^2}{2}Q_j^{(3)}(\alpha_0) + \frac{(\widehat{\alpha}-\alpha_0)^3}{6}Q_j^{(4)}(\alpha^*) = 0. \quad (10)$$

We use the following asymptotic statements:

$$\sqrt{T}\left(\frac{1}{T}\sum e_t e_{t-j} - \gamma_j\right) \Rightarrow \xi_j \sim N(0, \sigma_j^2), \quad (11)$$

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{[Ts]} e_t \Rightarrow \omega W(s), \quad (12)$$

$$\frac{1}{T}\sum e_{t-j}y_{t-1} \Rightarrow \gamma_0 + ... + \gamma_{j-1} + \omega^2\int_0^1 J_c(s)dW(s), \quad (13)$$

$$\frac{1}{T}\sum e_t y_{t-1-j} \Rightarrow -\gamma_1 + ... - \gamma_{j-1} + \omega^2\int_0^1 J_c(s)dW(s), \quad (14)$$

$$\frac{1}{T^2}\sum y_{t-1}y_{t-j-1} \Rightarrow \omega^2\int_0^1 J_c^2(s)ds. \quad (15)$$

Here (11) is a regular Central Limit Theorem, (12) is a Functional Central Limit Theorem, (13)-(15) are a slight modification of local-to-unity asymptotic results (see for example, Phillips (1987)). The variable ξ_j are mutually independent and independent from $W(\cdot)$, $\sigma_j^2 = \sigma^4$ is the long-run variance of the sequence $e_t e_{t-j}$, $\omega = \gamma_0 = \sigma^2$ is the long-run variance of e_t, $J_c(s) = \int_0^s e^{c(t-s)} dW(t)$ is an Ornstein-Uhlenbeck process.

From (11) - (15) it is easy to find the order of magnitude of different terms:

$$T^{1/2} Q_j^{(1)}(\alpha_0) =^a -2\sqrt{T} \left(\frac{1}{T} \sum e_t e_{t-j} - \gamma_j \right) \sum \left[e_t y_{t-j-1} + e_{t-j} y_{t-1} \right] \Rightarrow$$

$$\Rightarrow -2\xi_j \left(\gamma_0 + 2\omega^2 \int_0^1 J_c(s) dW(s) \right);$$

$$T^{-1/2} Q_j^{(2)}(\alpha_0) =^a 4\sqrt{T} \left(\frac{1}{T} \sum e_t e_{t-j} - \gamma_j \right) \frac{1}{T} \left(\frac{1}{T} \sum y_{t-1} y_{t-j-1} \right) \Rightarrow$$

$$\Rightarrow 4\xi_j \omega^2 \int_0^1 J_c^2(s) ds;$$

$$T^{-1} Q_j^{(3)}(\alpha_0) =^a -12 \frac{1}{T} \left(\frac{1}{T} \sum y_{t-1} y_{t-j-1} \right) \left(\frac{1}{T} \sum \left[e_t y_{t-j-1} + e_{t-j} y_{t-1} \right] \right) \Rightarrow$$

$$\Rightarrow -12\omega^2 \int_0^1 J_c^2(s) ds \left(\gamma_0 + 2\omega^2 \int_0^1 J_c(s) dW(s) \right)$$

$$T^{-2} Q_j^{(4)}(\alpha) = 24 \left(\frac{1}{T^2} \sum_1^T y_{t-1} y_{t-2} \right)^2 \Rightarrow 24 \left(\omega^2 \int_0^1 J_c^2(s) ds \right)^2 > 0$$

To summarize:

$$Q_j^{(1)}(\alpha_0) = O(T^{-1/2}), Q_j^{(2)}(\alpha_0) = O(T^{1/2}), Q_j^{(3)}(\alpha_0) = O(T^1), Q_j^{(4)}(\alpha) = O(T^2)$$

Assume that $T^\gamma(\widehat{\alpha} - \alpha_0) = O(1)$, and calculate the order of magnitude of different terms:

$$Q^{(1)}(\alpha_0) = O(T^{-1/2}), Q^{(2)}(\alpha_0)(\widehat{\alpha} - \alpha_0) = O(T^{1/2-\gamma}),$$

$$Q^{(3)}(\alpha_0)(\widehat{\alpha} - \alpha_0)^2 = O(T^{1-2\gamma}), Q^{(4)}(\alpha_0)(\widehat{\alpha} - \alpha_0)^3 = O(T^{2-3\gamma})$$

It allows us to find what terms are dominating in the Taylor expansion (10):

- if $\gamma < 3/4$ then $Q^{(4)}(\alpha_0)(\widehat{\alpha} - \alpha_0)^3$ is the only leading term;
- if $\gamma = 3/4$ then $Q^{(4)}(\alpha_0)(\widehat{\alpha} - \alpha_0)^3$ and $Q^{(2)}(\alpha_0)(\widehat{\alpha} - \alpha_0)$ are leading and of the same order;
- if $3/4 < \gamma < 1$ then $Q^{(2)}(\alpha_0)(\widehat{\alpha} - \alpha_0)$ is the only leading term;
- if $\gamma = 1$ then $Q^{(2)}(\alpha_0)(\widehat{\alpha} - \alpha_0)$ and $Q^{(1)}(\alpha_0)$ are leading and of the same order;
- if $\gamma > 1$ then $Q^{(1)}(\alpha_0)$ is the only leading term.

Given that we have to solve equation (10), the only suspects are $\gamma = 3/4$ and $\gamma = 1$.

If $\gamma = 3/4$ then asymptotically we have the following equation:

$$(\widehat{\alpha} - \alpha_0)Q_j^{(2)}(\alpha_0) + \frac{(\widehat{\alpha} - \alpha_0)^3}{6} Q_j^{(4)}(\alpha^*) = 0.$$

or

$$T^{3/2}(\widehat{\alpha} - \alpha_0)^2 = -T^{3/2} \frac{Q_j^{(2)}(\alpha_0)}{Q_j^{(4)}(\alpha^*)} \Rightarrow \frac{-\xi_j}{\omega^2 \int_0^1 J_c^2(s)ds}$$

the equation has solution only for $\xi_j < 0$.

If $\gamma = 1$ then asymptotically we have the following equation:

$$Q_j^{(1)}(\alpha_0) + (\widehat{\alpha} - \alpha_0)Q_j^{(2)}(\alpha_0) = 0.$$

or

$$T(\widehat{\alpha} - \alpha_0) = T \frac{Q_j^{(1)}(\alpha_0)}{Q_j^{(2)}(\alpha_0)} \Rightarrow \frac{\gamma_0 + 2\omega^2 \int_0^1 J_c(s)dW(s)}{2\omega^2 \int_0^1 J_c^2(s)ds}$$

But if $\xi_j < 0$ then at the above mentioned solution the second derivative $Q_j^{(2)}(\widehat{\alpha})$ will be of the wrong sign (the solution is a local maximum of Q rather than a local minimum). That is, we proved that (3) holds.

Now, let us consider a possibility that several covariances are matched, that is:

$$Q(\alpha) = \sum_{j=1}^{J} Q_j(\alpha), \quad \widehat{\alpha}^{QD_0} = \arg\min_{\alpha} Q(\alpha).$$

All statements analogous to those above stay valid and lead to (3) with ξ_j being replaced with $\sum_{j=1}^{J} \xi_j/J \sim N(0, \sigma^4/J)$. So, one can see that in the AR(1) case matching more than one autocovariance leads to increase in efficiency.

Properties of QD The QD estimator is $\widehat{\alpha} = \arg\min_\alpha g_j(\alpha)^2$,

$$g_j(\alpha) = \Big[\widehat{\gamma}_j(\alpha) - \widehat{\gamma}_0(\alpha)\Big] - \Big[\gamma_j(\alpha) - \gamma_0(\alpha)\Big].$$

Again in our case $\gamma_j(\alpha) = 0$ for all $j > 0$, and $\gamma_0 = \sigma^2$.

$$\begin{aligned} \widehat{\gamma}_j(\alpha) &= \frac{1}{T}\sum_{t=1}^{T} e_t e_{t-j} + (\alpha_0 - \alpha)\frac{1}{T}\sum_{t=1}^{T}\Big(y_{t-1}e_{t-j} + y_{t-j-1}e_t\Big) + (\alpha_0 - \alpha)^2 \frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_{t-j-1} \\ \widehat{\gamma}_0(\alpha) &= \frac{1}{T}\sum_{t=1}^{T} e_t^2 + 2(\alpha_0 - \alpha)\frac{1}{T}\sum_{t=1}^{T} y_{t-1}e_t + (\alpha_0 - \alpha)^2 \frac{1}{T}\sum_{t=1}^{T} y_{t-1}^2. \end{aligned}$$

$$\frac{1}{T}\sum[y_{t-1}y_{t-j-1} - y_{t-1}^2] = \frac{1}{T}\sum y_{t-1}y_{t-j-1} - \frac{1}{T}\sum y_{t-1}(\alpha_0^j y_{t-j-1} + \sum_{k=0}^{j-1}\alpha_0^k e_{t-k-1})$$
$$= (1-\alpha_0^j)\frac{1}{T}\sum_t y_{t-1}y_{t-j-1} - \frac{1}{T}\sum_{k=0}^{j-1}\alpha_0^k\left(\sum_t y_{t-1}e_{t-k-1}\right).$$

If $\alpha_0 = 1 + c/T$, and j is fixed, then (13) and (15) imply

$$\frac{1}{T}\sum[y_{t-1}y_{t-j-1} - y_{t-1}^2] \Rightarrow j\left(-c\omega^2\int_0^1 J_c^2(s)ds - \gamma_0 - \omega^2\int_0^1 J_c(s)dW(s)\right) = O_p(1).$$

It implies that $\widehat{\alpha}$ is $\sqrt{T}$ consistent, and

$$\sqrt{T}(\widehat{\gamma}_j(\widehat{\alpha}) - \widehat{\gamma}_0(\widehat{\alpha})) = \xi_j - \xi_0 + \sqrt{T}(\widehat{\alpha}-\alpha_0)\left(\gamma_0 + 2\omega^2\int J_c(r)dW(r) - 2\omega^2\int J_c(r)dW(r)\right) + o_p(1).$$

That is,

$$\sqrt{T}g(\widehat{\alpha}) = \xi_j - \xi_0 + \gamma_0\sqrt{T}(\widehat{\alpha}-\alpha_0) + o_p(1).$$

This gives $\sqrt{T}(\widehat{\alpha}-\alpha_0) \sim N(0,2)$.

If we use the classical assumption that $|\alpha_0| < 1$ is fixed, then again the second order term

$$T^{-1}\sum[y_{t-1}y_{t-j-1} - y_{t-1}^2] = O_p(1)$$

The term on $(\alpha_0 - \alpha)$ is

$$\frac{1}{T}\sum_{t=1}^{T}\left(y_{t-1}e_{t-j} + y_{t-j-1}e_t\right) \to^p \alpha_0^{j-1}\sigma^2$$

The estimate $\widehat{\alpha}$ is $\sqrt{T}$ consistent,

$$\sqrt{T}(\widehat{\gamma}_j(\widehat{\alpha}) - \widehat{\gamma}_0(\widehat{\alpha})) = \xi_j - \xi_0 + \sqrt{T}(\widehat{\alpha}-\alpha_0)\alpha_0^{j-1}\sigma^2 + o_p(1),$$

and $\sqrt{T}(\widehat{\alpha}-\alpha_0) \Rightarrow N(0, 2/\alpha_0^{2(j-1)})$.

Now assume that we match a fixed number of autocovariances:

$$\widehat{\alpha}_J^{QD} = \arg\min_\alpha \sum_{j=1}^{J} g_j^2(\alpha).$$

We assume that J stays fixed. The first order condition is $\sum_{j=1}^J g_j(\widehat{\alpha})\frac{\partial g_j}{\partial\alpha}(\widehat{\alpha}) = 0$. In local to unity asymptotic framework we have $\frac{\partial g_j}{\partial \alpha}(\widehat{\alpha}) = 2\gamma_0 + o_p(1)$ and as before $\sqrt{T}g(\widehat{\alpha}) = \xi_j - \xi_0 + \gamma_0\sqrt{T}(\widehat{\alpha} - \alpha_0) + o_p(1)$. That is, our problem is asymptotically equivalent to

$$\sum_{j=1}^{J}(\xi_j - \xi_0) + \sqrt{T}(\widehat{\alpha}-\alpha_0)J\gamma_0 + o_p(1) = 0.$$

That is, $\sqrt{T}(\widehat{\alpha}-\alpha_0) \Rightarrow N(0, 1+1/J)$.

In stationary asymptotics: $\frac{\partial g_j}{\partial \alpha}(\widehat{\alpha}) = 2\sigma^2\alpha^{j-1} + o_p(1)$ and as before $\sqrt{T}g(\widehat{\alpha}) = \xi_j - \xi_0 + \alpha_0^{j-1}\sigma^2\sqrt{T}(\widehat{\alpha}-\alpha_0) + o_p(1)$. As a result:

$$\sqrt{T}(\widehat{\alpha}-\alpha_0) = -\frac{\sum_{j=1}^{J}(\xi_j-\xi_0)\alpha_0^{j-1}}{\sigma^2\sum_{j=1}^{J}\alpha_0^{2(j-1)}} \Rightarrow N(0,\sigma_J^2)$$

Where

$$\sigma_J^2 = \frac{\left(\sum_{j=1}^{J}\alpha_0^{(j-1)}\right)^2 + \sum_{j=1}^{J}\alpha_0^{2(j-1)}}{\left(\sum_{j=1}^{J}\alpha_0^{2(j-1)}\right)^2}$$

The FD Estimator

Consider the FD estimator:

$$\widehat{\alpha}^{FD} - \alpha_0 = \frac{\sum_{t=1}^{T} e_t \Delta y_{t-1}}{\sum_{t=1}^{T} y_{t-1}\Delta y_{t-1}}.$$

Since $\Delta y_{-1} = e_{t-1}$ when $\alpha_0 = 1$, the numerator satisfies $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_t e_{t-1} \xrightarrow{d} N(0,\sigma^4)$. The denominator is $y_{t-1}e_{t-1} = (y_{t-2}+e_{t-1})e_{t-1}$. Now $T^{-1}\sum_{t=1}^{T} y_{t-2}e_{t-1} \Rightarrow \frac{\sigma^2}{2}(J_c(1)^2-1)$ and $T^{-1}\sum_{t=1}^{T} e_{t-1}^2 \xrightarrow{p} \sigma^2$. Thus, the denominator converges to $\frac{\sigma^2}{2}(J_c(1)^2+1)$.

To show that the t statistic is asymptotically normal even though $\widehat{\alpha}$ is only conditionally normal, we need to show that the numerator and the denominator of the t statistic are independent. For this, we need to consider the joint distribution of

$$\left[T^{-1/2}\sum_{t=1}^{T} e_{t-1}e_t \quad T^{-1}\sum_{t=1}^{T} y_{t-1}e_{t-1}\right].$$

Their covariance is

$$T^{-3/2}E\Bigg[\sum_{s=1}^{T} e_{s-1}e_s \sum_{t=1}^{T} y_{t-1}e_{t-1}\Bigg].$$

Since $y_{t-1} = \sum_{j=1}^{t} e_{j-1}$, this covariance is non-zero only if $s = t$. In this case,

$$T^{-3/2}\sum_{s=1}^{T} E(e_s^2 e_{s-1}^2) = T^{-3/2}\sum_{s=1}^{T}\sigma^4 = \sigma^4 T^{-1/2} \xrightarrow{p} 0.$$

The two terms are asymptotically uncorrelated. By the Brownian motion property of the denominator and asymptotic property of the numerator, the two terms are also asymptotically independent. By the continuous mapping theorem, the ratio has a limit $2(1+J_c(1)^2)^{-1}N(0,1)$.

When $|\alpha_0| < 1$, the denominator is

$$\begin{aligned} T^{-1}\sum_{t-1}^{T} y_{t-1}\Delta y_{t-1} &= T^{-1}\sum_{t=1}^{T} y_{t-1}^2 - y_{t-1}y_{t-2} = \widehat{\gamma}_y(0) - \widehat{\gamma}_y(1) \\ &\xrightarrow{p} \gamma_y(0) - \gamma_y(1) = \frac{\sigma^2(1-\alpha_0)}{1-\alpha_0^2} = \frac{\sigma^2}{1+\alpha_0}. \end{aligned}$$

Now consider the numerator: $T^{-1}\sum_{t=1}^{T} e_t\Delta y_{t-1} \xrightarrow{p} E(e_t\Delta y_{t-1}) = 0$ by the law of iterated projection and $\text{var}(\Delta y_t) = 2\gamma_y(0) - 2\gamma_y(1) = 2\frac{\sigma^2}{(1+\alpha_0)}$. It follows that $\text{var}(e_t\Delta y_{t-1}) = 2\sigma^4/(1+\alpha_0)$. Combining the results, $\sqrt{T}(\widehat{\alpha}^{FD} - \alpha_0) \xrightarrow{d} \frac{(1+\alpha_0)}{\sigma^2} N(0, \frac{2\sigma^4}{(1+\alpha_0)}) = N(0, 2(1+\alpha_0))$. Arguments similar to FD apply to HD estimator. Details are omitted.

References

Canjels, E. and Watson, M. W. 1997, Estimating Deterministic Trends in the Presence of Serially Correlated Errors, *Review of Economics and Statistics* **May**, 184–200.

Choi, I. 1993, Asymptotic Normality of the Least Squares Estimates for Higher Order Autoregressive Integrated Process with some Applications, *Econometric Theory* **9:3**, 263–282.

Gorodnichenko, Y. and Ng, S. 2007, Estimation of DSGE Models when the Data are Persistent, mimeo,Columbia University.

Han, C., Phillips, P. and Sul, D. 2009a, Uniform Asymptotic Normality in Stationary and Unit Root Autoregression, unpublished manuscript.

Han, C., Phillips, P. and Sul, D. 2009b, X-Differencinng and Fully Aggregated Estimation of Panel Models, Working Paper, Yale University.

Laroque, B. and Salanie, B. 1997, Normal Estimators for Cointegrating Relationships, *Economics Letters* **55**, 185–189.

Pesavento, E. and Rossi, B. 2006, Small Sample Confidence Interevals for Multivariate Impulse Response Functions at Long lags, *Journal of Applied Econometrics* **21:8**, 1135–1155.

Phillips, P. and Lee, C. 1996, Efficiency Gains from Quasi-differencing Under Non-stationarity, *in* P. Robinson and M. Rosenblatt (eds), *Athens Conference on Applied Probability and Time Series: Volume II Time Series Analysis in Honor of E.J. Hannan.*

Phillips, P. C. B. 1987, Time Series Regression with Unit Roots, *Econometrica* **55**, 277–302.

Phillips, P. C. B. and Han, C. 2008, Gaussian Inference in AR(1) Time Series With or Without a Unit Root, *Econometric Theory* **63**, 1023–1078.

So, B. and Shin, D. 1999, Cauchy Estimators for Autoregressive Procseses with Applications to Unit Root Tests and Confidence Intervals, *Econometric Theory* **15**, 166–176.

Table 1: Mean Estimates of ρ and Rejection Rates of J Test.

ρ_0	M	OLS	QD	QD_0	HD	FD	QD	QD_0	HD	FD
DGP 1(a)		Mean estimates					J test			
Intercept model: T=200										
1.00	2	0.973	1.002	1.004	0.975	0.961	0.071	0.090	0.062	0.065
0.98	2	0.956	0.983	0.998	0.962	0.958	0.072	0.080	0.062	0.065
0.95	2	0.927	0.954	0.962	0.933	0.932	0.072	0.063	0.061	0.065
0.92	2	0.898	0.925	0.930	0.904	0.903	0.072	0.070	0.059	0.065
0.90	2	0.878	0.906	0.909	0.884	0.883	0.073	0.071	0.060	0.066
0.85	2	0.827	0.858	0.855	0.834	0.834	0.075	0.082	0.059	0.067
0.50	2	0.474	0.507	0.493	0.486	0.488	0.084	0.109	0.062	0.070
-0.50	2	-0.539	-0.511	-0.499	-0.495	-0.491	0.114	0.107	0.068	0.028
Intercept model: T=500										
1.00	2	0.989	1.004	0.998	0.990	0.974	0.054	0.079	0.056	0.055
0.98	2	0.971	0.985	0.996	0.974	0.974	0.059	0.055	0.057	0.056
0.95	2	0.941	0.955	0.960	0.944	0.945	0.061	0.046	0.055	0.055
0.92	2	0.911	0.925	0.929	0.914	0.915	0.061	0.055	0.054	0.055
0.90	2	0.891	0.906	0.907	0.894	0.895	0.060	0.058	0.055	0.055
0.85	2	0.841	0.856	0.851	0.844	0.845	0.060	0.068	0.054	0.054
0.50	2	0.489	0.504	0.497	0.495	0.496	0.072	0.074	0.056	0.062
-0.50	2	-0.513	-0.501	-0.498	-0.497	-0.495	0.086	0.075	0.059	0.031
Linear trend model: T=200										
1.00	2	0.947	1.002	1.006	0.963	0.968	0.072	0.083	0.063	0.065
0.98	2	0.938	0.982	0.996	0.953	0.958	0.074	0.075	0.064	0.065
0.95	2	0.912	0.954	0.962	0.926	0.932	0.074	0.057	0.063	0.066
0.92	2	0.884	0.925	0.928	0.897	0.903	0.073	0.065	0.061	0.065
0.90	2	0.864	0.906	0.905	0.877	0.883	0.073	0.072	0.060	0.066
0.85	2	0.814	0.858	0.851	0.827	0.833	0.076	0.085	0.061	0.066
0.50	2	0.458	0.507	0.487	0.479	0.487	0.086	0.114	0.061	0.070
-0.50	2	-0.563	-0.511	-0.501	-0.497	-0.491	0.113	0.118	0.073	0.028
Linear trend model: T=500										
1.00	2	0.979	1.005	0.999	0.986	0.988	0.059	0.071	0.056	0.057
0.98	2	0.965	0.985	0.996	0.971	0.974	0.061	0.052	0.056	0.056
0.95	2	0.936	0.955	0.959	0.942	0.944	0.061	0.043	0.055	0.055
0.92	2	0.906	0.925	0.927	0.912	0.914	0.061	0.053	0.054	0.055
0.90	2	0.886	0.906	0.905	0.892	0.894	0.060	0.059	0.055	0.056
0.85	2	0.836	0.856	0.849	0.842	0.844	0.060	0.070	0.055	0.054
0.50	2	0.483	0.504	0.495	0.492	0.496	0.072	0.078	0.057	0.063
-0.50	2	-0.524	-0.501	-0.499	-0.498	-0.495	0.086	0.078	0.064	0.031

Table 2: Finite Sample Power: $M = 2$.

	$t_1 : H_0 : \rho = \rho_0$ $H_1 : \rho < \rho_0$					$t_2 : H_0 : \rho_0 - .05$ $H_1 : \rho > \rho_0 - .05$					$t_3 : H_0 : \rho = \rho_0 - .10$ $H_1 : \rho > \rho_0 - .1$				
ρ_0	OLS	QD	QD_0	HD	FD	OLS	QD	QD_0	HD	FD	OLS	QD	QD_0	HD	FD
	T=200														
1.00	0.454	0.061	0.113	0.124	0.029	0.539	0.168	0.392	0.158	0.067	0.912	0.319	0.641	0.568	0.251
0.98	0.225	0.063	0.085	0.109	0.032	0.466	0.194	0.376	0.215	0.094	0.883	0.333	0.597	0.614	0.289
0.95	0.147	0.069	0.134	0.107	0.042	0.374	0.197	0.246	0.229	0.124	0.805	0.329	0.538	0.617	0.309
0.92	0.117	0.072	0.132	0.101	0.047	0.308	0.198	0.198	0.227	0.130	0.725	0.327	0.529	0.606	0.304
0.90	0.110	0.074	0.128	0.100	0.050	0.281	0.198	0.177	0.223	0.132	0.674	0.322	0.531	0.600	0.305
0.85	0.097	0.082	0.115	0.097	0.051	0.222	0.192	0.160	0.210	0.134	0.563	0.313	0.547	0.574	0.297
0.50	0.076	0.099	0.091	0.084	0.046	0.106	0.165	0.204	0.172	0.125	0.252	0.287	0.480	0.432	0.265
-0.50	0.070	0.114	0.065	0.052	0.074	0.057	0.203	0.231	0.223	0.204	0.095	0.460	0.515	0.522	0.407
	Intercept model: T=500														
1.00	0.453	0.041	0.171	0.073	0.014	0.960	0.240	0.554	0.474	0.174	1.000	0.488	0.868	0.933	0.493
0.98	0.153	0.050	0.065	0.076	0.028	0.913	0.248	0.497	0.518	0.232	0.999	0.478	0.954	0.956	0.579
0.95	0.114	0.052	0.106	0.073	0.040	0.777	0.248	0.338	0.504	0.233	0.994	0.461	0.967	0.953	0.571
0.92	0.095	0.053	0.093	0.074	0.041	0.675	0.243	0.351	0.480	0.233	0.984	0.447	0.949	0.940	0.564
0.90	0.094	0.056	0.088	0.074	0.043	0.617	0.243	0.376	0.470	0.228	0.969	0.442	0.937	0.935	0.563
0.85	0.094	0.061	0.079	0.072	0.043	0.492	0.237	0.379	0.450	0.222	0.906	0.438	0.910	0.912	0.561
0.50	0.067	0.076	0.062	0.067	0.048	0.203	0.211	0.341	0.330	0.197	0.496	0.452	0.789	0.781	0.503
-0.50	0.062	0.074	0.052	0.047	0.063	0.100	0.335	0.366	0.389	0.296	0.187	0.774	0.829	0.856	0.677
	Linear trend model: T=200														
1.00	0.768	0.068	0.108	0.226	0.055	0.146	0.188	0.361	0.106	0.061	0.664	0.332	0.507	0.461	0.226
0.98	0.429	0.070	0.111	0.157	0.047	0.216	0.197	0.352	0.169	0.101	0.710	0.334	0.502	0.538	0.291
0.95	0.278	0.069	0.150	0.144	0.052	0.207	0.197	0.255	0.192	0.121	0.658	0.326	0.468	0.552	0.313
0.92	0.206	0.073	0.150	0.136	0.056	0.180	0.195	0.200	0.189	0.129	0.589	0.325	0.447	0.543	0.309
0.90	0.184	0.076	0.147	0.132	0.053	0.165	0.195	0.169	0.186	0.128	0.552	0.322	0.449	0.532	0.304
0.85	0.159	0.082	0.135	0.128	0.056	0.138	0.192	0.139	0.177	0.132	0.455	0.311	0.477	0.514	0.292
0.50	0.102	0.099	0.106	0.103	0.048	0.076	0.162	0.184	0.146	0.125	0.196	0.287	0.437	0.392	0.268
-0.50	0.082	0.113	0.069	0.056	0.075	0.048	0.203	0.220	0.214	0.203	0.078	0.459	0.499	0.506	0.406
	Linear trend model: T=500														
1.00	0.775	0.046	0.143	0.128	0.034	0.826	0.249	0.420	0.397	0.172	0.997	0.489	0.817	0.916	0.523
0.98	0.275	0.049	0.078	0.096	0.033	0.822	0.251	0.456	0.470	0.236	0.996	0.478	0.934	0.943	0.570
0.95	0.182	0.053	0.125	0.089	0.043	0.691	0.248	0.303	0.458	0.232	0.990	0.461	0.947	0.937	0.571
0.92	0.144	0.054	0.105	0.089	0.043	0.592	0.243	0.303	0.445	0.234	0.968	0.446	0.929	0.925	0.568
0.90	0.134	0.056	0.099	0.088	0.045	0.538	0.244	0.330	0.436	0.229	0.948	0.443	0.911	0.919	0.563
0.85	0.124	0.061	0.088	0.090	0.043	0.416	0.236	0.342	0.417	0.221	0.876	0.437	0.886	0.899	0.559
0.50	0.080	0.076	0.070	0.074	0.048	0.180	0.211	0.321	0.304	0.198	0.451	0.451	0.766	0.758	0.505
-0.50	0.073	0.074	0.056	0.050	0.063	0.089	0.335	0.357	0.376	0.296	0.164	0.774	0.826	0.849	0.676

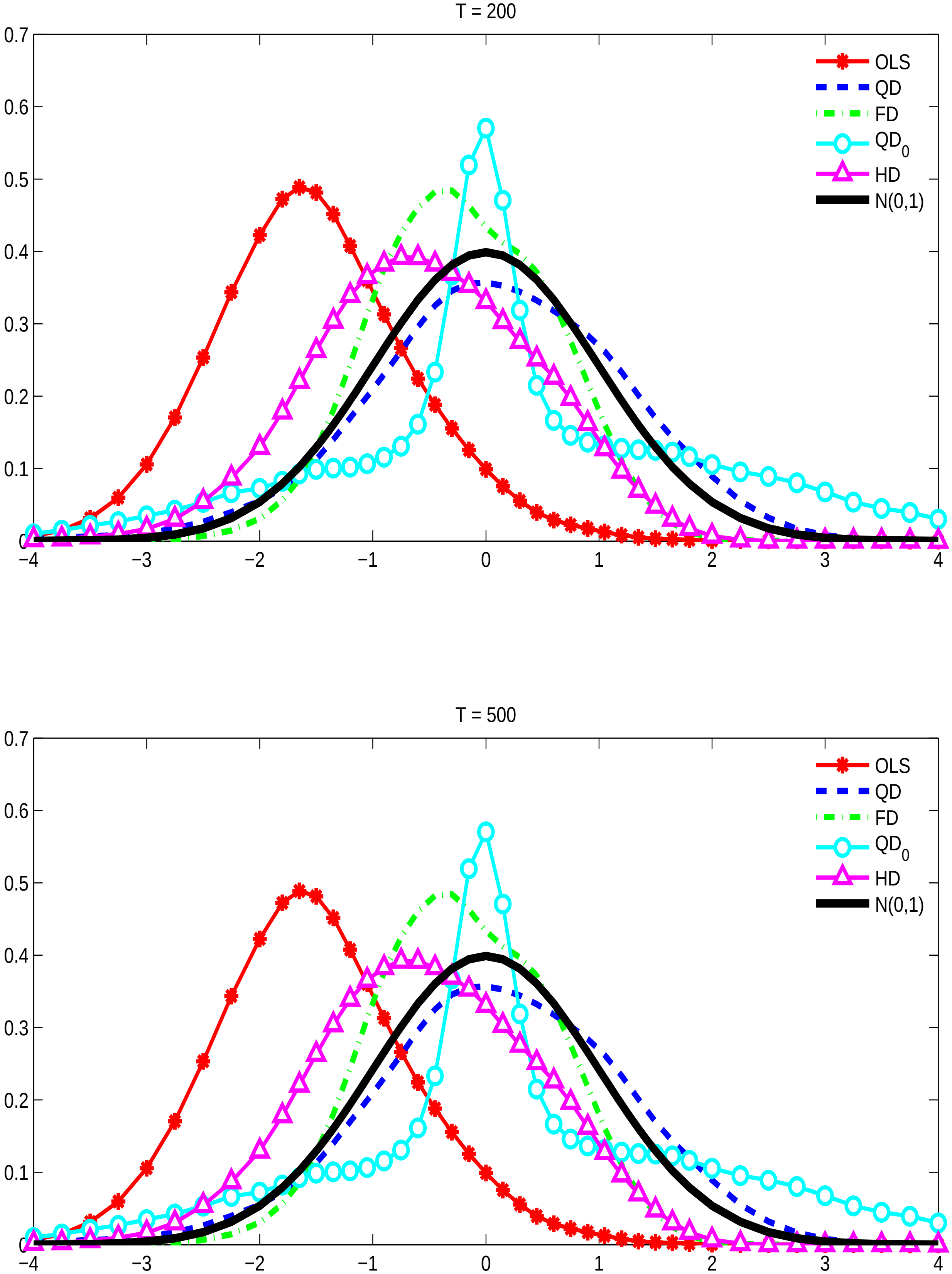

Figure 1: Distribution of the t-statistic for the largest autoregressive root in the intercept-only model with $K = 1$,$M = 2$, $\alpha_0 = 1$.