A Simple Test for Nonstationarity in Mixed Panels

Serena Ng
Department of Economics, University of Michigan, Ann Arbor, MI 48109 (*Serena.Ng@umich.edu*)

This article proposes a simple estimator that is consistent for the fraction of a panel that has an autoregressive unit root. Given such an estimate, $\hat{\theta}$, we can test the null hypothesis that $\theta = \theta^0$ for any value of $\theta^0 \in (0, 1]$. The test is asymptotically standard normal and is valid whether or not the panel is cross-sectionally correlated. The main insight is that in a panel in which some units are stationary and some have unit roots, the cross-sectional variance of the mixed panel is dominated by a linear trend that grows at rate θ, where θ is precisely the fraction of the panel with a unit root. Averaging the change in cross-sectional variance over time then gives a $\sqrt{N}$ consistent estimate of θ as $N, T \to \infty$. Simulations show that the estimator has good finite-sample properties when $T \geq 100$, even with N as small as 30.

1. INTRODUCTION

This article considers the case where the units of a panel have heterogeneous dynamics and some (and possibly all) of the units are cross-sectionally correlated. We provide a simple procedure to consistently estimate and test what fraction of the units in the panel has an autoregressive unit root. We denote this fraction by θ. We can test whether $\theta = \theta^0$ for any $\theta^0 \in (0, 1]$. The test is asymptotically standard normal. The critical values are the same whether or not there is cross-sectional dependence, provided that the cross-sectional variation is stationary.

Panel nonstationarity tests often maintain as the null hypothesis that every series in the panel has an autoregressive unit root and test it against the alternative that at least one series is stationary. Rejections of the null hypothesis leave the researcher with no information as to whether the rejection is due to just a few or to most of the series not agreeing with the null hypothesis. Chang and Song (2002) tested the null hypothesis that at least one unit is nonstationary against the alternative that all units are stationary. Nonetheless, the hypothesis can be rejected for many configurations of the dynamic parameters. Panel stationarity tests are subject to the same criticism.

Our proposed test is different in two ways. First, we are not confined to testing the two extreme hypotheses that all units are nonstationary or stationary. Second, we quantify the fraction of the units that are nonstationary, making it possible to be more precise about the extent of heterogeneity in the dynamics of the units in the panel. The basic idea behind the estimator is that if all series in the panel are nonstationary, then the cross-sectional variance has a linear trend. The linear trend in variance has been described as the "fanny-out" phenomenon by Deaton and Paxson (1994), who pointed out that if the permanent income hypothesis holds and consumption follows a random walk, then consumption inequality must increase as agents of the same cohort move through the life cycle. Lucas (2003) noted that the fanning out over time of the earnings and consumption distributions within a cohort found by Deaton and Paxson is "striking evidence of a sizeable, uninsurable random walk component in earnings." On the other hand, if all series are stationary, then the cross-section variance will be nontrending. In a panel in which some series are nonstationary and some are stationary, the cross-sectional variance of the mixed panel will have a linear trend that increases at rate $\theta < 1$, where θ is precisely the fraction of the panel that is nonstationary.

Our analysis can be applied in various contexts. First, nonstationarity of every series in a system is a precondition for the presence of a common stochastic trend. It is convenient to use $\theta = 1$ to test whether the common trend representation is valid, especially when the number of variables in the system is large. Second, many tests have been developed to ascertain whether all units in a panel of data are nonstationary (or all stationary). Recent surveys of panel unit root tests have been given in Maddala and Wu (1999) and Baltagi and Kao (2001). The early test developed by Quah (1994) imposed substantial homogeneity in the cross-sectional dimension. Subsequent tests, such as those of Levin, Lin, and Chu (2002), Hadri (2000), and Im, Pesaran, and Shin (2003), allow for heterogeneous intercepts and slopes while maintaining the assumption of independence across units. These tests were initially motivated by their potential for power gains over univariate tests. However, it is now understood that unless the units are cross-sectionally uncorrelated, some power gains are in fact the consequence of size distortions. O'Connell (1998) showed that cross-sectional correlation leads the standard pooled test to overreject the null hypothesis. Tests that control for cross-sectional dependence often assume a factor structure designed to model strong form dependence. There remains no satisfactory way to handle the case when the units are neither independent nor strongly cross-sectionally dependent and N is large.

Third, explicit modeling of the behavior of the micro units is increasingly recognized as being necessary to better understand aggregate behavior. In the case of consumption, the typical heterogeneous agent model assumes that household units differ according to the ex post realizations of income shocks, but that their earnings evolve according to the same law of motion. Following the seminal work of MaCurdy (1981), the earnings process is often specified as the sum of a permanent and a transitory component, implicitly assuming that earnings has a unit root (see Carroll and Samwick 1997, among many others). However, when the parameters governing the dynamics are freely estimated, the evidence for a unit root is not altogether convincing. For Panel Study of Income Dynamics (PSID) data, Baker (1997), Alvarez, Browning, and Ejrnaes (2001), and Guvenen (2005) reported estimates of the autoregressive parameter

Journal of Business & Economic Statistics
January 2008, Vol. 26, No. 1
DOI 10.1198/073500106000000675

to be between .5 and .85. In contrast, Storesletten, Telmer, and Yaron (2004) obtained an estimate of .98 and subsequently imposed it to be 1. This lack of consensus concerning the magnitude of the autoregressive parameter is problematic, because the dynamics of earnings is crucial to explaining any excess sensitivity and excess smoothness observed in consumption data. For example, a consumer with a stationary earnings process should have a much smaller consumption response to an earnings innovation than a consumer whose earnings have a unit root, because for the latter consumer, all earnings changes are permanent (see Campbell and Deaton 1989; Ludvigson and Michaelides 2001). Recently, Reis (2005) also pointed out that the costs of fluctuations are much smaller when income is stationary than when it has a unit root. But to better understand the dynamics of the micro units, the assumption of homogeneous dynamics must be questioned. As forcefully argued by Alvarez et al. (2001), there is no economic reason why earnings at the individual level should have the same dynamic structure. For example, Choi (2004) found that the panel of data that he examined is a mixture of unit root and stationary processes. Our analysis sheds light on this issue, because unless θ is estimated to be 0 or 1, imposing homogeneous earnings dynamics would be inappropriate.

Fourth, the seminal work of Hall (1978) predicts that under rational expectations and perfect capital markets, the marginal utility of consumption should be a martingale. Many explanations (including liquidity constraints, myopic behavior, and nonseparability between consumption and leisure) have been used to rationalize rejections of the martingale hypothesis by the data. But these studies often test the null hypothesis against one specific alternative. Because the permanent income hypothesis can be rejected for more than one reason, it instead might be useful to take as a starting point the fact that all departures from the permanent income hypothesis imply that changes in marginal utility are predictable. Finding the fraction of the sample for which the permanent income hypothesis holds amounts to determining the fraction in the sample for which the marginal utility of consumption has a unit root. Our analysis provides a simple way to estimate this fraction.

Our use of panel data information is quite different from other panel nonstationarity tests, because we aggregate rather than pool cross-sectional information. Various authors have studied the dynamics of a process constructed from aggregating heterogeneous micro units. Granger and Morris (1976) showed that the sum of N independent AR(1) processes,

$$y_{it} = \alpha_i y_{it-1} + e_{it}, \tag{1}$$

is an ARMA$(N, N-1)$. Granger (1980) assumed a beta density for α_i and showed the parameter restrictions under which the aggregate process will be stationary with long memory. Zaffaroni (2004) assumed that α_i has a semiparametric density $c(1-\alpha)^b$ with support on $[0, 1]$ and showed that the aggregate process is stationary if $b > -1/2$. Lewbel (1994) showed that if $\alpha_i \in [0, 1]$ and the innovation variance is the same across units, then the coefficients of the autoregression in the aggregate data can be related to the moments of α_i. In particular, the first two coefficients of the autoregression in the aggregate data are the mean and the variance of α_i. The aggregate data need not be stationary, because in Lewbel's analysis, $\alpha_i \in [0, 1]$, so that some units can be stationary while others are not. Like Lewbel (1994), we are also interested in this case of a mixed panel, but our approach is to use the cross-sectional variance to quantify the extent of dynamic heterogeneity, defined as the fraction of units in the sample with $\alpha_i = 1$. Furthermore, we also allow the innovation variance to differ across units.

As a matter of terminology, a nonstationary series in this article is taken to mean that the series has an autoregressive unit root. An "I(1) unit" is differenced stationary and thus nonstationary. In contrast, an "I(0) unit" is stationary in the sense of having autoregressive roots strictly less than unity. The objective is to estimate what fraction of the units in a panel is I(1).

2. MOTIVATION

To fix ideas, we first consider the following AR(1) model, where for $i = 1, \dots, N$ and $t = 1, \dots, T$:

$$y_{it} = \lambda_i + u_{it}$$

and

$$u_{it} = \alpha_i u_{it-1} + e_{it},$$

which we can also write as

$$y_{it} = \lambda_i + \alpha_i^t u_{i0} + \sum_{j=0}^{t-1} \alpha_i^j e_{it-j}. \tag{2}$$

In this model y_{it} is observed, but λ_i and u_{it} are not observed. We assume the following:

A1. $\alpha_i \in [0, 1]$ and is independent of e_{jt} for i, j, and all t.
A2. $0 \le \lambda_i < \infty$ for all i, and λ_i is independent of e_{it} for all i and for every t. Furthermore, $0 \le \text{var}_{i,N}(\lambda_i) < \infty$, where $\text{var}_{i,N}(\lambda_i) = \frac{1}{N}\sum_{i=1}^N (\lambda_i^2 - \Lambda_N)^2$ and $\Lambda_N = \frac{1}{N}\sum_{i=1}^N \lambda_i$.
A3. For each i, e_{it} is iid over i and t with $u_{i0} = O_p(1)$.

Assumptions A1 and A2 are used throughout the analysis. The parameter λ_i is an individual-specific intercept that is independent of α_i and need not be random, although if it is, then we have a random-effects model. The assumption that e_{it} is iid with mean 0, has unit variance, and is serially uncorrelated is made for simplicity. We allow for heteroscedastic errors, higher-order dynamics, and cross-sectional correlation later. Restricting attention to $\alpha_i \in [0, 1]$ is without loss of generality, and for economic problems, this is the parameter space of interest.

Define $m_{j,\infty} = E_i(\alpha_i^j)$ to be the jth raw moment of α_i with $m_{0,\infty} = 1$, where E_i is taken to mean plim $\frac{1}{N}\sum_{i=1}^N$. In this notation, $E_i(\alpha_i) = m_{1,\infty}$ and $\text{var}_{i,\infty}(\alpha_i) = m_{2,\infty} - m_{1,\infty}^2$ are the cross-sectional mean and variance of α_i. Because $\alpha_i \in [0, 1]$, all moments of α_i are finite. Note also that $m_{j,\infty} \ge m_{r,\infty}$ when $r > j$ except when $\alpha_i = 1$ for every i, in which case $m_{j,\infty} = 1\ \forall j$. Let

$$Y_{t,N} = \frac{1}{N}\sum_{i=1}^N y_{it} \qquad \text{and} \qquad U_{t,N} = \frac{1}{N}\sum_{i=1}^N u_{it}$$

be the sample cross-sectional means of y_{it} and u_{it} for a given t. The corresponding cross-sectional variances are defined as

$$V_{t,N} = \frac{1}{N}\sum_{i=1}^{N}(y_{it} - Y_{t,N})^2 \qquad \text{and}$$

$$s_{t,N} = \frac{1}{N}\sum_{i=1}^{N}(u_{it} - U_{t,N})^2.$$

Then the population mean and variance of y_{it} at time t are

$$Y_{t,\infty} = E_i(y_{it}) \qquad \text{and} \qquad V_{t,\infty} = \text{var}_i(y_{it}).$$

Likewise, define $\epsilon_{t,N} = \frac{1}{N}\sum_{i=1}^{N} e_{it}$ and $\epsilon_{t,\infty} = E_i(e_{it})$. Hereinafter, an aggregate variable with superscript of "1" is used to denote quantities belonging to the I(1) units, whereas a superscript "0" denotes quantities corresponding to the I(0) units. In this notation, Y^1_{t,N_1} and V^1_{t,N_1} are the cross-sectional mean and variance of a sample of I(1) observations of size N_1, whereas $m^0_{t,\infty}$ is the tth raw moment of α_i of the stationary subpopulation.

Aggregating over i and for large N, (2) gives

$$Y_{t,\infty} = \Lambda_\infty + m_{t,\infty}U_{0,\infty} + \sum_{k=0}^{t-1} m_{k,\infty}\epsilon_{t-k,\infty} \qquad (3)$$

and

$$V_{t,\infty} = \text{var}_{i,\infty}(\lambda_i) + m_{2t,\infty}\,\text{var}_{i,\infty}(u_{i0}) + \sum_{j=0}^{t-1} m_{2j,\infty}. \qquad (4)$$

To be precise about $Y_{t,\infty}$ and $V_{t,\infty}$ in the mixed-panel case, we first consider the two homogeneous cases separately. The first case arises when all units are stationary; the second occurs when all units are nonstationary.

In the stationary homogeneous panel case, $\alpha_i \in [0,\gamma]$, $\gamma < 1$ for every i. For $t > t^*$ such that $m_{t^*,\infty}$ is negligible, it is easy to see that $Y^0_{t,\infty}$ is an asymptotically stationary process with a finite mean Λ_∞ and a finite variance. Furthermore, its autocovariance at lag k tends toward 0 as $k \to \infty$, because $m^0_{k,\infty} \to 0$ as $k \to \infty$.

In the nonstationary homogeneous case, $\alpha_i = 1$ for all i, and thus $m_{j,\infty} = 1$ for all j. The average of individually I(1) processes is itself a unit root process. Then (3) and (4) become

$$Y^1_{t,\infty} = \Lambda_\infty + U^1_{0,\infty} + \sum_{j=1}^{t-1}\epsilon_{t-j,\infty} \qquad (5)$$

and

$$V^1_{t,\infty} = \text{var}_{i,\infty}(\lambda_i) + \text{var}^1_{i,\infty}(u_{i0}) + t. \qquad (6)$$

Notably, the cross-sectional mean is not time-invariant, and the cross-sectional variance $V^1_{t,\infty}$ has a linear trend component. The fact that $Y_{t,\infty}$ has a time-invariant variance only when all of the underlying units are stationary suggests that in a mixed panel when α_i is unity for some units, the cross-sectional variance will still have some form of a trend. We now make this precise.

2.1 Mixed Panel

Suppose that a random variable is drawn from a distribution with mean Y^1 and variance V^1 with probability θ, or a distribution with mean Y^0 and variance V^0 with probability $1-\theta$. It is well known that the mixture distribution has mean and variance

$$Y = \theta Y^1 + (1-\theta)Y^0$$

and

$$V = \theta V^1 + (1-\theta)V^0 + \theta(1-\theta)(Y^1 - Y^0)^2.$$

When applied to our mixed panel, the first result implies that the aggregate variable will be nonstationarity, because a weighted sum of an I(1) process and an I(0) process remains nonstationary. Our point of departure is the second result that the cross-sectional variance of any mixed panel is a mixture of the within-group variance, $\theta V^1 + (1-\theta)V^0$, and the between-group variance, $(Y^1 - Y^0)^2$. The following lemma characterizes the properties of the two components when the panel is mixed with nonstationary and stationary units.

Lemma 1. Let $y_{it} = \lambda_i + u_{it}$, and $u_{it} = \alpha_i u_{it-1} + e_{it}$, and $e_{it} \sim \text{iid}(0,1)$. Suppose that $\alpha_i = 1$ with probability θ, and $\alpha_i < 1$ with probability $1-\theta$. For a given $t > t^*$ such that $m^0_{t^*,\infty} \approx 0$, the cross-sectional variance of the mixed population is

$$V_{t,\infty} \approx \Lambda_\infty + \theta \cdot t + c,$$

where $c > 0$ is constant.

In a population mixed with stationary and nonstationary units, the cross-sectional variance has a linear trend component that grows over time at rate θ, the fraction of the panel that is nonstationary. Importantly, this trending property is unique to mixing stationary and nonstationary units. It does not exist if α_i is drawn from two distributions both with upper bounds strictly less than unity.

To understand Lemma 1, note first that $V^j_{t,\infty} = \text{var}_{i,\infty}(\lambda_i) + s^j_{t,\infty}$ for $j = 0, 1$, and thus

$$\begin{aligned} V_{t,\infty} &= \theta V^1_{t,\infty} + (1-\theta)V^0_{t,\infty} + \theta(1-\theta)(Y^1_{t,\infty} - Y^0_{t,\infty})^2 \\ &= \text{var}_{i,\infty}(\lambda_i) + \theta[\text{var}^1_{i,\infty}(u_{i0}) + t] \\ &\quad + (1-\theta)s^0_{t,\infty} + \theta(1-\theta)(Y^1_{t,\infty} - Y^0_{t,\infty})^2. \end{aligned}$$

Because $\text{var}_{i,\infty}(\lambda_i)$ and $\text{var}^1_{i,\infty}(u_{i0})$ do not depend on t, Lemma 1 thus implies that $s^0_{t,\infty}$ and $(Y^1_{t,\infty} - Y^0_{t,\infty})^2$ are constant when $t > t^*$. That $s^0_{t,\infty}$ is asymptotically constant follows from stationarity because $m_{t,\infty}$ is decreasing in t. Now $(Y^1_{t,\infty} - Y^0_{t,\infty}) = (U^1_{t,\infty} - U^0_{t,\infty})$. Because $U^0_{t,\infty} \approx 0$ for $t > t^*$, it remains to consider $U^1_{t,\infty}$.

It is well known that if $U^1_{t,N_1} = U^1_{t-1,N_1} + \epsilon^1_{t,N_1}$ is a random walk and $\epsilon^1_{t,N_1} \sim \text{N}(0, \sigma^2_{\epsilon_1,N_1})$, then $(U^1_{t,N_1})^2$ is a heteroscedastic random walk with drift, where the drift equals $\sigma^2_{\epsilon_1,N_1}$ and the innovation variance of the random walk is $4\sigma^4_{\epsilon_1,N_1} \cdot t$. But by definition, ϵ^1_{t,N_1} is the average of e^1_{it} in the nonstationary subsample and obeys a central limit theory. Thus U^1_{t,N_1} is approximately a Gaussian random walk with innovation variance $\sigma^2_{\epsilon_1,N_1}$ that decreases as N_1 increases. Furthermore, $(U^1_{t,N_1})^2$ is a random

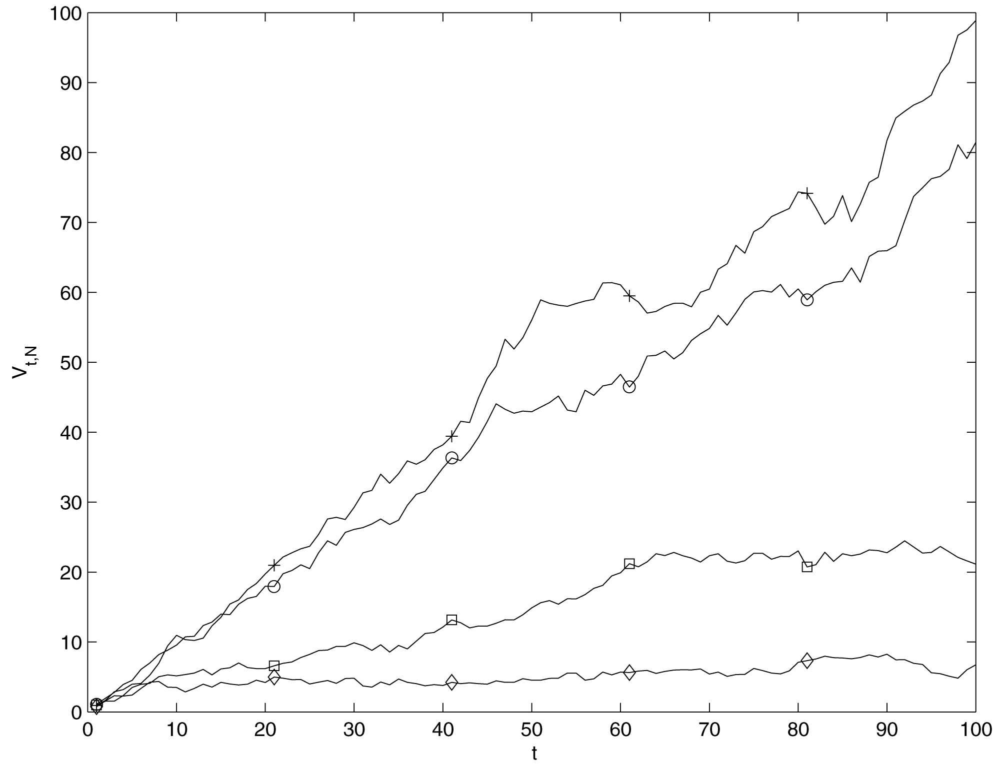

Figure 1. Cross-section variance for different θ, $n = 100$ ($\diamond$, .05; $\square$, .25; $\circ$, .75; +, 1).

walk with a drift of order $1/N_1$ and a variance of order t/N_1^2. Notably, both terms degenerate as $N_1 \to \infty$ for a given t. It follows that $U^1_{t,\infty} - U^0_{t,\infty}$ is degenerate and Lemma 1 holds for $t > t^*$. Importantly, Lemma 1 implies that

$$\Delta V_{t,\infty} = V_{t,\infty} - V_{t-1,\infty} = \theta.$$

That is, when there is a mixture of I(1) and I(0) units, the population cross-sectional variance grows at rate θ, where θ is the fraction of I(1) units in the panel. This is a population result. To see that the result continues to hold true in finite samples, Figure 1 presents the cross-sectional variance for $\boldsymbol{\theta} = (.05, .25., .75., 1)$ with $(N, T) = (100, 100)$. The stationary units are assumed to be AR(1) processes with the AR(1) coefficient drawn from a uniform distribution over the support [.5, .99]. As we can see, when $\theta = 1$, $V_{t,N}$ is roughly linear in t. The slope of $V_{t,N}$ flattens as θ moves toward 0. This suggests that θ remains the slope of $V_{t,N}$ in finite samples. The next section suggests procedures for estimating θ.

3. ESTIMATION AND INFERENCE

We observe y_{it}, $i = 1, \dots, N$, $t = 1, \dots, T$, of which N_0 observations have $\alpha_i < 1$ and N_1 observations have $\alpha_i = 1$. Then $\theta = N_1/N$ and $N = N_0 + N_1$. Without loss of generality, the data are ordered such that the N_1 nonstationary units come first. We do not know this ordering, nor do we know N_0 or N_1, but we can compute, at each t,

$$Y_{t,N} = \frac{1}{N}\sum_{i=1}^{N} y_{it}$$

and

$$V_{t,N} = \frac{1}{N}\sum_{i=1}^{N}(y_{it} - Y_{t,N})^2.$$

Theorem 1. Suppose that for $i = 1, \dots, N$, $t = 1, \dots, T$, y_{it} is generated as $y_{it} = \lambda_i + u_{it}$, $u_{it} = \alpha_i u_{i-1} + e_{it}$, and that assumptions A1–A3 hold. Let θ be the fraction of the mixed population with $\alpha_i = 1$. Define $\hat\theta = \frac{1}{T}\sum_{t=1}^{T} \Delta V_{t,N}$. As $N \to \infty$ and then $T \to \infty$,

$$\sqrt{N}(\hat\theta - \theta) \xrightarrow{d} \mathrm{N}(0, 2\theta).$$

Note that the limiting distribution of $\hat\theta$ is discontinuous at $\theta = 0$. When all units are stationary, the cross-sectional variance does not grow and the theory is no longer valid.

The sample cross-sectional variance, $V_{t,N}$, is $\sqrt{N}$ consistent for the population cross-sectional variance $V_{t,\infty}$. Lemma 1 suggests that the time average of $\Delta V_{t,N}$ will estimate $\Delta V_{t,\infty}$, which is θ. However, as shown in the Appendix, the variance of $\Delta V_{t,N}$ increases with t. Thus $\hat\theta$ is only $\sqrt{N}$-consistent even though $N, T \to \infty$.

One might wonder why we estimate the mean of $\Delta V^1_{t,N}$ and not the trend of $V^1_{t,N}$ directly. We do this because for fixed N, $(Y^1_{t,N_1} - Y^0_{t,N_0})$ is a unit root process and will render a regression of $V_{t,N}$ on t spurious when left as an error in the regression model. But this stochastic trend can be removed by first differencing $V_{t,N}$ as suggested in Theorem 1.

3.1 Higher-Order Serial Correlation

Theorem 1 provides the motivation for $\hat\theta$, but the underlying assumptions are restrictive. We now consider more general dynamics and heteroscedastic errors. Suppose that

$$y_{it} = \lambda_i + u_{it}, \qquad \alpha_i(L)u_{it} = e_{it}, \quad \text{and} \quad e_{it} \sim \text{iid}(0, \sigma_i^2),$$

where $\alpha_i(L) = 1 - \alpha_{i1}L - \cdots - \alpha_{ip}L^p$, p finite. Let $\phi_{i1} > \phi_{i2} > \cdots > \phi_{ip}$ be the reciprocal of the roots of $\alpha_i(L) = 0$. We can rewrite the model as

$$y_{it} = \lambda_i + \frac{e_{it}}{(1 - \phi_{i1}L)(1 - \phi_{i2}L)\cdots(1 - \phi_{ip}L)}.$$

By the method of partial fractions,

$$y_{it} = \lambda_i + u_{i0}\sum_{j=1}^{p}\phi_{ij}^t + \frac{1}{D_i}\sum_{j=1}^{t-1}[A_{i1}\phi_{i1}^j + A_{i2}\phi_{i2}^j + \cdots + A_{ip}\phi_{ip}^j]e_{it-j}, \quad (7)$$

where $D_i < \infty$ is the common denominator of the partial fractions and is discussed later. Direct calculations show that $A_{i1} = \phi_{i1}^{p-1}$. Whether or not a unit is stationary or nonstationary units is distinguished by the size of the largest autoregressive root, ϕ_{i1}. If $\phi_{i1} = 1$ and all remaining roots are strictly <1, then $A_{i1} = 1$ and $A_{ik} < 1$ for every $k = 2, \ldots, p$. For such a series, its variance will be $\frac{\sigma_i^2}{D_i^2}t$ plus terms that grow slower than t. It follows that if all units in the panel are nonstationary, then the cross-sectional variance of the rescaled process $\frac{D_i y_{it}}{\sigma_i}$ series will be dominated by t. Thus higher-order dynamics does not change our main result that if a fraction θ of the panel has an unit autoregressive root, then the cross-sectional variance grows at rate θ, provided that the cross-sectional variance is defined on the rescaled process. This suggests that θ can be estimated as follows.

Estimator A.

1. For each i, let $\hat{e}_{it}$ be the residuals from a least squares regression of the model
$$y_{it} = \alpha_0 + \alpha_1 y_{it-1} + \cdots + \alpha_p y_{it-p} + e_{it}. \quad (8)$$
Denote the coefficient estimates by $(\hat{\alpha}_0, \hat{\alpha}_{i1}, \ldots, \hat{\alpha}_{ip})$.
2. Let $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}^2$.
3. Let $\hat{\phi}_{i1}, \ldots, \hat{\phi}_{ip}$ be the reciprocal of the roots of $(1 - \hat{\alpha}_{i1}L - \cdots - \hat{\alpha}_{ip}L^p) = 0$.
4. Construct $\hat{D}_i$ and $\hat{V}_{t,N}$, the cross-sectional variance of $\hat{y}_{it} = \frac{\hat{D}_i y_{it}}{\hat{\sigma}_i}$.
5. Let $\hat{\theta} = \frac{1}{T}\sum_{t=1}^{T}\Delta\hat{V}_{t,N}$ and $\hat{\eta}_{t,N} = \Delta\hat{V}_{t,N} - \hat{\theta}$. Let $\hat{\gamma}_\eta(k) = \frac{1}{T}\sum_{t=1}^{T-1}\hat{\eta}_{t,N}\hat{\eta}_{t-k,N}$ and $\omega_{\hat{\theta}}^2 = \hat{\gamma}_\eta(0) + 2\sum_{s=1}^{M}K(s,M)\hat{\gamma}_\eta(s)$, where $K(s,M)$ is a kernel with a truncation point of M.

Then $\hat{\theta} = \frac{1}{T}\sum_{t=1}^{T}\Delta\hat{V}_{t,N}$ is an estimate of θ, just as in Theorem 1. The studentized statistic

$$t = \frac{(\hat{\theta} - \theta)}{\omega_{\hat{\theta}}} \approx \text{N}(0, 1)$$

can be used to test the null hypothesis that $\theta = \theta^0$ for any $\theta^0 \in (0, 1]$.

Step 1 requires estimation of an autoregressive model of order p. If y_{it} is stationary, then $\hat{\alpha}_{i1}, \ldots, \hat{\alpha}_{ip}$ are all $\sqrt{T}$ consistent. If y_{it} is nonstationary, then $\hat{\alpha}_i(1)$ is super-consistent. By continuous mapping, $\hat{\phi}_i \xrightarrow{p} \phi_i$ and $\hat{\sigma}_i^2 \xrightarrow{p} E(e_{it}^2)$. Because the raw data have different innovation variances and dynamics, the cross-sectional variance must be constructed on the rescaled data. Step 4 performs this task. In finite samples, terms that should be degenerate when t and N are large may still have time variation. Consequently, $\hat{\eta}_{t,N}$ may be serially correlated. For this reason, we use robust standard errors to conduct inference, as indicated in step 5.

It should be noted that Theorem 1 assumes that ϕ_1 is known and there are no nuisance parameters to estimate. Therefore, once consistent estimates of the cross-sectional variance are obtained, time-averaging of $\Delta V_{t,N}$ yields a $\sqrt{N}$-consistent estimate of θ. Estimator A, however, requires $\sqrt{T}$ consistent estimates of D_i and σ_i^2 before proceeding to estimate the cross-sectional variances, and heteroscedasticity autocorrelation-consistent standard errors are also needed. For this reason, we do not state Estimator A as a theorem. Instead, we only say that the standardized test is approximately normal, and turn to simulations to evaluate the properties of Estimator A.

It remains to make precise how to construct $\hat{\mathbf{D}}_i$. Suppose that $\phi_{i1}, \ldots, \phi_{ip}$ are observed. For a pth-order autoregression, $\mathbf{D}_i = \mathbf{A}^{-1}\mathbf{B}$, where $\mathbf{B}$ is $p \times 1$ vector with 1 in the first entry and 0 otherwise. The matrix $\mathbf{A}$ is $p \times p$ and can be constructed as follows. Row 1 is a unit vector. The ith column of row 2 is $\sum_{j\neq i}\phi_i$. The ith column of row 3 is $\prod_{j,k\neq i}\phi_j\phi_k$. The ith column of row 4 is $\prod_{j,k,m\neq i}\phi_j\phi_k\phi_m$. The remaining rows can be obtained by deduction. For $p = 4$, we have

$$\mathbf{D}_i = \begin{pmatrix} 1 & 1 & 1 & 1 \\ \phi_2 + \phi_3 + \phi_4 & \phi_1 + \phi_3 + \phi_4 & \phi_1 + \phi_2 + \phi_4 & \phi_1 + \phi_2 + \phi_3 \\ \phi_2\phi_3 + \phi_2\phi_4 + \phi_3\phi_4 & \phi_1\phi_3 + \phi_1\phi_4 + \phi_3\phi_4 & \phi_1\phi_2 + \phi_1\phi_4 + \phi_2\phi_4 & \phi_1\phi_2 + \phi_1\phi_3 + \phi_2\phi_3 \\ \phi_2\phi_3\phi_4 & \phi_1\phi_3\phi_4 & \phi_1\phi_2\phi_4 & \phi_1\phi_2\phi_3 \end{pmatrix}^{-1} \times \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

For convenience, we present $\mathbf{D}_i$ for $p = 1$–4:

p	D_i	A_{i1}
1	1	1
2	$\phi_{i1} - \phi_{i2}$	ϕ_{i1}
3	$\phi_{i1}^2 - \phi_{i1}\phi_{i2} - \phi_{i2}\phi_{i3} - \phi_{i1}\phi_{i3}$	ϕ_{i1}^2
4	$\phi_{i1}^3 - \phi_{i1}^2(\phi_{i2} + \phi_{i3} + \phi_{i4})$ $+\phi_{i1}(\phi_{i2}\phi_{i4} + \phi_{i2}\phi_{i3} + \phi_{i3}\phi_{i4}) - \phi_{i2} - \phi_{i3} - \phi_{i4}$	ϕ_{i1}^3

Replacing $\phi_{i1}, \ldots, \phi_{ip}$ by their consistent estimates yields $\hat{\mathbf{D}}_i$.

3.2 Cross-Sectional Correlation

Panel unit root and stationarity tests developed under the assumption of cross-sectional independence are not robust to cross-sectional correlation in the errors. A large literature has been developed to remedy this problem. Tests of Moon and Perron (2004) and Bai and Ng (2002) assume a factor structure, but power loss can be expected when the errors are in fact not strongly cross-sectionally correlated. Chang (2002) developed a panel unit root test, and Harris, Leybourne, and McCabe (2004) developed a panel stationarity test to allow for general cross-sectional dependence. But the former maintained as a null hypothesis that all units are nonstationary, whereas the latter assumed that all units are stationary under the null. Researchers might find it useful to know just how overwhelming the evidence against the null hypothesis is. An estimate of θ provides this information.

Suppose that the data-generating process is now

$$y_{it} = \lambda_i F_t + u_{it} \tag{9}$$

and

$$\alpha_i(L)u_{it} = e_{it}, \tag{10}$$

where $\alpha_i(L)$ is a finite pth-order polynomial in L. The variable F_t in (9) may or may not be observed; if it is not observed, then a good proxy is assumed to be available. Two approaches that have been used in the literature are the principal component of y_{it} and the average of y_{it}. This variable, F_t, generates correlation between two series i and j whenever $\lambda_i, \lambda_j \neq 0$. In the test of Bai and Ng (2002), the largest eigenvalue of the $N \times N$ population covariance matrix of y_{it} increases with N, so that F_t is a pervasive factor. Here we do not impose such a restriction to permit cross-sectional correlation to be strong or weak. We only assume, as in A2, that $\text{var}_{i,\infty}(\lambda_i) > 0$. In addition to A1, A2, and A3, we further assume the following:

A4. $\beta(L)F_t = v_t$, where the roots of $\beta(L) = 0$ are outside the unit circle.

A5. $E(u_{it}v_s) = 0$ for all i, t, and s; $E(\alpha_i v_s) = 0$ for all i and s; and $E(\lambda_i \alpha_j) = 0$ for all i and j.

The source of covariation, F_t, may or may not be observed. However, it must be stationary by A4. To understand how F_t affects the estimation of θ, consider, without loss of generality, $p = 1$ and $e_{it} \sim \text{iid}(0, 1)$. The population cross-sectional variance of y_{it} under (9) is

$$V_{t,\infty} = \text{var}_{i,\infty}(\lambda_i)F_t^2 + \text{var}_i(u_{i0})m_{2t} + \sum_{j=0}^{t-1} m_{2j}.$$

The time series properties of the cross-sectional variance evidently depend on the within-group variance of λ_i and on the properties of α_i and F_t. If $\text{var}_{i,\infty}(\lambda_i) = 0$, then, trivially, $V_{t,\infty}$ is approximately constant if the entire panel is stationary and $t > t^*$, whereas $V_{t,\infty} = t$ if every unit in the panel has $\alpha_i = 1$, whatever the properties of F_t. An identical response to a common source of variation does not induce additional cross-section variance over the case considered in Theorem 1. Not surprisingly, Theorem 1 continues to hold.

More generally, there will be variation in λ_i. But if F_t is stationary, then F_t^2 will remain stationary. Consequently, the cross-sectional variance of the nonstationary panel will still be dominated by the linear trend component. To make this point clear, suppose that $F_t = \beta F_{t-1} + v_t$ and $\alpha_i(L)u_{it} = e_{it}$. Then (7) now becomes

$$y_{it} = u_{i0}\sum_{j=1}^{p}\phi_{ij}^t + \frac{1}{D_i}\sum_{j=1}^{t-1}[A_{i1}\phi_{i1}^j + A_{i2}\phi_{i2}^j + \cdots + A_{ip}\phi_{ip}^j]e_{it-j} + \lambda_i\left[\sum_{j=0}^{t}\beta^j v_{t-j}\right]. \tag{11}$$

If $\phi_{i1} = 1$ and $\beta < 1$, then the variance of the series is still $\frac{\sigma_i^2 t}{D_i}$ plus terms that grow at a slower rate. The cross-sectional variance of the rescaled panel $\frac{D_i y_{it}}{\sigma_i}$ will still increase over time at rate θ. Therefore, Estimator A remains valid if ΔF_t^2 is treated as a residual in $\Delta V_{t,N} - \hat{\theta}$. But because F_t can be serially correlated, the use of robust standard errors to conduct inference will be essential in this situation. This way of controlling for cross-sectional correlation can be appealing if F_t is not observed and no suitable estimate of it is available. However, controlling for F_t when constructing D_i may yield more precise estimates of the largest autoregressive root. Toward this end, consider an alternative estimator that generalizes Estimator A to allow for cross-sectional correlation.

Estimator B.

1. For each i, let $\hat{e}_{it}$ be the residuals from a least squares regression of the model

$$y_{it} = \alpha_0 + \alpha_1 y_{it-1} + \cdots + \alpha_p y_{it-p} + \lambda_{i0}\hat{F}_t + \cdots + \lambda_{iq}\hat{F}_{t-q} + e_{it}, \tag{12}$$

where $\hat{F}_t$ can be an aggregate variable or the first principal component of a large panel. Denote the coefficient estimates by $(\hat{\alpha}_0, \hat{\alpha}_{i1}, \ldots, \hat{\alpha}_{ip}, \hat{\lambda}_{i0}, \ldots, \hat{\lambda}_{iq})$.

Then follow steps 2–6 in Estimator A. For large N and large T, $\hat{\theta} = \frac{1}{T}\sum_{t=1}^{T}\Delta\hat{V}_{t,N} \xrightarrow{p} \theta$ and $(\hat{\theta} - \theta)/\omega_{\hat{\theta}} \sim \text{N}(0, 1)$.

4. FINITE–SAMPLE PROPERTIES

We use two models to evaluate the finite-sample properties of our estimator for θ. In each case, the error process is specified as

$$u_{it} = (\alpha_{i1} + \alpha_{i2})u_{it-1} - \alpha_{i2}u_{it-2} + e_{it} + \psi e_{it-1}$$

and

$$e_{it} \sim \text{N}(0, \sigma_i^2).$$

The nonstationary units are generated by letting $\alpha_{i1} = 1$, so that by construction, the largest autoregressive root is unity. The stationary units are generated with $\alpha_{i1} \sim \text{U}[.5, \gamma]$. Many configurations were considered. To conserve space, we report results only for the following parameterizations:

- Model 1: $y_{it} = \lambda_i + u_{it}$
  - Model 1a: $\lambda_i \sim \text{U}[-1, 1]$, $\alpha_{i2} \sim \text{U}[0, .2]$, $\psi = 0$ $\sigma_i \sim \text{U}[.5, 2]$

- Model 1b: $\lambda_i \sim$ U[−1, 1]; $\alpha_{i2} = 0$, $\psi_i \sim$ U(−.8, 0), $\sigma_i^2 = 1\ \forall i$
- Model 2: $y_{it} = \lambda_i F_t + u_{it}$
  - $\alpha_{i2} \sim$ U[0, .2], $\psi_i = 0\ \forall i$, $\sigma_i \sim$ U[.5, 2], $\beta = .5$
  - Model 2a: Pr($\lambda_i = 0$) = .5, Pr($\lambda_i \sim$ N(0, 1)) = .5
  - Model 2b: $\lambda_i \sim$ U(0, 1).

Model 1 has no cross-sectional correlation. Whereas the stationary units in Model 1a are AR(2) processes, those in Model 1b are MA(1) processes. Cross-sectional dependence is allowed in Model 2. Models 2a and 2b have the same dynamic error structure as Model 1a. Note that because α_i varies across units, there is unconditional cross-sectional heteroscedasticity in every case.

Simulations were conducted with $\gamma = .8$ and .99, where γ is the upper bound on the persistence parameter of the stationary units. The results are similar, and we report only those for $\gamma = .99$. Various combinations of N and T were considered. Although in theory N and T both need to be large, simulations reveal that the estimator works well with $N \geq 30$. The results are much more sensitive to T. Here, we report results for $N = 30, 60$ and $T = 100, 200$. We note, however, that increasing N from 60 to 100 yields little gain, but increasing T from 50 to 100 drastically reduces the variance of $\hat{\theta}$. With $T \geq 200$ (not reported), the estimator is approximately mean and median unbiased. As a matter of practice, the estimator is more suited for panel data with $T \geq 100$.

Because N_1 must be an integer, θ is defined in the simulations as

$$\theta = \frac{\text{ceil}(\theta^* N)}{N},$$

with $\theta^* = [.01, .05, .5, .95, 1]$. This means that $N_1 = \text{ceil}(\theta^* N)$. For example, if $N = 30$, then $\theta^* = .05$, $N_1 = 2$, and $\theta = .067$. Because N_1 must be at least 1 (as θ must be >0 for the analysis to be valid), the smallest θ considered in the simulations is .017, which occurs when $N = 60$ and $N_1 = 1$.

In the simulations, we fix p to 2 to assess the adequacy of the asymptotic approximation provided by Theorem 1. However, in applications, a data-dependent method, such as the Bayes information criterion (BIC), can be used. The Newey–West kernel, $K(s, M) = 1 - \frac{s}{M+1}$, with a lag length of $M = 2$, is used to construct robust t statistics. Results for $M = 1$ and 4 are similar. Three hypothesis are being tested:

A. $H_0^A : \theta = \theta^0$; $H_1^A : \theta < \theta^0$; $H_2^A : \theta \neq \theta_0$,
B. $H_0^B : \theta = .01$; $H_1^B : \theta > .01$, and
C. $H_0^C : \theta = 1$; $H_1^C : \theta < 1$.

Hypothesis C tests that all units are I(1) against the alternative that at least some unit is stationary. Although we cannot test the hypothesis that all units are stationary, we can test whether θ is slightly larger than 0, as we do in Hypothesis B. If we reject (B), then we also must reject the hypothesis that $\theta = 0$. Hypothesis A allows the researcher to test hypotheses other than the two extremes. In the simulations, we test whether $\theta = \theta^0$, the true θ. The tests are based on the 5% asymptotic critical value. Thus, for hypotheses (A and C), the critical value is −1.64, and for B, it is 1.64. The critical value for the two-sided test under A is 1.96.

Table 1 reports results for Model 1 using Estimator A. The finite-sample rejection rates, the mean and median estimates of θ, and the variance of $\hat{\theta}$ in the simulations (scaled by N/θ) are provided. At $(N, T) = (30, 100)$, the finite-sample variance of the estimator is larger than suggested by theory, resulting in some size distortion in the studentized tests. However, the variance of the estimator tends to the theoretical value of 2 as N and T increase. Univariate unit root tests tend to exhibit enormous size distortions when the moving average parameter is negative, and panel tests also tend to have this problem, albeit less severe. Our estimates of θ are not sensitive to the presence of a negative moving average parameter, as shown in the lower part of Table 1. We also estimated Models 1a and 1b using Estimator B instead of Estimator A, that is, when cross-sectional correlation was not present but was controlled for. The results are very similar to those in Table 1 and thus are not reported. For example, when $(T, N) = (200, 60)$ and $\theta = .5$, $\hat{\theta}$ is .535 on average, compared with .518 in Table 1. This is not surprising, because the inclusion of irrelevant regressors does not affect consistent estimation of the dynamic parameters and thus of D_i.

The two-sided test for the null hypothesis that $\theta = \theta^0$ is more precise than the one-sided test. Note that when Hypothesis B (that $\theta = .01$) is tested, the rejection rate is only around .1 even though the true θ is .068. In addition, when Hypothesis C (that $\theta = 1$) is tested and the true θ is .95, the test has a rejection rate of around .1, even though the hypothesis is false. This shows that the power of testing $\theta = \theta^0$ against close alternatives can be low, a result shared by univariate unit root and stationarity tests. When the true θ is .5 and the hypothesis that $\theta = .01$ or $\theta = 1$ is tested, rejection rates are well below 1, even with larger (N, T). More importantly, it is not clear how a researcher should proceed when Hypotheses B and C are both rejected. In contrast, $\hat{\theta}$ is informative about the extent of heterogeneity among the panel units.

Table 2 presents results for the DGP 2 using Estimator B; Table 3, those using Estimator A. The difference is that Estimator B explicitly controls for cross-sectional correlation. Pesaran (2005) suggested that $\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{it}$ and its lags can approximate the latent stationary common factor well when N is large. We use $(\Delta\bar{y}_t, \Delta\bar{y}_{t-1})$ to approximate F_t because stationarity of F_t is a crucial aspect of the analysis. The estimator remains fairly precise whether the cross-sectional correlation is strong (bottom part) or partial (top part). An alternative is to proxy F_t by the first principal component of Δy_{it}. The results are similar and thus are not reported.

Table 3 reveals that $\hat{\theta}$ tends to be downward-biased when cross-sectional correlation is omitted. The bias can be nontrivial when the cross-sectional correlation is strong. For example, when $(T, N) = (100, 60)$, $\theta = .5$, and there is a common factor (bottom part), the mean and median of $\hat{\theta}$ are .464 and .449, instead of .498 and .482 as reported in the bottom part of Table 2. Although the bias is smaller when we omit weak correlation (top part), in practice we often do not know whether cross-sectional correlation is strong or weak or present at all. Controlling for cross-sectional correlation when none is present provides more robust inference than ignoring the correlation when it is in fact strong.

The proposed estimator is $\sqrt{N}$ consistent for the fraction of units that are nonstationary. Then $\hat{N}_1 = [\hat{\theta} \cdot N]$ is the estimated number of nonstationary units, where $[\cdot]$ denotes the

Table 1. DGP: $y_{it} = \lambda_i + u_{it}$, $u_{it} = (\alpha_{i1} + \alpha_{i2})u_{it-1} - \alpha_{i2}u_{it-2} + e_{it} + \psi e_{it-1}$, $e_{it} \sim \mathrm{N}(0, \sigma_i^2)$

T	N	θ^0	$\hat{\theta}$	med($\hat{\theta}$)	$N\,\mathrm{var}(\hat{\theta})/\theta$	$H_0: \theta = \theta^0$ $H_1: \theta < \theta^0$	$H_0: \theta = \theta^0$ $H_1: \theta \neq \theta^0$	$H_0: \theta = .01$ $H_1: \theta > \theta^0$	$H_0: \theta = 1$ $H_1: \theta < \theta^0$
Estimator A: $\sigma_i \sim \mathrm{U}[.5, 2]$, $\alpha_{i2} \sim \mathrm{U}[0, \alpha_2^b]$, $\alpha_2^b, = .2$, $\psi = 0$									
100	30	.033	.073	.056	3.119	0	.023	.102	1.000
100	30	.067	.087	.062	2.946	.004	.026	.128	.999
100	30	.500	.517	.474	2.996	.116	.114	.931	.783
100	30	.967	.996	.957	2.776	.094	.114	1.000	.119
100	30	1.000	1.035	1.002	2.923	.099	.103	1.000	.099
200	30	.033	.037	.017	2.348	.001	.016	.057	1.000
200	30	.067	.073	.053	1.903	.018	.012	.135	1.000
200	30	.500	.516	.492	2.362	.082	.079	.934	.805
200	30	.967	.961	.924	2.403	.094	.087	.999	.123
200	30	1.000	.999	.970	2.247	.089	.070	1.000	.089
100	60	.017	.037	.028	3.752	0	.027	.058	1.000
100	60	.050	.072	.058	3.086	0	.031	.229	1.000
100	60	.500	.531	.513	2.916	.082	.114	1.000	.922
100	60	.950	.962	.953	2.657	.097	.111	1.000	.156
100	60	1.000	1.017	1.010	2.751	.106	.120	1.000	.106
200	60	.017	.027	.019	2.219	0	.013	.029	1.000
200	60	.050	.056	.045	2.322	.026	.031	.174	1.000
200	60	.500	.509	.497	2.318	.074	.085	1.000	.946
200	60	.950	.964	.953	2.414	.074	.087	1.000	.125
200	60	1.000	1.011	.999	2.148	.059	.066	1.000	.059
Estimator A: $\psi \sim \mathrm{U}(-.80, 0)$, $\alpha_2^b = 0$, $\sigma_i = 1$									
100	30	.033	.043	.024	2.909	0	.015	.049	1.000
100	30	.067	.074	.053	2.530	.025	.024	.115	1.000
100	30	.500	.514	.494	2.579	.103	.092	.924	.798
100	30	.967	.989	.956	2.706	.084	.092	.996	.115
100	30	1.000	1.008	.972	2.666	.089	.088	.998	.089
200	30	.033	.040	.023	1.855	0	.009	.041	1.000
200	30	.067	.070	.047	2.421	.080	.054	.118	1.000
200	30	.500	.507	.474	2.209	.068	.063	.933	.814
200	30	.967	.961	.949	2.229	.089	.069	.999	.109
200	30	1.000	.995	.977	2.258	.089	.071	1.000	.089
100	60	.017	.026	.017	2.702	0	.008	.034	1.000
100	60	.050	.064	.052	2.625	0	.019	.175	1.000
100	60	.500	.506	.492	2.396	.092	.089	1.000	.950
100	60	.950	.961	.949	2.650	.091	.095	1.000	.139
100	60	1.000	1.015	1.001	2.505	.085	.083	1.000	.085
200	60	.017	.025	.018	1.822	0	.004	.023	1.000
200	60	.050	.055	.043	2.385	.016	.014	.142	1.000
200	60	.500	.505	.493	2.361	.080	.077	1.000	.942
200	60	.950	.961	.946	2.230	.071	.067	1.000	.115
200	60	1.000	1.011	1.003	2.263	.067	.070	1.000	.067

integer part. A natural question to ask is whether we can recover the identity of those units that are nonstationary. Two possibilities come to mind. The first is to consider all subsets of size $\hat{N}_1 = [\hat{\theta} \cdot N]$ for which panel unit root tests cannot reject. This is the approach followed by Kapetanios (2003), who proposed carrying out a sequence of panel unit root tests on a reduced dataset, where the reduction is achieved by dropping series deemed to be stationary. For large $\hat{N}_1$, this can be computationally challenging, and, more importantly, the solution set may not be unique.

A second approach is to take as starting point that units with an autoregressive unit root have variance that increases with time. Because these autoregressive roots (i.e., $\phi_{i1}, \ldots, \phi_{ip}$) are calculated when constructing $\hat{D}_i$, it seems natural to take advantage of this information. More precisely, we order the units by size $|\hat{\phi}_{i1}|$, the estimated largest autoregressive root, and take the first $\hat{N}_1$ units as the nonstationary ones. Note that whether a unit is classified to be in the stationary or the nonstationary panel is determined by the ranking of $|\hat{\phi}_{i1}|$ and $\hat{\theta}$. Individual unit root tests are not conducted. In simple experiments, this scheme works very well. The correct classification rate is around .6 even when N and T are quite small. A formal proof that the nonstationary units can be recovered is beyond the scope of this analysis. But the point to be emphasized is that in mixed panels, θ bears information about the cross-sectional properties of α_i that standard panel unit root tests cannot deliver.

Table 2. DGP: $y_{it} = \lambda_i F_t + u_{it}$, $u_{it} = \alpha_{i1} u_{it-1} + e_{it}$, $e_{it} \sim \text{N}(0, \sigma_i^2)$, $F_t = \beta F_{t-1} + w_t$, $w_t \sim \text{N}(0, 1)$

T	N	θ^0	$\hat{\theta}$	med($\hat{\theta}$)	$N\,\text{var}(\hat{\theta})/\theta$	$H_0: \theta = \theta^0$ $H_1: \theta < \theta^0$	$H_0: \theta = \theta^0$ $H_1: \theta \neq \theta^0$	$H_0: \theta = .01$ $H_1: \theta > \theta^0$	$H_0: \theta = 1$ $H_1: \theta < \theta^0$
Estimator B: $\beta = .5$, $\lambda_i \sim .5\,\text{N}(0, 1) + .5(0)$									
100	30	.033	.049	.032	2.535	0	.010	.045	1.000
100	30	.067	.088	.069	2.206	.001	.011	.094	1.000
100	30	.500	.499	.458	2.511	.085	.080	.901	.810
100	30	.967	.926	.888	2.607	.124	.112	.994	.151
100	30	1.000	1.042	1.004	2.924	.079	.092	1.000	.079
200	30	.033	.048	.031	2.516	0	.009	.042	1.000
200	30	.067	.095	.074	2.688	0	.020	.099	1.000
200	30	.500	.478	.453	2.011	.094	.066	.909	.848
200	30	.967	.961	.933	2.345	.070	.069	.998	.092
200	30	1.000	1.001	.981	2.328	.081	.068	.998	.081
100	60	.017	.041	.033	3.537	0	.012	.050	1.000
100	60	.050	.079	.070	2.576	0	.022	.184	1.000
100	60	.500	.496	.483	2.516	.091	.088	.999	.947
100	60	.950	.992	.974	2.819	.065	.100	1.000	.098
100	60	1.000	1.030	1.017	2.868	.074	.094	1.000	.074
200	60	.017	.029	.020	2.503	0	.010	.040	1.000
200	60	.050	.048	.039	1.402	.002	.005	.086	1.000
200	60	.500	.474	.460	2.112	.110	.068	.996	.971
200	60	.950	.956	.937	2.319	.064	.061	1.000	.102
200	60	1.000	.995	.979	2.242	.080	.064	1.000	.080
Estimator B: $\beta = .5$, $\lambda_i \sim \text{N}(0, 1)$									
100	30	.033	.040	.024	2.225	0	.007	.033	1.000
100	30	.067	.071	.053	1.924	0	.011	.071	1.000
100	30	.500	.476	.451	2.201	.117	.093	.901	.842
100	30	.967	1.014	.986	2.882	.065	.080	.998	.091
100	30	1.000	1.045	1.011	3.019	.074	.096	.997	.074
200	30	.033	.037	.018	2.519	0	.015	.043	1.000
200	30	.067	.048	.030	1.173	.087	.036	.053	1.000
200	30	.500	.440	.420	1.841	.145	.093	.859	.886
200	30	.967	.902	.875	2.276	.119	.090	.996	.154
200	30	1.000	.960	.931	2.404	.096	.089	.998	.096
100	60	.017	.031	.022	2.731	0	.015	.036	1.000
100	60	.050	.054	.043	2.195	.003	.019	.121	1.000
100	60	.500	.481	.462	2.409	.120	.089	.995	.953
100	60	.950	.936	.930	2.467	.115	.092	1.000	.161
100	60	1.000	1.001	.987	3.023	.098	.114	1.000	.098
200	60	.017	.020	.014	1.159	0	0	.009	1.000
200	60	.050	.053	.042	2.056	.011	.014	.122	1.000
200	60	.500	.469	.459	1.959	.110	.080	.999	.977
200	60	.950	.926	.912	2.082	.085	.064	1.000	.146
200	60	1.000	.957	.951	2.130	.117	.072	1.000	.117

5. NONSTATIONARY F_t

We use assumption A4 to rule out the case that F_t is nonstationary for two reasons. First, if F_t is pervasive and nonstationary, then every series in the panel is nonstationary. In such a case, θ is a misleading indicator of how many units in the panel are nonstationary, because it has nothing to do with the fraction of units in the mixed panel having $\alpha_i = 1$. Instead, we should estimate the number of common trends in the panel using the PANIC procedure developed by Bai and Ng (2004). Assumption A4 also rules out time-specific effects of the form $F_t = f(t)$, of which $F_t = t$ is the case of special interest. To understand why, consider

$$y_{it} = \lambda_i t + u_{it},$$

where $u_{it} = \alpha_i u_{it-1} + e_{it}$, $e_{it} \sim \text{iid}(0, \sigma^2)$. The cross-sectional variances are

$$V^1_{t,\infty} = \text{var}_i(\lambda_i) t^2 + s^1_{t,\infty}$$

and

$$V^0_{t,\infty} = \text{var}(\lambda_i) t^2 + s^0_{t,\infty}.$$

Although a linear trend that increases at the rate of θ still remains in the mixed panel, we will need to control for the quadratic trend to estimate θ.

Theorem 2. Suppose that $y_{it} = \lambda_i t + u_{it}$, λ_i is independent of e_{it} for all i and t, $u_{it} = \alpha_i u_{i-1} + e_{it}$, and $e_{it} \sim \text{iid}(0, 1)$. Let θ

Table 3. DGP: $y_{it} = \lambda_i F_t + u_{it}$, $u_{it} = \alpha_{i1} u_{it-1} + e_{it}$, $e_{it} \sim \mathrm{N}(0, \sigma_i^2)$, $F_t = \beta F_{t-1} + w_t$, $w_t \sim \mathrm{N}(0, 1)$

T	N	θ^0	$\hat{\theta}$	med($\hat{\theta}$)	$N\,\mathrm{var}(\hat{\theta})/\theta$	$H_0: \theta = \theta^0$ $H_1: \theta < \theta^0$	$H_0: \theta = \theta^0$ $H_1: \theta \neq \theta^0$	$H_0: \theta = .01$ $H_1: \theta > \theta^0$	$H_0: \theta = 1$ $H_1: \theta < \theta^0$
Estimator A: $\beta = .5$, $\lambda_i \sim .5\,\mathrm{N}(0, 1) + .5(0)$									
100	30	.033	.046	.030	2.307	0	.009	.043	1.000
100	30	.067	.083	.065	1.985	.002	.010	.091	1.000
100	30	.500	.475	.436	2.307	.114	.089	.905	.852
100	30	.967	.862	.827	2.279	.181	.140	.994	.223
100	30	1.000	.966	.929	2.501	.131	.105	1.000	.131
200	30	.033	.045	.029	2.296	0	.009	.041	1.000
200	30	.067	.092	.071	2.593	0	.021	.103	1.000
200	30	.500	.457	.433	1.847	.113	.084	.909	.885
200	30	.967	.913	.883	2.120	.106	.082	.997	.146
200	30	1.000	.950	.925	2.098	.116	.087	.999	.116
100	60	.017	.039	.032	3.245	0	.011	.050	1.000
100	60	.050	.076	.067	2.386	0	.021	.186	1.000
100	60	.500	.475	.461	2.329	.124	.105	.999	.965
100	60	.950	.941	.925	2.520	.103	.096	1.000	.166
100	60	1.000	.978	.962	2.588	.132	.107	1.000	.132
200	60	.017	.028	.020	2.395	0	.010	.040	1.000
200	60	.050	.047	.037	1.353	.008	.005	.086	1.000
200	60	.500	.463	.449	2.019	.129	.077	.997	.978
200	60	.950	.924	.903	2.163	.094	.075	1.000	.145
200	60	1.000	.962	.945	2.096	.109	.074	1.000	.109
Estimator A: $\beta = .5$, $\lambda_i \sim \mathrm{N}(0, 1)$									
100	30	.033	.038	.023	2.048	.001	.008	.033	1.000
100	30	.067	.068	.050	1.784	.001	.011	.071	1.000
100	30	.500	.452	.427	1.971	.150	.118	.903	.885
100	30	.967	.945	.915	2.476	.114	.095	.998	.141
100	30	1.000	.971	.944	2.599	.126	.117	.996	.126
200	30	.033	.036	.017	2.397	.001	.015	.042	1.000
200	30	.067	.047	.029	1.113	.107	.050	.050	1.000
200	30	.500	.410	.390	1.650	.206	.145	.854	.925
200	30	.967	.859	.829	2.068	.174	.118	.996	.217
200	30	1.000	.908	.881	2.154	.152	.108	.998	.152
100	60	.017	.029	.021	2.574	0	.017	.037	1.000
100	60	.050	.052	.040	2.022	.006	.017	.116	1.000
100	60	.500	.459	.442	2.200	.156	.115	.995	.973
100	60	.950	.890	.885	2.197	.159	.129	1.000	.232
100	60	1.000	.937	.926	2.642	.178	.140	1.000	.178
200	60	.017	.019	.014	1.077	0	0	.009	1.000
200	60	.050	.052	.041	1.965	.016	.016	.126	1.000
200	60	.500	.456	.446	1.851	.131	.099	.999	.983
200	60	.950	.896	.882	1.950	.123	.075	1.000	.187
200	60	1.000	.926	.919	1.986	.148	.106	1.000	.148

be the fraction of the panel with $\alpha_i = 1$. Let $\hat{\theta}$ be obtained from least squares estimation of

$$\Delta \hat{V}_{t,N} = \theta + \beta \Delta t^2 + \eta_{t,N}.$$

If $N, T \to \infty$, then

$$\frac{\sqrt{N}}{\sqrt{T}}(\hat{\theta} - \theta) \xrightarrow{d} \mathrm{N}\left(0, \frac{32\theta\,\mathrm{var}_{i,\infty}(\lambda_i)}{15}\right).$$

Because the regression error variance increases more quickly when there are individual specific time effects, the variance of the estimator vanishes only if $\frac{T}{N} \to 0$. Consistency of the estimator is not assured as in the earlier case when incidental trends were absent. Moon, Perron, and Phillips (2005) also found that panel unit root tests have lower power in the presence of "incidental trends." In our setting, the more heterogeneous the response to the common time effect, the more imprecise the estimated fraction of the panel that is nonstationary. However, $\hat{\beta}$ is $\sqrt{NT}$ consistent, where $\beta = \mathrm{var}(\lambda_i)$. Although not the focus, the present framework also provides an estimate of the variance of the incidental trends.

Implicit in Theorem 2 is the assumption that the nonstationary units are the only source of variation that is of order t. If there exists F_t (stochastic or deterministic) with variances proportional to t (e.g., $F_t = \sqrt{t}$), then we will not be able to identify θ. In that case, $\hat{\theta}$ estimates only an upper bound for θ. But, if this is the case, then all existing unit root tests (panel or univariate) also will be misspecified, because they too do not allow for such variations in the data.

As it stands, Theorem 2 assumes that e_{it} is iid across i and t. We want to allow for higher-order dynamics, heteroscedastic errors, and cross-sectional correlation. A slight variation to Estimator B leads to the following estimator.

Estimator C.

1. Let $\hat{e}_{it}$ be the residuals from a least squares regression of y_{it} on a constant, p lags of y_{it}, t, and possibly F_t or its proxies. Denote the autoregressive coefficient estimates by $(\hat{\alpha}_{i1}, \ldots, \hat{\alpha}_{ip})$.
2. Define $\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_{it}^2$.
3. Let $\hat{\phi}_{i1}, \ldots, \hat{\phi}_{ip}$ be the roots of $\hat{\alpha}_i(L) = 0$ and construct $\hat{D}_i$.
4. Let $\hat{y}_{it} = y_{it}/\hat{\sigma}_i$, and let $\hat{V}_{t,N}$ be the cross-sectional variance of $\hat{y}_{it}$.
5. Consider least squares regression of
$$\Delta \hat{V}_{t,N} = \theta + \beta \Delta t^2 + \eta_{t,N}. \tag{13}$$
6. Let $\omega_{\hat{\theta}}$ be the heteroscedastic autocorrelation-consistent standard errors. Then $t = \frac{(\hat{\theta}-\theta)}{\omega_{\hat{\theta}}} \approx \mathrm{N}(0,1)$ for large N and T.

Because we are interested only in the component in the cross-sectional variance that is linear in time, Estimator C is no longer the sample mean of $\Delta \hat{V}_{t,N}$. Instead, one must control for that component in $\hat{V}_{t,N}$ that grows quadratically with time. Nonetheless, the studentized statistic can be used for inference.

To verify the properties of Estimator C, we simulate data as follows:

- Model 3: $y_{it} = \lambda_i t + u_{it}$
  - Model 3a: $\lambda_i = 0$ with probability δ and $\lambda_i \sim \mathrm{U}(-.1, .1)$ with probability $1 - \delta$
  - Model 3b: $\lambda_i \sim \mathrm{U}(-.1, .1)$,

where u_{it} is the same as in Model 1a. The results are reported in Table 4. The estimator exhibits small mean and median bias, but the variance is large even with $T = 200$. We report results for $T = 300$ and 600 with $N = 200$ and 400, much larger than in the previous tables. Even with this many observations, the asymptotic approximation is not very good, especially when θ is small. Although the two-sided test is quite precise, the sample size required to further reduce the variability of the estimator severely restricts the estimator's usefulness.

6. APPLICATIONS

In the first application, we consider a panel of real exchange rates and estimate what fraction of the panel is nonstationary. Data for nominal exchange rates and the consumer price indices are obtained from the International Finance Statistics. We use data from 1974:1–1997:4 for 21 countries: Canada, Austria, New Zealand, Australia, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Netherlands, Norway, Spain, Sweden, Switzerland, U.K., Japan, Korea, Singapore, and Thailand. We use the United States as the numeraire country. Figure 2 plots the cross-sectional variance of the log real exchange rates. The cross-sectional variance does not exhibit a positive trend. If anything, the trend line is negatively sloped. With p is chosen by the BIC, Estimator A yields a $\hat{\theta}$ of $-.123$, and the t statistic for testing $\theta = .01$ is .114. Estimator B, using the averaged changes in real exchange rate to control for cross-sectional correlation, yields a $\hat{\theta}$ of $-.888$. The t statistic for testing $\theta = .01$ is $-.189$. Using the growth in U.S. industrial production as ΔF_t yields the same conclusion, that $\hat{\theta}$ is not statistically different from .01.

In a second application, we consider earnings of male household head in the PSID for whom 25 years of data are available from 1976 onward. This gives $N = 104$ and $T = 25$. Following the earnings dynamics literature, life cycle and observed heterogeneity effects are removed by regressing log earnings on age, age-squared, and education dummies, and the residuals (which we call income) are used for analysis. Covariance structure estimation using the moments of the residuals gives ARMA estimates of .856 and $-.219$, which are similar to results reported in the literature. However, such estimation ignores the possibility of heterogeneity in the dynamic parameters. The cross-sectional variance of income, plotted in Figure 3, seems to be drifting upward. We again select p using the BIC. Estimator A yields a $\hat{\theta}$ of .149, and the t statistic for testing $\theta = .01$ is 1.36. Estimator B yields a $\hat{\theta}$ of .188, with a t statistic of 1.486. The one-tailed 10% critical value is 1.281, and we reject the null hypothesis that $\theta = .01$ at the 10% level. However, we cannot reject the hypothesis that $\theta = .2$ or $\theta = .35$ whether we use Estimator A or B. Estimator C yields a $\hat{\theta}$ of $-.072$ and the hypothesis that $\theta = .01$ cannot be rejected. However, var(λ_i) is estimated to be .01 and not statistically different from 0, casting doubt on the validity of Estimator C. With the caveat of small T in mind, the data suggest that earnings for one-fifth to as many as one-third of the household heads in the sample may have stochastic trends.

7. CONCLUSION

This article exploits the fact that the cross-sectional variance of nonstationary units has a linear time trend to obtain an estimate of the fraction of units in a sample that are differenced stationary processes. The procedure is valid under cross-sectional correlation, provided that the correlation is stationary. An advantage of this approach is that the point estimate of θ is informative about the properties of the panel. In contrast, rejecting the null hypothesis that all units in the panel are nonstationary often leaves the researcher with little guidance on how to alternatively characterize the data.

ACKNOWLEDGMENTS

This article was presented at the CIREQ Time Series conference in Montreal. The author thanks conference participants, Benoit Perron, Chi-Young Choi, Nickolay Iskrev, Jushan Bai, two anonymous referees, and the associate editor for helpful comments, and Chang-Ching Lin for data assistance. Financial support from the National Science Foundation (grant SES-0345237) is gratefully acknowledged.

APPENDIX: PROOFS

Proof of Lemma 1

The DGP is $y_{it} = \lambda_i + u_{it}$, $u_{it} = \alpha_i u_{it-1} + e_{it}$ with $e_{it} \sim \mathrm{N}(0, \sigma_e^2)$ $\forall i$. Without loss of generality, $\sigma_e^2 = 1$. For each $t =$

Table 4. DGP: $y_{it} = \lambda_i t + u_{it}$, $u_{it} = \alpha_{i1} u_{it-1} + e_{it}$, $e_{it} \sim \mathrm{N}(0, 1)$

T	N	θ^0	$\hat{\theta}$	med($\hat{\theta}$)	$N\,\mathrm{var}(\hat{\theta})/\theta$	$H_0: \theta = \theta^0$ $H_1: \theta < \theta^0$	$H_0: \theta = \theta^0$ $H_1: \theta \neq \theta^0$	$H_0: \theta = .01$ $H_1: \theta > \theta^0$	$H_0: \theta = 1$ $H_1: \theta < \theta^0$
Estimator C: $\beta = .5$, $\lambda_i \sim .5\,\mathrm{N}(0, 1) + .5(0)$									
300	200	.010	.036	.034	18.592	.001	.016	.046	1.000
300	200	.050	.077	.074	5.725	0	.023	.261	1.000
300	200	.500	.493	.492	5.213	.079	.076	.988	.985
300	200	.950	.930	.926	4.699	.097	.087	1.000	.148
300	200	1.000	.962	.962	4.184	.106	.079	1.000	.106
600	200	.010	.027	.025	9.415	0	.003	.008	1.000
600	200	.050	.070	.072	9.570	.034	.051	.181	1.000
600	200	.500	.505	.505	2.604	.061	.072	.991	.983
600	200	.950	.937	.938	2.698	.094	.081	1.000	.154
600	200	1.000	.991	.988	2.464	.070	.067	1.000	.070
300	400	.010	.039	.038	16.035	0	.044	.096	1.000
300	400	.050	.075	.073	6.171	.003	.051	.488	1.000
300	400	.500	.500	.498	5.157	.070	.071	1.000	1.000
300	400	.950	.919	.913	3.998	.103	.068	1.000	.226
300	400	1.000	.967	.966	4.375	.124	.091	1.000	.124
600	400	.010	.026	.025	5.453	0	.001	.005	1.000
600	400	.050	.065	.063	6.391	.010	.038	.262	1.000
600	400	.500	.493	.489	2.586	.071	.070	1.000	.999
600	400	.950	.940	.940	2.614	.085	.065	1.000	.166
600	400	1.000	.984	.982	2.552	.088	.070	1.000	.088
Estimator C: $\beta = .5$, $\lambda_i \sim \mathrm{U}[-.1, .1]$									
300	200	.010	.043	.042	15.445	0	.012	.039	1.000
300	200	.050	.075	.071	12.111	.006	.031	.193	1.000
300	200	.500	.499	.493	7.288	.081	.080	.990	.977
300	200	.950	.916	.910	6.925	.104	.095	1.000	.176
300	200	1.000	.977	.977	6.254	.096	.083	1.000	.096
600	200	.010	.029	.029	6.923	0	0	.001	1.000
600	200	.050	.067	.064	8.117	.005	.015	.100	1.000
600	200	.500	.503	.496	5.299	.083	.091	.952	.939
600	200	.950	.942	.939	3.613	.079	.065	1.000	.120
600	200	1.000	.998	1.010	4.407	.094	.094	1.000	.094
300	400	.010	.035	.032	17.039	0	.013	.055	1.000
300	400	.050	.069	.069	12.498	.005	.047	.362	1.000
300	400	.500	.502	.502	7.504	.071	.081	1.000	1.000
300	400	.950	.927	.929	6.566	.106	.081	1.000	.201
300	400	1.000	.974	.973	6.023	.118	.095	1.000	.118
600	400	.010	.028	.027	8.160	0	.001	.001	1.000
600	400	.050	.067	.067	7.300	.004	.008	.151	1.000
600	400	.500	.504	.501	5.233	.077	.079	.998	.998
600	400	.950	.940	.935	3.952	.079	.078	1.000	.145
600	400	1.000	.996	.999	4.500	.082	.091	1.000	.082

$1, \ldots, T$,

$$V_{t,N} = \frac{1}{N}\sum_{i=1}^{N}(y_{it} - Y_{t,N})^2$$
$$= \theta V^1_{t,N_1} + (1-\theta)V^0_{t,N_0} + \theta(1-\theta)g^2_{t,N} \qquad \text{(A.1)}$$

where $g_{t,N} = (Y^1_{t,N} - Y^0_{t,N})$, N_0 is the number of units with $\alpha_i < 1$, N_1 is the number of units with $\alpha_i = 1$, and $N = N_1 + N_0$. Here $\theta = N_1/N$ is treated as a fixed parameter so that $N_1 = \theta N \to \infty$ as $N \to \infty$, and likewise for $N_0 = (1-\theta)N$. The observations are ordered such that the first N_1 units have $\alpha_i = 1$. The proof consists of showing that for fixed t and as $N \to \infty$, $\Delta V_{t,N} \xrightarrow{p} \Delta V_{t,\infty} = \theta \Delta V^1_{t,\infty}$.

We begin with $\Delta V^1_{t,N_1}$. Because $\Delta V^1_{t,N_1} = \Delta\,\mathrm{var}^1_{i,N_1}(\lambda_i) + \Delta s^1_{t,N_1}$ and $\mathrm{var}^1_{i,N_1}(\lambda_i) \xrightarrow{p} \mathrm{var}_{i,\infty}(\lambda_i)$ for all t regardless of α_i, by A2, $\Delta\,\mathrm{var}^1_{i,N_1}(\lambda_i) \xrightarrow{p} 0$. It remains to consider $\Delta s^1_{t,N_1}$. By direct calculations,

$$s^1_{t,N_1} = \frac{1}{N_1}\sum_{i=1}^{N_1} u^2_{it} - \left(\frac{1}{N_1}\sum_{i=1}^{N_1} u_{it}\right)^2$$
$$= \frac{1}{N_1}\sum_{i=1}^{N_1}(u^2_{it-1} + e^2_{it} + 2u_{it-1}e_{it})$$
$$- \left(\frac{1}{N_1}\sum_{i=1}^{N_1} u_{it-1} + e_{it}\right)^2$$

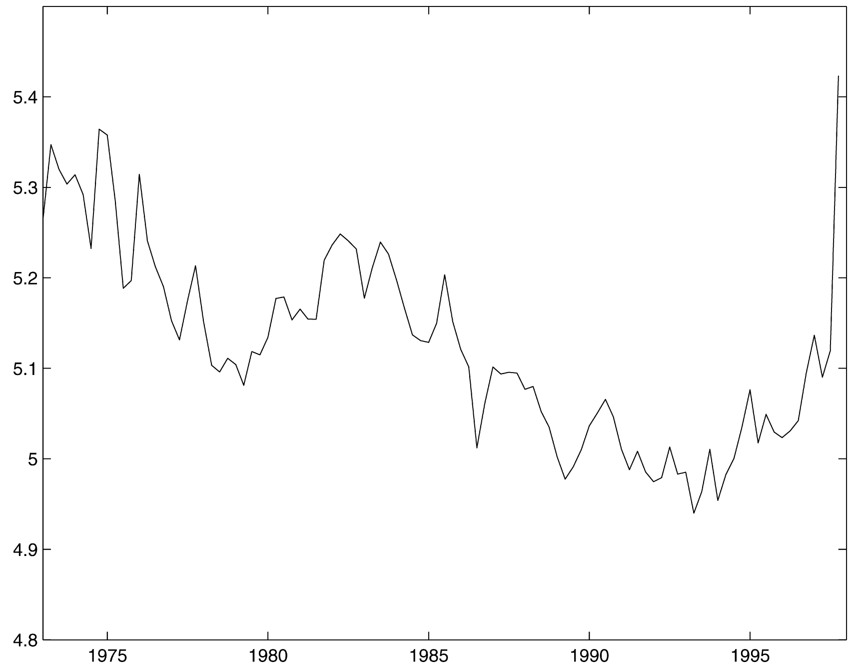

Figure 2. Cross-sectional variance of real exchange rates.

$$= \left[\frac{1}{N_1}\sum_{i=1}^{N_1} u_{it-1}^2 - \left(\frac{1}{N_1}\sum_{i=t}^{N_1} u_{it-1}\right)^2\right]$$
$$+ \left[\frac{1}{N_1}\sum_{i=1}^{N_1} e_{it}^2 - \left(\frac{1}{N_1}\sum_{i=1}^{N_1} e_{it}\right)^2\right]$$
$$+ 2\left[\frac{1}{N_1}\sum_{i=1}^{N_1} u_{it}e_{it} - \left(\frac{1}{N_1}\sum_{i=1}^{N_1} u_{it-1}\right)\left(\frac{1}{N_1}\sum_{i=1}^{N_1} e_{it}\right)\right]$$
$$\equiv s_{t-1,N_1}^1 + (\sigma_{N_1}^1)^2 + 2\,\mathrm{cov}_{i,N_1}^1(u_{it-1}, e_{it}).$$

Let $\eta_{t,N_1}^1 = \mathrm{cov}_{i,N_1}^1(u_{it}, e_{it})$, which is a sample covariance. For fixed t, $\eta_{t,N_1}^1 \xrightarrow{p} \eta_{t,\infty}^1 = E_i(u_{it-1}e_{it}) = 0$. Thus, as $N_1 \to \infty$,

$$\Delta s_{t,N_1}^1 = \left(\sigma_{N_1}^1\right)^2 + 2\eta_{t,N_1} \xrightarrow{p} \Delta s_{t,\infty}^1 = 1 = \Delta V_{t,\infty}^1.$$

It remains to consider the last two terms of (A.1). Now $s_{t,N_0}^0 \xrightarrow{p} s_{t,\infty}^0$ and $s_{t,\infty}^0 \approx 0$ for any $t > t^*$ such that $m_{t^*} \to 0$. Finally, $g_{t,N} = Y_{t,N_1}^1 - Y_{t,N_0}^0$. Because λ_i is independent α_i, the mean and variance of λ_i are the same in the two samples. This implies that $g_{t,N} = Y_{t,N_1}^1 - Y_{t,N_0}^0 \approx U_{t,N_1}^1 - U_{t,N_0}^0$. But $U_{t,N_0}^0 \xrightarrow{p}$

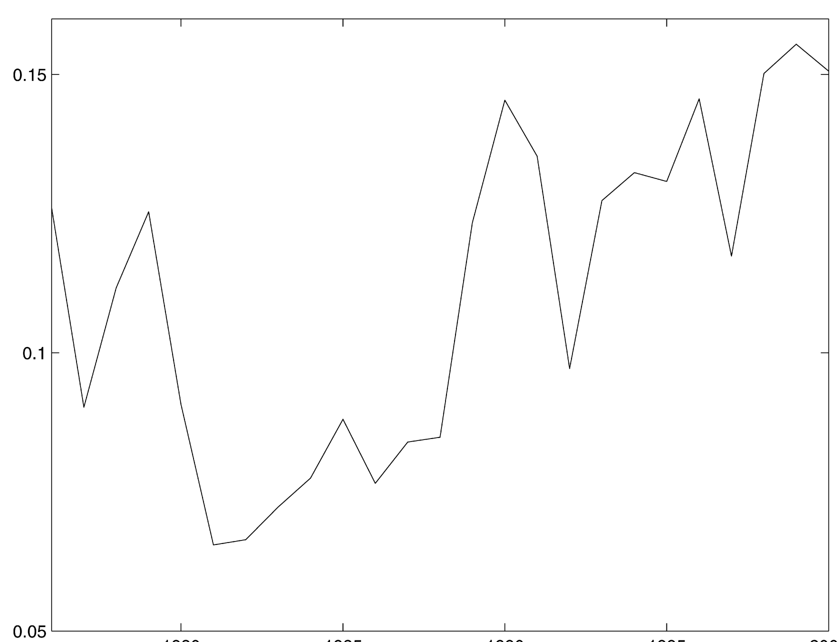

Figure 3. Cross-sectional variance of PSID earnings.

0 for $t > t^*$ and $U^1_{t,N_1} \to U^1_{0,\infty}$, which does not depend on t. Thus $g_{t,N} \to g_{t,\infty} = U^1_{0,\infty}$ and $V_{t,N} \xrightarrow{p} V_{t,\infty} = \theta V^1_{t,\infty} + c$. Because $c = \theta(1-\theta)U^1_{0,\infty}$ does not depend on t, $\Delta V_{t,\infty} = \theta$.

The following result is used in the proof of Theorem 1.

Lemma A.1. Suppose that $g_t = g_{t-1} + v_t$, $v_t \sim \mathrm{N}(0, \sigma_v^2)$. Then g_t^2 is a heteroscedastic random walk with a drift of σ_v^2 and conditional variance of $4\sigma_v^4 \cdot t - 2\sigma_v^4$.

By definition, $g_t^2 = g_{t-1}^2 + v_t^2 + 2g_{t-1}v_t$. To establish that g_t^2 is a unit root with drift and heteroscedastic error, we need to show that $z_t = v_t^2 + 2g_{t-1}v_t$ has a nonzero conditional mean and a trend component in the conditional variance. Let I_{t-1} be the past history of g_t. Because $v_t \sim \mathrm{N}(0, \sigma_v^2)$, $E(z_t|I_{t-1}) = \sigma_v^2$, which constitutes the drift of g_t^2. Furthermore, $E(z_t^2|I_{t-1}) = 3\sigma_v^4 + 4(t-1)\sigma_v^4$, so that $\mathrm{var}(z_t|I_{t-1}) = 4\sigma_v^4 \cdot t - 2\sigma_v^4$, which is the innovation variance of g_t^2.

Let $g_{t,N} = Y^1_{t,N_1} - Y^0_{t,N_0} = U^1_{t,N_1} - U^0_{t,N_0} + o_p(1)$. Thus $g^2_{t,N} = (U^1_{t,N_1})^2 + (U^0_{t,N_0})^2 - 2U^1_{t,N_1}U^0_{t,N_0} + o_p(1)$, which is dominated by $(U^1_{t,N_1})^2$. Because $\Delta U^1_{t,N_1} = \epsilon^1_{t,N_1}$ and ϵ^1_{t,N_1} is a sample average of e_{it}, ϵ^1_{t,N_1} is approximately normally distributed. Thus U^1_{t,N_1} is approximately a Gaussian random walk. By Lemma A.1, $\Delta(U^1_{t,N_1})^2$ has a mean of $(\sigma^1_{\epsilon,N_1})^2$ and variance of order $4t(\sigma^1_{\epsilon,N_1})^4$. But $(\sigma^1_{\epsilon,N_1})^2 = O_p(N_1^{-1})$. Thus for a given t, (U^1_{t,N_1}) has a drift and variance of order $1/N_1$ and t/N_1^2. Summing over t, $\frac{1}{T}\sum_{t=1}^T (g^2_{t,N})$ has mean and variance of order $1/N_1$ and $1/N_1^2$, which tend to 0 as $N_1 \to \infty$. We use this result in what follows.

Proof of Theorem 1

Let $\eta^1_{t,N_1} = \mathrm{cov}^1_{i,N_1}(u_{it-1}e_{it})$. The estimator is

$$\begin{aligned}\hat\theta &= \frac{1}{T}\sum_{t=1}^T \Delta V_{t,N}\\ &= \theta\frac{1}{T}\sum_{t=1}^T \Delta V^1_{t,N_1} + (1-\theta)\frac{1}{T}\sum_{t=1}^T V^0_{t,N_0}\\ &\quad + \theta(1-\theta)\frac{1}{T}\sum_{t=1}^T g^2_{t,N}.\end{aligned}$$

Because $\Delta V^0_{t,N_0}$ is negligible,

$$\begin{aligned}\hat\theta &= \theta\frac{1}{T}\sum_{t=1}^T \Delta V^1_{t,N_1} + \theta(1-\theta)\frac{1}{T}\sum_{t=1}^T g^2_{t,N} + o_p(1)\\ &= \theta\frac{1}{T}\sum_{t=1}^T\big[\Delta\,\mathrm{var}^1_{i,N}(\lambda_i) + (\sigma^1_{N_1})^2 + 2\,\mathrm{cov}^1_{i,N_1}(u_{it-1}e_{it})\big]\\ &\quad + \theta(1-\theta)\frac{1}{T}\sum_{t=1}^T g^2_{t,N} + o_p(1)\\ &= \theta\frac{1}{T}\sum_{t=1}^T\big[1 + 2\eta^1_{t,N_1}\big] + o_p(1),\end{aligned}$$

because $(\sigma^1_{N_1})^2 \xrightarrow{p} 1$ and $\Delta\,\mathrm{var}^1_{i,N}(\lambda_i) \xrightarrow{p} 0$, and using the implication of Lemma A.1. Multiplying by $\sqrt{N}$,

$$\sqrt{N}(\hat\theta - \theta) = \frac{\sqrt{N}}{T}\sum_{t=1}^T 2\theta\eta^1_{t,N_1} + o_p(1).$$

Now η^1_{t,N_1} is a cross-sectional sample covariance. It converges to $E_i(y_{it-1}e_{it}) = 0$ for all t with $N_1\,\mathrm{var}(\eta^1_{t,N_1}) \xrightarrow{p} (t-1)\sigma_e^4 = (t-1)$ because $\sigma_e^2 = 1$. Thus, as $T \to \infty$,

$$N_1\,\mathrm{var}\left(\frac{1}{T}\sum_{t=1}^T 2\theta\eta^1_{t,N_1}\right) = 4\theta^2\frac{1}{T^2}\sum_{t=1}^T t + o(1) \to 2\theta^2,$$

and because $\theta = N_1/N$, by definition,

$$N\,\mathrm{var}\left(\frac{1}{T}\sum_{t=1}^T 2\theta\eta^1_{t,N}\right) = \frac{N_1}{\theta}\,\mathrm{var}\left(\frac{1}{T}\sum_{t=1}^T 2\theta\eta^1_{t,N_1}\right) \xrightarrow{p} 2\theta.$$

Because the variance of η^1_{t,N_1} is linear in t, time averaging does not accelerate the rate of convergence. Thus we have $\sqrt{N}(\hat\theta - \theta) \xrightarrow{d} \mathrm{N}(0, 2\theta)$ as $N \to \infty$, and then $T \to \infty$.

Proof of Theorem 2

The DGP is $y_{it} = \lambda_i \cdot t + u_{it}$ with $u_{it} = \alpha_i u_{it-1} + e_{it}$, $e_{it} \sim \mathrm{iid}(0, \sigma_e^2)$, $\sigma_e^2 = 1$, and λ_i independent of α_i and e_{it} for all i and t. We again need to evaluate

$$V_{t,N} = \theta V^1_{t,N_1} + (1-\theta)V^0_{t,N_0} + \theta(1-\theta)g^2_{t,N}.$$

For this model,

$$V^1_{t,N_1} = s^1_{t,N_1} + \mathrm{var}^1_{i,N_1}(\lambda_i)t^2 + 2\,\mathrm{cov}^1_{i,N_1}(\lambda_i t, u_{it}).$$

First differencing and substituting in $\Delta s^1_{t,N_1}$, from Theorem 1, we have

$$\begin{aligned}\Delta V^1_{t,N_1} &= \Delta s^1_{t,N_1} + \mathrm{var}^1_{i,N_1}(\lambda_i)\Delta t^2 + 2\Delta\,\mathrm{cov}^1_{i,N_1}(\lambda_i t, u_{it})\\ &= (\sigma^1_{N_1})^2 + \mathrm{var}^1_{i,N_1}(\lambda_i)\Delta t^2 + 2\,\mathrm{cov}^1_{i,N_1}(u_{it-1}, e_{it})\\ &\quad + 2\Delta\,\mathrm{cov}^1_{i,N_1}(\lambda_i t, u_{it})\\ &= 1 + \mathrm{var}^1_{i,N_1}(\lambda_i)\Delta t^2 + \eta^1_{t,N_1} + o_p(1), \qquad \text{(A.2)}\end{aligned}$$

where $\eta^1_{t,N_1} = 2\,\mathrm{cov}^1_{i,N_1}(u_{it-1}, e_{it}) + 2\Delta\,\mathrm{cov}^1_{t,N_1}(\lambda_i t, u_{it})$ and we have used the fact that $(\sigma^1_{i,N_1})^2 \xrightarrow{p} 1$. Similarly, $V^0_{t,N_0} = s^0_{t,N_0} + \mathrm{var}^0_{i,N_0}(\lambda_i)t^2 + 2\,\mathrm{cov}^0_{i,N_0}(\lambda_i t, u_{it})$. First differencing gives

$$\Delta V^0_{t,N_0} = \mathrm{var}^0_{i,N_0}(\lambda_i)\Delta t^2 + \eta^0_{t,N_0} + o_p(1), \qquad \text{(A.3)}$$

where $\eta^0_{t,N_0} = 2\Delta\,\mathrm{cov}^0_{i,N_0}(\lambda_i, u_{it})$, and we have used the fact that $\Delta s^0_{t,N_0}$ is approximately 0.

By assumption, the distribution of λ_i does not depend on α_i. This implies that $g_{t,N} \approx U^1_{t,N_1} - U^0_{t,N_0}$, and the variance of λ_i in the two samples can be replaced by the common $\mathrm{var}_{i,\infty}(\lambda_i)$ for large N. Then (A.2) and (A.3), can be used to rewrite (A.1) as

$$\begin{aligned}\Delta V_{t,N} &= \theta + \mathrm{var}_{i,N}(\lambda_i)\Delta t^2 + \eta_{t,N} + o_p(1)\\ &= \theta + \beta\Delta t^2 + \eta_{t,N} + o_p(1), \qquad \text{(A.4)}\end{aligned}$$

where $\beta = \mathrm{var}_{i,\infty}(\lambda_i)$ and $\eta_{t,N} = \theta\eta^1_{t,N_1} + (1-\theta)\eta^0_{t,N_0}$ is dominated by $2\theta\Delta\,\mathrm{cov}^1_{i,N_1}(\lambda_i t, u_{it})$. To derive its properties, first note that $\text{cov}^1_{i,N_1}(\lambda_i t, u_{it})$ is a sample moment, and a central limit theorem applies. In particular, $\sqrt{N_1}\,\text{cov}^1_{i,N_1}(\lambda_i t, u_{it}) \xrightarrow{d} \text{N}(0, \text{var}_{i,\infty}(\lambda_i)t^3\sigma_e^2)$. Because e_{it} is serially uncorrelated by assumption and $\sigma_e^2 = 1$, $\text{var}(\Delta\sqrt{N_1}\,\text{cov}^1_{i,N_1}(\lambda_i t, u_{it})) = \text{var}_{i,\infty}(\lambda_i)t^2 + o(t)$. Thus, for fixed t, $\sqrt{N}2\theta\,\Delta\,\text{cov}^1_{i,N_1}(\lambda_i t, u_{it}) \xrightarrow{d} \text{N}(0, 4\theta\,\text{var}_{i,\infty}(\lambda_i)t^2 + o(t))$ and

$$\sqrt{N}\eta_{t,N} \xrightarrow{d} \text{N}\big(0, 4\theta\,\text{var}_{i,\infty}(\lambda_i)t^2 + o(t)\big). \tag{A.5}$$

Consider a least squares regression of $\Delta V_{t,N}$ on $z_t = (1, \Delta t^2)$, and let $(\hat{\theta}, \hat{\beta})'$ be the coefficient estimates. Define $D_T = \text{diag}(T^{1/2}T^{3/2})$. Then

$$\frac{\sqrt{N}}{T}D_T\left(\begin{pmatrix}\hat{\theta}\\ \hat{\beta}\end{pmatrix} - \begin{pmatrix}\theta\\ \beta\end{pmatrix}\right) = \left[D_T^{-1}\sum_{t=1}^{T} z_t z_t' D_T^{-1}\right]^{-1}\begin{pmatrix}T^{-3/2}\\ T^{-5/2}\end{pmatrix}\sum_{t=1}^{T}\sqrt{N}z_t\eta_{tN},$$

noting that

$$D_T^{-1}\sum_{t=1}^{T} z_t z_t' D_T^{-1} \xrightarrow{p} \begin{pmatrix}1 & 1\\ 1 & 4/3\end{pmatrix} = Q.$$

Using (A.5),

$$\text{var}\left(\frac{\sqrt{N}}{T^{3/2}}\sum_{t=1}^{T}\eta_{t,N}\right) = 4\theta T^{-3}\sum_{t=1}^{T}t^2 = 4\theta\,\text{var}_{i,\infty}(\lambda_i)/3$$

and

$$\text{var}\left(\frac{\sqrt{N}}{T^{5/2}}\sum_{t=1}^{T}\Delta t^2\cdot\eta_{tN}\right) = 16\theta T^{-5}\sum_{t=1}^{T}t^4 = 16\theta\,\text{var}_{i,\infty}(\lambda_i)/5.$$

Thus

$$\begin{pmatrix}T^{-3/2}\\ T^{-5/2}\end{pmatrix}\sum_{t=1}^{T}z_t\sqrt{N}\eta_{t,N} \xrightarrow{d} \text{N}(0, Z),$$

where

$$Z = \theta\,\text{var}_{i,\infty}(\lambda_i)\begin{pmatrix}4/3 & 2\\ 2 & 16/5\end{pmatrix}.$$

Putting the results together,

$$\frac{\sqrt{N}}{T}D_T\left(\begin{pmatrix}\hat{\theta}\\ \hat{\beta}\end{pmatrix} - \begin{pmatrix}\theta\\ \beta\end{pmatrix}\right) \xrightarrow{d} \text{N}(0, Q^{-1}ZQ^{-1}),$$

where

$$Q^{-1}ZQ^{-1} = \theta\,\text{var}_{i,\infty}(\lambda_i)\begin{pmatrix}32/15 & -14/15\\ -14/15 & 24/5\end{pmatrix}.$$

Thus $\frac{\sqrt{N}}{\sqrt{T}}(\hat{\theta} - \theta) \sim \text{N}(0, 32\theta\,\text{var}_{i,\infty}(\lambda_i)/15)$ and $\sqrt{NT}(\hat{\beta} - \beta) \xrightarrow{d} \text{N}(0, 24/5)$.

[Received December 2005. Revised October 2006.]

REFERENCES

Alvarez, J., Browning, M., and Ejrnaes, M. (2001), "Modelling Income Processes With Lots of Heterogeneity," University of Copenhagen, **???**.

Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221.

——— (2004), "A PANIC Attack on Unit Roots and Cointegration," *Econometrica*, 72, 1127–1177.

Baker, M. (1997), "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics*, 15, 338–375.

Baltagi, B., and Kao, C. (2001), "Nonstationary Panels, Cointegration in Panels, and Dynamic Panels," Center for Policy Research, Syracuse University.

Campbell, J., and Deaton, A. (1989), "Why Is Consumption so Smooth," *Review of Economic Studies*, 56, 357–374.

Carroll, C. D., and Samwick, A. (1997), "How Important Is Precautionary Wealth," *Journal of Monetary Economics*, 40, 41–72.

Chang, Y. (2002), "Nonlinear IV Unit Root Tests in Panels With Cross-Section Dependency," *Journal of Econometrics*, 110, 261–292.

Chang, Y., and Song, W. (2002), "Panel Unit Root Tests in the Presence of Cross-Section Heterogeneity," unpublished manuscript, Rice University, **???**.

Choi, C. Y. (2004), "Searching for Evidence of Long-Run PPP From a Post–Bretton Woods Panel: Separating the Wheat From the Chaff," *Journal of International Money and Finance*, 23, 1159–1186.

Deaton, A. S., and Paxson, C. (1994), "Intertemporal Choice and Inequality," *Journal of Political Economy*, CII, 437–467.

Granger, C. (1980), "Long-Memory Relationships and the Aggregation of Dynamic Models," *Journal of Econometrics*, 14, 227–238.

Granger, C. W., and Morris, M. J. (1976), "Time Series Modelling and Interpretation," *Journal of the Royal Statistical Association*, Ser. A, 139, 246–257.

Guvenen, F. (2005), "An Empirical Investigation of Labor Income Process," University of Rochester, **???**.

Hadri, K. (2000), "Testing for Stationarity in Heterogeneous Panel Data," *Econometrics Journal*, 3, 148–161.

Hall, R. (1978), "Stochastic Implications of the Life Cyle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86, 971–987.

Harris, D., Leybourne, S., and McCabe, B. (2004), "Panel Stationarity Tests for Purchasing Power Parity With Cross-Section Dependence," mimeo, University of Nottingham, **???**.

Im, K., Pesaran, M., and Shin, Y. (2003), "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, 115, 53–74.

Kapetanios, G. (2003), "Determining the Stationary Properties of Individual Series in Panel Datasets," mimeo, Queen Mary, University of London, **???**.

Levin, A., Lin, C. F., and Chu, J. (2002), "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties," *Journal of Econometrics*, 98, 1–24.

Lewbel, A. (1994), "Aggregation and Simple Dynamics," *American Economic Review*, 59, 635–642.

Lucas, R. E. (2003), "Macroeconomic Priorities," *American Economic Review*, 93, 1–14.

Ludvigson, S., and Michaelides, A. (2001), "Can Buffer Stock Saving Explain Consumption Excesses," *American Economic Review*, 91, 631–647.

MaCurdy, T. E. (1981), "The Use of Time Series Processes to Model the Error Structure of Earnings in Longitudinal Data Analysis," *Journal of Econometrics*, 18, 83–114.

Maddala, G. S., and Wu, S. (1999), "A Comparative Study of Unit Root Tests With Panel Data and a New Simple Test," *Oxford Bulletin of Economics and Statistics*, **??**, 631–652.

Moon, R., and Perron, B. (2004), "Testing for a Unit Root in Panels With Dynamic Factors," *Journal of Econometrics*, 122, 81–126.

Moon, R., Perron, B., and Phillips, P. (2005), "Incidental Trends and the Power of Panel Unit Root Tests," mimeo, **???**.

O'Connell, P. (1998), "The Overvaluation of Purchasing Power Parity," *Journal of International Economics*, 44, 1–19.

Pesaran, H. (2005), "A Simple Panel Unit Root Test in the Presence of Cross-Section Dependence," mimeo, Cambridge University, **???**.

Quah, D. (1994), "Exploiting Cross-Section Variations for Unit Root Inference in Dynamic Panels," *Economics Letters*, 44, 1–9.

Reis, R. (2005), "The Time Series Properties of Aggregate Consumption: Implications for the Costs of Fluctuations," Working Paper 11297, National Bureau of Economic Research.

Storesletten, K., Telmer, C., and Yaron, A. (2004), "Consumption and Risk Sharing Over the Life Cycle," *Journal of Monetary Economics*, 51, 609–633.

Zaffaroni, P. (2004), "Contemporaneous Aggregation of Linear Dynamic Models in Large Economies," *Journal of Econometrics*, 120, 75–102.