Testing for Unit Roots in Panel Data:

An Exploration Using Real and Simulated Data

Bronwyn H. HALL (*UC Berkeley; IFS; and NBER)*

Jacques MAIRESSE (*CREST-INSEE; EHESS; and NBER)*

Version of December 2002

ABSTRACT

This paper presents the results of a Monte Carlo study that compares the small sample performance of various unit root tests in short panels using simulated data that mimic the time series and cross sectional properties of commonly used firm level variables. Our conclusion is that in the presence of firm-level heteroskedasticity two methods are preferred, depending on the nature of the preferred alternative: the simplest method based on the ordinary least squares regression of the variable under consideration on its own lag and a version with a more complex alternative hypothesis suggested by Im, Pesaran, and Shin. The paper also reports the results of using these tests for sales, employment, investment, R&D and cash-flow in three panels of large French, Japanese and US manufacturing firms. In most cases our data reject the presence of a unit root in favor of a first order autoregressive model with a very high autoregressive coefficient, so high that fixed effects are of negligible additional importance in the model.

We are grateful to Karim Abadir, Stephen Bond, Andrew Chesher, Hugo Kruiniger and Neil Shepard for helpful comments on earlier versions of this paper. The paper has also benefited greatly from remarks and useful suggestions from James Stock and two referees.

Keywords: panel data, unit roots, investment, R&D, firm data, international comparisons

Testing for Unit Roots in Panel Data:

An Exploration Using Real and Simulated Data

Bronwyn H. Hall and Jacques Mairesse

1 Introduction

In this paper, we investigate the properties of several unit root tests in short panel data models using simulated data that look like the data typically encountered in studies on firm behavior. This investigation arose from a previous exploration of a simple question: could we find a simple parsimonious model that accounts for the time series properties of key observable variables characterizing the behavior of individual firms: sales, employment, investment, R&D, and cash flow or profits in France, Japan and the United states.[1] We started from a fairly general autoregressive model in the spirit of Holtz-Eakin, Newey, and Rosen (1988) where the heterogeneity across firms is accounted by an individual-specific intercept or firm fixed effect and a firm-specific variance of the random disturbance. We proceeded in estimation by using the GMM methodology. Our estimates, however, were both imprecise and suggestive of the presence of finite sample bias.[2] We therefore investigated the properties of our estimator using two very simple but quite different data generating processes that approximated our data fairly well (random walk vs. fixed effect with no autoregression) and concluded that the first step in

---

[1] This exploration (see Hall and Mairesse, 2001) was itself a follow up on Hall, Mairesse, Branstetter and Crepon (1999).

[2] These estimates are documented in Hall and Mairesse (2001).

constructing a parsimonious univariate model for such data should probably be a test for stationarity, because the presence of a unit root will invalidate the commonly used GMM specification.[3]

Testing for stationarity in panel data models is also per se a matter of interest and it can be more directly motivated. It seems fairly intuitive that, within the general class of models where heterogeneity is restricted to an individual fixed effect, the times series behavior of an individual variable should often be well approximated either as an autoregressive process with a small positive coefficient and large fixed effects or as an autoregressive process with a near-unit root and negligible individual fixed effects. Both alternatives can be nested in a single model, in which the test of the former against the latter is a panel data unit root test. One expects, however, that such test might not perform well in a short panel, owing in particular to the problem of unobserved initial conditions and incidental parameter estimation. Trying to assess the properties of the available tests in a realistic setting is therefore of practical importance.

In recent years the econometrics literature has proposed a number of tests for unit roots in panel data. We confine our attention to the six of them that are valid when the number T of time periods (years in our case) is small and the number N of individuals (firms in our case) is large, that is that are consistent when T is fixed and $N \rightarrow \infty$.[4] We describe these six tests in detail in section 2 of the paper. They vary in several dimensions: 1) the degree of heterogeneity across

[3] For a more complete discussion of the problems with GMM estimation when the data are nearly nonstationary, see Blundell and Bond (1998).

[4] We have omitted all the tests that rely on the $T \rightarrow \infty$ assumption for validity, because such tests are inappropriate for the usual data on firms. See Quah (1994) and Levin and Lin (1993) for examples of these kinds of tests. The six tests we consider are those that are appropriate for the fixed T, large N case, and were known to us as of the time of writing (2001).

individuals that is allowed for; 2) serial correlation, heteroskedasticity, and robustness to non-normality; and 3) whether they follow the Wald, likelihood ratio, or Lagrange Multiplier (LM) testing principles (see Table 1). All of them treat the presence of a unit root, implying nonstationarity, as the null hypothesis, and the absence of unit root, or stationarity, as the alternative hypothesis.

The first test we will consider is based on CMLE (conditional maximum likelihood estimation) and is the most restrictive in terms of the assumptions necessary for validity. Then comes the HT (Harris-Tzavalis) test, which is based on bias-adjusted least squares dummy variable (LSDV) or within estimation and therefore allows non-normality but not heteroskedasticity.[5] We also consider a version of CMLE suggested by Kruiniger (1999b) which allows for heteroskedasticity across units and time separately and is slightly more general than H-T. The next test, which we will label OLS, allows for heteroskedasticity and non-normality, and takes a very different approach by viewing the panel data regression as a system of T year regressions. It is based on the fact that ordinary least squares is a consistent estimator for the model with a lagged dependent variable and no fixed effects.[6] The IPS (Im-Pesaran-Shin) test is the last one we consider. It also takes a different approach from the foregoing, in that it views the panel data regression as a system of N individual regressions and is based on the combination of independent Dickey-Fuller tests for these N regressions. Besides allowing heteroskedasticity,

[5] In fact, if we interpret the CMLE as a quasi-likelihood method, using it to construct a test is no more or less restrictive than the HT test. Both require homoskedasticity but not normality, and in principle, either one could be modified to yield a test robust to heteroskedasticity, as we do in the case of CML estimation.

[6] This test is implicit in early work by Macurdy (1985). It was suggested to us by Steve Bond (see Bond, Nauges, and Windmeijer, 2002).

serial correlation, and non-normality, this test also allows for heterogeneity of trends and of the lag coefficient under the alternative hypothesis of no unit root.

In the paper, we present the results of a Monte Carlo study that compares the small sample performance of these tests using simulated data mimicking the time series and cross sectional properties of the firm sales, employment, investment, R&D, and cash flow variables in three panel data samples for French, Japanese and US manufacturing firms. The design and calibration of the simulations, which are based on the most persistent of these series, the R&D in the US, is explained in section 3. The results of the eight different Monte Carlo experiments are presented in section 4. Our tentative conclusion is that the simplest method, the OLS test based on the ordinary least squares regression of the variable considered on its own lag, may actually be the best for micro-data panels similar to ours. The OLS estimator is unbiased under the null of a unit root (when the fixed effect vanishes) and its estimated standard error can easily be corrected for both serial correlation and heteroskedasticity of the disturbances.

In section 5, we also report the results of using all six tests for the five variables in our three samples. In most cases our data reject nonstationarity in favor of stationarity, but with a very high autoregressive coefficient, so high that it is not necessary to include fixed effects in the model. We very briefly conclude in section 6.

2 Testing for Unit Roots in Panel Data: an Overview

The most general form of the model considered in this paper can be written as follows:

$$
\begin{aligned}
y_{it} &= \alpha_i + \delta t + u_{it} \qquad t = 1,\ldots,T; i = 1,\ldots,N \\
u_{it} &= \rho u_{i,t-1} + \varepsilon_{it} \qquad \varepsilon_{\mathrm{it}} \sim [0, \sigma_t^2 \sigma_\varepsilon^2(i)] \\
&\Rightarrow y_{it} = (1-\rho)\alpha_i + (1-\rho)\delta t + \rho y_{i,t-1} + \varepsilon_{it} \qquad |\rho| < 1 \\
\text{or} \quad y_{it} &= y_{i,t-1} + \delta + \varepsilon_{it} \qquad \varepsilon_{\mathrm{it}} \sim [0, \sigma_t^2 \sigma_\varepsilon^2(i)] \qquad \rho = 1
\end{aligned} \tag{1}
$$

That is, we consider the possibility of either an autoregressive model with a fixed effect or a random walk with drift. In both cases, we allow for individual and time-varying heteroskedasticity of a proportional form in addition. In some cases, the various tests described below are valid only for more restrictive versions of the model in equation (1).

Table 1 provides a schematic view of the various unit root tests we consider and the assumptions under which they are valid. All the tests assume conditional independence across the units, and all except the OLS test allow for individual-specific means in estimation. The CMLE, IPS, and OLS tests can potentially accommodate a flexible correlation structure among the disturbances, as long as it is the same for all units. However, in this paper we have assumed throughout that the disturbances are serially uncorrelated (in the presence of the lagged dependent variable) and constructed our tests accordingly. With the possible exception of the investment and cash flow series, this assumption is satisfied by our real data series.[12] In the text that follows, we indicate how to modify the tests to accommodate serial correlation.

---

[12] The autocorrelograms of the level and first-differenced series are shown in Figures 1 and 2 of Appendix B. The autocorrelation of the first differences at lag one is less than 0.25 for most of the series. For our "model" series, U.S. log R&D, the autocorrelation is -0.04.

The estimators associated with these tests allow for various degrees of heterogeneity in addition to the individual-specific means. In particular, all of them except the Harris-Tzavalis test and the homoskedastic version of the CMLE test allow the variance of the disturbances to be different for each unit.[13] The IPS tests, which are based on N individual regressions, allow both the trend and the serial correlation coefficient to vary across the units under the alternative, in addition to the mean and variance. We now describe the tests in somewhat more detail.

2.1 Maximum Likelihood Methods with Homoskedastic Errors

Lancaster and Lindenhovius (1996), Kruiniger (1999b), and Binder, Hsiao, and Pesaran (2000) have independently pointed out that the conditional maximum likelihood estimate of the linear model with individual effects and a lagged dependent variable is well-identified and consistent even when there is a unit root, that is, even when the coefficient is one, although this value is on the boundary of the parameter space. This fact can be used to construct a likelihood ratio test of $\rho = 1$ versus $\rho < 1$.[14]

The model to be estimated is the one given in equation (1), but with homoskedastic disturbances and without the time trend:[15]

---

[13] The H-T test could probably be modified to accommodate heteroskedasticity also, but the version we use here does not.

[14] The consistency result is of considerable interest in its own right because the corresponding least squares (LSDV) estimator is neither consistent (as $N \rightarrow \infty$, T fixed) nor unbiased when $\rho = 1$. Appendix C contains a table of results for the OLS-levels, LSDV, and first-differenced OLS and IV for our simulated data. Except for level estimates of the models with no effects, the estimates are very far away from the true values.

[15] For simplicity of presentation, we omit the overall time trend in the presentation that follows. In practice, we removed year-specific means from the data before estimation.

$$
\begin{aligned}
&y_{it} = \alpha_i + u_{it} \qquad && u_{it} = \rho u_{i,t-1} + \varepsilon_{it} \qquad && |\rho| < 1 \\
&y_{it} = y_{i,t-1} + \varepsilon_{it} && && \rho = 1 \\
&\varepsilon_{\text{it}} \sim N[0, \sigma_\varepsilon^2] && &&
\end{aligned}
\tag{2}
$$

The null hypothesis is $\rho = 1$ and therefore no fixed effects. If we denote the vector of T observations for an individual as $y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})$ and similarly for u_i and $y_{i,-1}$, we can write this model in vector form as

$$
\begin{aligned}
&y_i = \alpha_i \iota + u_i \\
&E[u_i u_i'] = \sigma_\varepsilon^2 V_\rho = \frac{\sigma_\varepsilon^2}{1-\rho^2}
\begin{bmatrix}
1 & \rho & \ldots & \rho^{T-1} \\
\rho & 1 & \ldots & . \\
. & . & \ldots & . \\
\rho^{T-1} & \rho^{T-2} & \ldots & 1
\end{bmatrix}
\end{aligned}
\tag{3}
$$

or, in differenced form,

$$
Dy_i = Du_i \tag{4}
$$

Given normal disturbances, Dy_i has the joint normal distribution with mean zero and variance-covariance matrix $\Sigma = \sigma_\varepsilon^2 DV\rho D' = \sigma_\varepsilon^2 \Phi$ and the joint log likelihood for this model is the following:[16]

$$
\begin{aligned}
\log L(\rho, \sigma^2 '\{y_{it}\}) &= -\frac{N(T-1)}{2}\log(2\pi) - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(Dy_i)'\Sigma^{-1}Dy_i \\
&= -\frac{N(T-1)}{2}\log(2\pi) - \frac{N(T-1)}{2}\log\sigma^2 - \frac{N}{2}\log|\Phi| - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(Dy_i)'\Phi^{-1}Dy_i
\end{aligned}
\tag{5}
$$

---

[16] Higher order serial correlation of the disturbances can be allowed for by assuming that u_{it} follows an autoregressive model with a unit root and an order $p<T$ and deriving the appropriate V_ρ matrix that corresponds to this model.

Kruiniger (1999b) gives conditions under which maximizing this likelihood over the parameter space $(\rho, \sigma_\varepsilon^2) \in (-1, 1] \times (0, \infty)$ will yield consistent estimates.[17] Under those conditions, a conventional t-test for $\rho = 1$ is a test for a unit root. Alternatively, one could construct a likelihood ratio test by comparing the likelihood evaluated at its unconstrained maximum with the likelihood evaluated at $\rho = 1$.[18]

2.2 Maximum Likelihood Methods with Heteroskedastic Errors

A common feature of data on firms, even in logarithms, is that the variances of the errors vary across firms, which implies that estimation using methods assuming homoskedasticity is likely to produce wrong standard errors, at the least.[19] Consider the following variation of (2), which omits the trend:

---

[17] Basically, he requires stationarity if $\rho < 1$ and boundedness of the initial condition if $\rho = 1$. Also note that one cannot evaluate this likelihood as written if $\rho = 1$. See Kruiniger for details of the form of the likelihood when there is a unit root; that version collapses to the random walk model under that condition.

[18] Lancaster and Lindenhovius (1996) took a slightly different approach, using the same model and likelihood, but considering the Bayesian estimator with a flat prior on the effects (which drops out due to the differencing) and a prior of $1/\sigma$ for σ. This yields the joint marginal posterior density

$$p(\rho, \sigma^2 \mid \{y_{it}\}) = -\frac{N(T-1)}{2}\log(2\pi) - \log\sigma - \frac{N}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{N}(Dy_i)'\Sigma^{-1}Dy_i$$

The mode of this density is consistent for ρ and σ as $N\to\infty$. They do not consider the case $\rho = 1$. In practice, we found that evaluating the mode of this posterior gave essentially the same answer as the CMLE for samples of our size, so we do not report simulation results for this test.

[19] In fact, this is one of the several reasons why researchers often prefer methods based on the GMM methodology.

$$\begin{aligned} y_{it} &= \alpha_i + u_{it} && i = 1,\ldots,N; t = 1,\ldots,T \\ u_{it} &= \rho u_{i,t-1} + \varepsilon_{it} && \varepsilon_{it} \sim i.i.d.N(0,\sigma_i^2\sigma_t^2) \end{aligned} \quad (6)$$

At first glance, it might appear that estimation of such a model using maximum likelihood methods would lead to an incidental parameter problem due to the fact that the number of firm level parameters σ_i^2 grows with the sample size *N*. However, Kruiniger (1999b) shows that maximum likelihood estimation of the structural parameters (ρ, σ_t^2, $t=1,.., T$) of this model is consistent. The likelihood function for this model is given by

$$\log L(\rho,\{\sigma_t^2\};\{y_{it}\}) = -\frac{(T-1)}{2}\log(2\pi) - \frac{(T-1)}{2}\sum_{i=1}^{N}\log(\sigma_i^2) - \frac{N}{2}\log|\Phi| - \sum_{i=1}^{N}\frac{1}{2\sigma_i^2}(Dy_i)'\Phi^{-1}Dy_i \quad (7)$$

where

$$\Phi = DPV_\rho PD' \qquad \text{and} \qquad P = diag(\sigma_t) \quad (8)$$

Thus Φ depends only on the structural parameters ρ and $\{\sigma_t^2\}$. Given values for these parameters, it is clear that the maximum likelihood estimate of the individual-specific variances has the usual form:

$$\widehat{\sigma_i^2} = \frac{1}{T-1}tr(\Phi^{-1}Z_i) \qquad \text{where} \qquad Z_i = Dy_i(Dy_i)' \quad (9)$$

We use this fact to concentrate the $\widehat{\sigma_i^2}$, $i=1,\ldots,N$ out of the likelihood function, which greatly simplifies estimation. See Appendix A for details of the estimation procedure.

2.3 Harris-Tzavalis Test

The test for unit roots in panel data proposed by Harris and Tzavalis (1999) begins with the observation that the "Nickell" bias in the estimated coefficient of the lagged endogenous variable using LSDV (within) estimation is of known magnitude under some simple assumptions about the data generating process. Using this fact, one can compute bias adjustments to both the estimated coefficient and its standard error analytically and use the corrected estimates to construct a test of known size for a unit root.

H-T consider the model in equation (2) and show that under the null hypothesis that ρ = 1, the least squares dummy variable estimator has a limiting normal distribution of the following form:

$$\sqrt{N}(\rho - 1 - B_2) \rightarrow N(0, C_2) \tag{10}$$

where $B_2 = -3/(T+1)$ and $C_2 = 3(17T^2 - 20T + 17)/[5(T-1)(T+1)^3]$. Using this fact, it is straightforward to base a t-test on the estimated ρ, standardized by its mean and variance. Like the CMLE test, this test requires homoskedasticity and no serial correlation in the disturbances, although because it is based on a least squares estimator, it does not require normality.[20]

2.4 OLS - pooled estimation under the null

Bond, Nauges, and Windmeijer (2001) suggest that a test based on the model estimated under the null of a unit root (that is, where OLS can be used because there are no "fixed effects") may have more power when the true ρ is near unity. The advantage of such a test is that it does not require bias adjustment and it is easy to allow for heteroskedasticity by using a seemingly unrelated regression framework with each year being an equation.[21] Because there are no

[20] Neither normality nor homoskedasticity are required for the test based on the CMLE to be consistent either, although if these assumptions fail the conventional standard error estimates will be inconsistent and a "sandwich" estimator should be used.

[21] As in the well-known Dickey-Fuller test, if the disturbances are serially correlated, it will be necessary to include enough lagged values of the differenced y in the regression to render the disturbances uncorrelated in order to achieve consistency of the estimator.

incidental parameters under the null, asymptotics in the *N* dimension are straightforward and the test relies on those.

We base our OLS test on the following model:

$$\begin{aligned} y_{it} &= \delta_t + \rho y_{i,t-1} + \varepsilon_{it} \qquad i = 1,\ldots,N; t = 1,\ldots,T \\ E[\varepsilon_i \varepsilon_i '] &= \Omega \end{aligned} \tag{11}$$

where $\varepsilon_i = (\varepsilon_{i1}\ \varepsilon_{i2},\ \ldots,\ \varepsilon_{iT})$. The method of estimation is seemingly unrelated regression with a weighting matrix based on the first stage estimate of Ω.[22] Although the estimation method assumes homoskedasticity, we report standard errors that are robust to heteroskedasticity across the firms.

2.5 The IPS Method

Recent work by Im, Pesaran, and Shin (1997, hereafter IPS) suggests another approach for testing for unit roots, one that allows for more heterogeneity of behavior than that allowed for by the conditional maximum likelihood or least squares dummy variable approach. They assume a heterogeneous version of the model in equation (1):

$$y_{it} = (1-\rho_i)\alpha_i + \rho_i y_{i,t-1} + \varepsilon_{it} \qquad i = 1,\ldots,N; t = 1,\ldots,T \tag{12}$$

where initial values y_{i0} are given, and they test for the null hypothesis that ρ_i is unity for all observations versus an alternative that some of the ρ_is are less than one. Under the null, there is no fixed effect, while under the alternative, each fixed effect is equal to $(1-\rho_i)\ \alpha_i$. They propose tests based on the average over the individual units of a Lagrange-multiplier test of the

---

[22] As discussed earlier, we have assumed a diagonal form for Ω. If the εs are serially correlated within individuals, lagged values of the differenced y's should be added to the model until the residuals are approximately uncorrelated, as in the augmented Dickey-Fuller test.

hypothesis that $\rho_i = 1$ as well as tests based on the average of the augmented Dickey-Fuller statistics, which they find to have somewhat better finite sample properties than the L-M test.

As in Dickey and Fuller's original work, IPS also propose tests based on a model with a deterministic trend:

$$y_{it} = (1-\rho_i)\alpha_i + (1-\rho_i)\delta_i t + \rho_i y_{i,t-1} + \varepsilon_{it} \qquad i = 1,...,N; t = 1,...,T \tag{13}$$

We will use both these tests for our data, since there is reason to believe that trends do exist in the real series. Note that an important difference between these models and the models considered in the previous sections is that both the lag coefficient and the trend coefficient are allowed to differ across firms under the alternative hypothesis of stationarity.

When we applied these tests to our simulated data, we found that allowing the data to choose the length of augmenting lag p invariably yielded a p of either 2 or 3, even though the data were in all cases generated from models where $p = 0$ was appropriate.[23] Because of this fact and the fact that the table of critical values supplied by IPS breaks down for the case where the number of observations is around 10 and the length of the augmenting lag is greater than zero, we chose to focus on the tests where $p = 0$ is imposed. This makes our IPS test comparable to the others reported in this paper, which do not allow for serial correlation in the disturbances.

[23] Previous versions of this paper reported the results of tests using $p = 2$ and/or $p = 3$ on our simulated data and concluded that they had low power and were inaccurate even when we increased the number of time series observations per firm to 20, especially when $p = 3$ was the "optimal" choice of augmenting lag.

3 Design and Calibration of Simulations

In Section 5 of the paper we apply the panel data unit root tests to five firm-level variables drawn from three countries: employment, sales, cash flow, investment and R&D in the France, Japan and the United States. In our previous explorations using these data, we found that the process which describes each of the variables is more similar across countries than across variables, and that the variables can be clearly ranked by their long run "persistence": sales, employment, and R&D on the one hand versus cash flow and investment on the other. The behavior of the latter variables most resembled that of a stationary process.

Figures 1 and 2 in Appendix B display the autocorrelograms of the levels and first differences of our series for the 3 countries. These confirm the high autocorrelation in levels and the low autocorrelation in differences that characterize these data. They also show that the series most likely to exhibit the properties of a random walk is the log R&D series for all three countries, which has essentially zero autocorrelation at all lags in first differences. Therefore, we chose to investigate the performance of these tests on simulated data calibrated to match the time series and cross-sectional characteristics of the log R&D series for the United States.[24]

The general form of model or data generation process (DGP) that we use in our simulations is the model in equation (1) with and without heteroskedasticity across individuals and no heteroskedasticity over time.[25] We considered eight cases: the two extreme cases of a

[24] The details of the construction of our datasets and the results of GMM estimation using these data are given in Hall et al (1999) and Hall and Mairesse (2001).

[25] In estimation, we removed year means before performing any of the tests. For fixed *T*, this makes no difference to the asymptotic properties of the tests. As shown by Binder et al. (2000), when T is fixed, allowing for

random walk with drift ($y_{it} = y_{it-1} + \delta + \varepsilon_{it} \Rightarrow \Delta y_{it} = \delta + \varepsilon_{it}$) and a pure fixed effects process ($y_{it} = \alpha_i + \delta t + \varepsilon_{it} \Rightarrow \Delta y_{it} = \delta + \Delta\varepsilon_{it}$), and the six intermediate cases of a dynamic panel with or without fixed effects, taking ρ = 0.3, 0.9, and 0.99 and allowing α_i to vary across all individual units or imposing it to be the same for all of them. For each of these eight DGPs, we also consider both a homoskedastic version with $\sigma_\varepsilon^2(i)$ constant across the units, and a heteroskedastic one with $\sigma_\varepsilon^2(i)$ varying across the units.

Except for the random walk case, when constructing the DGPs we ensured that the resulting process satisfied covariance stationarity, in order to guarantee the consistency of the maximum likelihood estimators.[26] The exact calibration of the DGPs we used was derived from the first and second moments of the log R&D series and its first differences, as described below. The values of these moments were the following:

$$\begin{array}{ll} E[y] = 2.50 & E[\Delta y] = .085 \\ V[y] = 4.599 & V[\Delta y] = .0672 \end{array} \tag{14}$$

The necessary parameters are δ, ρ, μ_α, σ_ε^2, and σ_α^2. Conditional on α_i (or unconditional, because it is differenced out), we have the following two equations:

$$\begin{array}{l} E(\Delta y) = \delta \\ V(\Delta y) = 2\sigma_\varepsilon^2/(1+\rho) \end{array} \tag{15}$$

time-specific effects in estimation has no effect on the estimates of the other parameters so that these effects may be removed from the observed series before estimation.

[26] This stationary version of the dynamic panel model with fixed effects is denoted as Model I by Nickell (1981) and Lancaster (1997).

Given a value for ρ and the moments of our data, these equations give values of δ and σ_ε^2. Once we have values for δ, ρ, and σ_ε^2, to obtain the mean and variance of the distribution of the α_is, we use the moments of the series in levels:

$$
\begin{aligned}
E(y) &= \mu_\alpha + \delta(T+1)/2 \\
V(y) &= \sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{1-\rho^2}
\end{aligned}
\tag{16}
$$

Values of the parameters derived from the moment estimators specified by equations (15) and (16) are used to generate the simulated data as follows:

$$
\begin{aligned}
&\begin{bmatrix} \alpha_i \\ y_{i0} \end{bmatrix} \sim N\left[\begin{pmatrix} \mu_\alpha \\ E(y) \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & V(y) \end{pmatrix}\right] \\
&\varepsilon_{it} \sim N[0, \sigma_\varepsilon^2] \\
&y_{it} = (1-\rho)\alpha_i + (1-\rho)\delta t + \rho y_{i,t-1} + \varepsilon_{it}
\end{aligned}
\tag{17}
$$

where $N(\mu,\sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. It is straightforward to show that the processes generated according to these DGPs are mean and covariance stationary as long as $|\rho|<1$.[27] The AR(1) models without individual-specific effects are generated simply by assuming that $\sigma_\alpha^2 = 0$.[28]

For the non-stationary random walk case, we used the following four equations to determine the parameters δ, σ_ε^2, μ_0, and σ_0^2:

---

[27] For the "fixed effect" model, covariance stationarity is ensured by requiring the covariance of the initial condition y_{i0} and the individual-specific effect α_i to be σ_α^2.

[28] In this case it was not possible to reproduce the first two moments of the level and differenced series exactly, due to the fact that we were simulating a process that did not match our real data series that well. In all the other cases, the first two moments exactly identified the parameters needed.

$$\begin{aligned} &E(\Delta y) = \delta \\ &V(\Delta y) = \sigma_\varepsilon^2 \\ &E(y) = \mu_0 + \delta(T+1)/2 \\ &V(y) = \sigma_0^2 + \sigma_\varepsilon^2(T+1)/2 \end{aligned} \tag{18}$$

and generated the process using this model:

$$\begin{aligned} &y_{i0} \sim N(\mu_0, \sigma_0^2) \qquad i = 1,...,N \\ &\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2) \qquad t = 1,...,T \\ &y_{it} = \delta + y_{i,t-1} + \varepsilon_{it} \end{aligned} \tag{19}$$

In the heteroskedastic case, we allowed the variance of the shock ε to vary across firms. Inspection of the data revealed that a lognormal distribution of this variance was appropriate and the DGP we used was the following:

$$\begin{aligned} &\sigma_\varepsilon^2(i) = (1-\rho^2)\sigma_\omega^2(i) \\ &\log \sigma_\omega^2(i) \sim N[-2.05, 1.33] \end{aligned} \tag{20}$$

4 Results of Simulations

Table 2a reports the results of simulations designed to explore the behavior of the t-test and likelihood ratio test based on CML estimates.[29] The likelihood function used is given in equation (5) and the null hypothesis is that $\rho = 1$. As described earlier, the data used for the simulation were generated by processes whose first and second moments were chosen to match

[29] Estimating this model by maximum likelihood requires computation using the *T-1* by *T-1* variance-covariance matrix, which is perhaps why the CMLE method has not been used much in the literature. We implemented the estimator as an MLPROC in TSP 4.5 and found it to be fairly well-behaved, converging in 5 or 6 iterations if a good estimate of σ^2 (one based on the actual data) was used as a starting value along with a positive ρ. The TSP code is available as an example at http://www.tspintl.com.

those of the log of real R&D for the United States. The table has two panels, one for data generated with homoskedastic disturbances and one for data generated with firm-specific variances as described in equation (20).

The first column of each panel gives the average value of ρ and its standard deviation that was estimated by CMLE. In both cases (homoskedastic and heteroskedastic), these are fairly close to the true value, with a hint of downward bias for very large values of ρ. The next two columns give the average t-statistic for the hypothesis that ρ = 1, its standard deviation, and the size or power of the test as measured by the number of rejections at the 5% level of significance. The two columns following give the average likelihood ratio statistic for the same hypothesis and its size or power. It is clear from the table that both tests have approximately the correct size and considerable power when applied to homoskedastic data, except when the autoregressive coefficient is near unity (equal to 0.99). Note that when the true ρ is at or near unity, occasionally estimation using the simulated data will converge to the boundary of the parameter space, that is, ρ =1. In this case, we consider the hypothesis to be accepted, but we record the probability that this happens in the table (about 25 per cent of the time for the random walk, and about 10 per cent of the time for ρ = 0.99).

The final 4 columns of the table repeat the same exercise, but this time using data that were simulated to have the heteroskedasticity visible in our empirical series. The results are similar, with the following two exceptions: The sizes of the tests are slightly too large and the power against the alternative with fixed effects and ρ = 0.99 is actually slightly greater (note that this test is not size-adjusted). Both results are presumably due to the same fact: introducing some heterogeneity into the process reduces the probability of accepting the very restrictive null model if we impose homoskedasticity where it does not exist.

Table 2b reports the results of testing the hypothesis that ρ =1 using the CML estimator that allows for heteroskedasticity on data generated by the same homoskedastic and heteroskedastic processes as were used for Table 2a. The results are similar, except that the size of the test is now much too large and its corresponding power against large ρ alternatives much greater. Also, the t-test on ρ now gives a result that is quite different from the likelihood ratio test in the large ρ case. It appears that estimating the individual variances leads to results that bias the estimated ρ downwards in samples of our size, in spite of the consistency result of Kruiniger.

The first panel of Table 3 shows the results of applying the H-T test for a unit root to our simulated data. Not surprisingly, the results are very similar to those for the homoskedastic CMLE, with good power except when ρ is near unity, and too large a size when applied to heteroskedastic data. Thus when N is large and T small, it makes little difference to the result whether we use the inconsistent LSDV estimator and bias-adjust the answer, or the consistent CMLE estimator, which does not require bias adjusting. The underlying model was the same in both cases, and both require homogeneity of the coefficients and variances under the null and the alternative.

The next test considered is the pooled OLS test, which relaxes the assumptions of constant variance across time and individuals. The results of this test conducted on our simulated data are shown in the final columns of Table 3. The size of the test is approximately correct for both homoskedastic and heteroskedastic data, and the power is considerably better than for the Harris-Tzavalis or CMLE tests when ρ is near one. In spite of this fact, but not surprisingly, the estimates of ρ are severely biased towards one when the data are generated under an alternative with a fixed firm effect. This test does almost as well on heteroskedastic data as on homoskedastic, reflecting the fact that both the estimator and the standard error estimates are consistent under the null in both cases.

The results of conducting the IPS test with and without individual-specific trends, but with a zero augmenting lag imposed are shown in Table 4. The statistic shown is the average of an augmented Dickey-Fuller statistic for the N unit root tests on the individual series, together with empirical size or power of the test, based on critical values given in the tables of the IPS paper. We present results for a model both with and without a firm-specific time trend; all results are for data with a single cross-sectional mean removed in each year (that is, a full set of time dummies), as suggested by IPS and as was done for the other tests considered in this paper. Because our simulated data have no time trend, we expect that removal of these means will make the two tests (with and without allowing for a time trend) equivalent. However, due to the small sample of time periods available, requiring estimation of another parameter (the trend) could be somewhat costly in terms of degrees of freedom and may reduce the power of the test for samples of our size.

The results using the simulated data confirm this: the test without a trend has more power to discriminate between a random walk and a fixed effect plus AR(1) model than that with a trend. In the latter case, the size is too large, and the power against an alternative with $\rho = 0.9$ considerably weaker, whether or not there are also fixed effects in the model. Not surprisingly, the results are similar for the simulated heteroskedastic data. Because the IPS test is based on individual-level Dickey-Fuller tests, it allows for firm-level heterogeneity in variances, so adding this feature to the data generating process has only a limited effect on the results of the tests.

In Table 5 we present a summary of our results from these various tests. The first three columns contain results from the tests that are invalid when there is firm-level heteroskedasticity and last five columns results from those tests that remain valid in that case. We note first that the size of the former group of tests is larger than the theoretical value in the presence of the kind of heteroskedasticity displayed by our data, implying that these tests for a unit root will reject the null too often. In addition, all the tests have very low power against a near-unit root autoregressive model with fixed effects; recall that in this case, the fixed effect itself is multiplied by $(1-\rho)$ and therefore very small, so this result is not that surprising.

Most of the other tests have good size properties, with the exception of the conditional maximum likelihood estimates that allow for firm-specific heteroskedasticity. The most likely reason for the problems with the t-test based on the CML-HS estimates is that our standard error estimates are conventional and it is necessary to use a "sandwich" estimator here; see Kruiniger (1999b). The empirical standard error for the results in Table 2b was approximately 25-50% greater than the estimated standard error. However, we note also that our estimate of ρ does seem to be slightly downward biased in this case (see Table 2b), in spite of the fact that it is consistent, which implies that the rate of asymptotic convergence may be slow.

Restricting attention to the tests with the correct size that are robust to heteroskedasticity, we are left with the OLS test and the IPS test without a trend. The results of these tests differ significantly, in that the OLS test has by far the greater power against near-unit root alternatives, whether or not there is a fixed effect. The difference in power is doubtless due to the difference in alternative hypotheses, in that the OLS test considers $\rho = 1$ versus a single value of $\rho < 1$ for all individuals, whereas the IPS test considers $\rho = 1$ for all individuals versus $\rho < 1$ for at least one individual. Our simulated alternatives were all closer to the former model than the latter, so it is not surprising that the test does better in this case.

5 Results of Unit Root Tests for Observed Data

We now turn to our results for the observed data; details on the construction of these datasets and their characteristics are given in Hall et al (1999).[30] Table 6 reports the results of the tests for unit roots on the real data, highlighting the tests which reject nonstationarity at the 5 percent level in bold. The H-T test, which assumes homogenous time series processes that have no residual serial correlation beyond the first lag give essentially the same result as the IPS test without a trend: sales and employment are nonstationary and the remaining series are stationary, except for R&D in the United States. The IPS test with a trend is somewhat more likely to find a unit root, but as we have seen, the power of this test is low when the first order serial correlation is high.

The final two columns show one of our preferred tests for these data, the OLS test. Unlike the others, this test, which we saw to have more power against the alternative of stationarity with a very high autocorrelation coefficient, rejects nonstationarity in all cases except sales and employment in Japan. The estimated AR(1) coefficients are very high, so it is not surprising that we encountered difficulties with the tests that allow for the presence of fixed effects. Using the estimated values of $\rho\text{-}1$, we conducted a small analysis of variance on these 15 numbers which showed that the coefficients for U. S. and Japan could not be distinguished, while those for France were slightly more negative (implying lower serial correlation). The most significant differences were between investment and cash flow on the one hand and sales,

[30] A description may also be found in an unpublished appendix to this paper, available at http://emlab.berkeley.edu/users/bhhall/index.html.

employment, and R&D on the other, with the latter having a differenced coefficient of almost zero, as we saw in Figures 1 and 2.

Table 7 shows the results for the tests based on the two different CML estimates. Those based on the homoskedastic estimator give results very similar to the H-T test, as they should, since they rely on the same set of assumptions about the DGP. As in the earlier table, these results clearly reject non-stationarity for investment and cash flow, and for R&D in France and Japan. However, almost all of the real series reject the presence of a unit root when the heteroskedastic version of the CMLE is used. We suspect that some of the rejection may be due to the fact that both the coefficient and the standard error estimates seem to be systematically biased downward for samples of our size. Finally we note that a likelihood ratio test for constancy of variances clearly rejects in all cases.[31]

6 Conclusions

We began this investigation with the question of whether it was possible to distinguish between a model with a unit root or a model with a fixed effect and low order serial correlation when describing univariate time series data. Our principal conclusion is that the preferred model for our data is neither; rather it is a model with an extremely high serial correlation coefficient, but one that is less than one.

[31] Strictly speaking, this test is not valid asymptotically, since it is a test based on a number of parameters that grows at the same rate as the sample. We report it mainly as a heuristic indicator of the large difference allowing for heteroskedasticity makes to the likelihood.

With respect to the menu of unit root tests for fixed T samples, we have learned several things: The first conclusion from our simulation study of unit root tests is that the pooled OLS test and the IPS test have good power against most alternatives, although results from these tests differ when the alternative includes a coefficient near unity, primarily because they consider two quite different alternatives. Second, CML estimation is surprisingly easy to perform, even in the presence of heteroskedasticity, and may be a useful addition to the panel data arsenal, even if it is not as robust as simple OLS for the very particular problem of unit root testing. Further investigation should explore the reasons for finite sample bias in the heteroskedastic version of the CML estimator.

Substantively, we concluded that a very simple autoregressive model with a coefficient on the lag dependent variable that is near unity is a more parsimonious description of our data than a model with fixed effects. In Table 5, we observed that the only test with power against the ρ = 0.99 alternative was the OLS test, and in Table 6, this is the only test that rejects non-stationarity in favor of stationarity with a very large auto-regressive coefficient for almost all the real series. An alternative interpretation of this result is possible: the OLS test may be inappropriate because the proper alternative is heterogeneous serial correlation across the firms, implying that the IPS test is more appropriate. We have favored the former conclusion, not because we do not believe in heterogeneity of this kind, but because the more parsimonious model seems to describe the data fairly well, and because when serial correlation is this high, whether homogeneous or heterogeneous, the presence or absence of fixed effects is of little import, since they are necessarily quite small.

This fact leads us to a somewhat more controversial view that short panels of firm data are better described as having highly varied and persistent initial conditions rather than permanent unobserved firm effects. This feature of the data has been described by some as "not-so-fixed" firm effects. We would prefer to shift the emphasis in our modeling towards the idea that firm level differences are better captured by the initial condition, with the apparent "permanence" of differences being ascribed to very high serial correlation rather than to some left-out unobserved and permanent difference. We believe that this view of the firm is closer to the reality of firm evolution.

With our results in mind, some future research questions suggest themselves. First is the possibility of testing for the presence of firm-specific drifts or trends. It is certainly feasible to construct a CMLE of the doubly-differenced model in order to test for these, although the data may not have enough power for estimation. Second, given the near unit root behavior of the series, it may be of interest to examine their cointegrating properties. Mairesse, Hall, and Mulkay (1999) have already shown that a well-behaved error-correcting version of an investment equation can be constructed using data to ours, which implies that sales and capital stock are co-integrated and move together in the "long run". The possible interpretive significance of such a result is to unify the commonly observed differences between cross sectional and time series estimates based on panel data into a single model.

Bibliography

Arellano, Manuel. 1988. "An Alternative Transformation for Fixed Effects Models with Predetermined Variables." Oxford University Applied Discussion Paper No. 57. Institute of Economics and Statistics, Oxford University.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58: 277-297.

Arellano, Manuel, and Olympia Bover. 1991. "Another Look at the Instrumental-Variable Estimation of Error-components Models." *Journal of Econometrics* 68: 29-52.

Bai, Jushan, and Serena Ng. 2002. "A New Look at Panel Testing of Stationarity and the PPP Hypothesis," this volume.

Baltagi, B., and C. Kao. 2000. "Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey." In B. Baltagi (ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics*, Vol. 15, JAI: Amsterdam.

Bartelsman, Eric, and Wayne Gray. 1994. "TFP: The Productivity Database." Cambridge, Mass.: National Bureau of Economic Research. http://www.nber.org.

Binder, Michael, Cheng Hsiao, and M. Hashem Pesaran. 2000. "Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration," Cambridge, England: University of Cambridge DAE Working Paper No.0003.

Blundell, Richard S., and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics* 87: 115-143..

Bond, Stephen, Julie Ann Elston, Jacques Mairesse, and Benoit Mulkay. 1995. "A Comparison of Empirical Investment Equations Using Company Panel Data for France, Germany, Belgium, and the UK." *Review of Economics and Statistics*, forthcoming 2003.

Bond, Stephen, Celine Nauges, and Frank Windmeijer. 2002. "Unit Roots and Identification in Autoregressive Panel Data Models: A Comparison of Alternative Tests." Institute of Fiscal Studies, London: Photocopied.

Breitung, Joerg. 2000. "The Local Power of Some Unit Root Tests for Panel Data." In B. Baltagi (ed.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Advances in Econometrics*, Vol. 15, JAI: Amsterdam, 2000, p. 161-178.

Crepon, Bruno, Francis Kramarz, and Alain Trognon. 1998. "Orthogonality Conditions and Autoregressive Error Component Models." *Journal of Econometrics* 82: 135-156.

Hall, Bronwyn H., and Jacques Mairesse. 2001. "A Cautionary Note on Modeling the Univariate Behavior of Panel Data for Firms," paper presented to the German Classification Society Annual Meeting, Munich, Germany, March 14-15.

Hall, Bronwyn H., Jacques Mairesse, Lee Branstetter, and Bruno Crepon. 1999. "Does Cash Flow Cause Investment and R&D: An Exploration using Panel Data for French, Japanese, and United States Firms in the Scientific Sector", in D. Audretsch and R. Thurik (eds.), *Innovation, Industry Evolution and Employment*, Cambridge: Cambridge University Press.

Hall, Bronwyn H. 1990. "The Manufacturing Sector Master File: 1959-1987." Cambridge, Mass.: National Bureau of Economic Research Working Paper No. 3366.

Harris, Richard D. F., and Elias Tzavalis. 1999. "Inference for Unit Roots in Dynamic Panels where the Time Dimension is Fixed," *Journal of Econometrics* 91: 201-226.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56: 1371-1395.

Im, Kyung So, M. Hashem Pesaran, and Yongchool Shin. 1995. "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics*, forthcoming 2002.

Keane, Michael P., and David E. Runkle. 1992. "On the Estimation of Panel Data Models with Serial Correlation When Instruments are not Strictly Exogenous," *Journal of Business and Economic Statistics* 10: 1-29.

Kruiniger, Hugo. 1999a. "GMM Estimation of Dynamic Panel Data Models with Persistent Data," London: Queen Mary and Westfield College Working Paper No. 428.

Kruiniger, Hugo. 1999b. "Conditional Maximum Likelihood Estimation of Dynamic Panel Data Models," London: Queen Mary and Westfield College Working Paper No. 429.

Lancaster, Tony. 2002. "Orthogonal Parameters and Panel Data," *Review of Economic Studies* 69(3): 647-666.

Lancaster, Tony, and Bernard Lindenhovius. 1996. "Biases in Dynamic Panel Data Models: A Reconsideration," Providence, RI: Brown University Department of Economics Working Paper No. 96-11.

Levin, Andrew, and Chien-Fu Lin. 1993. "Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties." San Diego: University of California San Diego Discussion Paper 92-23.

Macurdy, Thomas E. 1985. "A Guide to Applying Time Series Models to Panel Data." Stanford, CA: Stanford University. Photocopied.

Mairesse, Jacques, and Bronwyn H. Hall. 1996. "Estimating the Productivity of Research and Development in French and United States Manufacturing Firms: An Exploration of

Simultaneity Issues with GMM." In van Ark, Bart, and Karin Wagner (eds.), *International Productivity Differences, Measurement, and Explanations*. Amsterdam: Elsevier-North Holland.

Mairesse, Jacques, and Alan Siu. 1984. "An Extended Accelerator Model of R&D and Physical Investment." In Griliches, Zvi (ed.), *R&D, Patents, and Productivity*. Chicago: Chicago University Press.

Nickell, Stephen. 1981. "Biases in dynamic models with fixed effects,'" *Econometrica* 49: 1399-1416.

OECD. 1991a. *OECD Economic Outlook: Historical Statistics, 1960-1989*. Paris: OECD.

OECD. 1991b. *Basic Science and Technology Statistics*. Paris: OECD.

Pesaran, M. Hashem, Yongchool Shin, and Richard P. Smith. 1997. "Pooled Estimation of Long-Run Relationships in Dynamic Heterogeneous Panels," University of Cambridge, Cambridge, England: Photocopied.

Quah, Danny, 1994. "Exploiting cross-section variations for unit root inference in dynamic data." *Economics Letters* 44, pp. 9-19.

World Bank. 1993. *World Tables 1993*. Baltimore, Maryland: Johns Hopkins University Press.

TABLE 1
Panel Data Unit Root Tests - Summary

Test	Testing principle	Description	Assumptions**	Heterogeneity across i	Heterogeneity across t**
CMLE	Wald, LR	Conditional Maximum Likelihood (t and LR tests)	independence across i, homoskedasticity, normality*	means	variances
CMLE-HS	Wald, LR	Conditional Maximum Likelihhod (t and LR tests)	independence across i, normality*	means; variances	variances
H-T	Wald	Bias and variance-corrected LSDV test	independence across i, homoskedasticity, no serial correlation	means	none
OLS	LM	Pooled OLS (SUR) regression with T equations	independence across i	variances	variances
IPS	Wald	Average of individual Dickey-Fuller tests without trend	independence across i	means, rho, variances	none
IPS-trend	Wald	Average of individual Dickey-Fuller tests with trend	independence across i	means, rho, trend, variances	none

*Could be interpreted as a quasi-maximum-likelihood estimator.

**The CMLE, SUR, and IPS tests can all be modified to allow for autocorrelation of the disturbances, as long as it is constant across individuals. However, the implementation in this paper assumes no serial correlation.

TABLE 2a

Testing for Nonstationarity using CML Estimator - Simulated Data

T=12; N=200

Estimation Method DGP	Conditional ML with homoskedastic variance Homoskedastic variance (1000 draws)					Conditional ML with homoskedastic variance Heteroskedastic variance (100 draws)				
Time series process for DGP	Estimated rho (s.d.)	Average t-statistic on rho (s.d.)	Empirical probability of rejection (#=1.0)**	Average Likelihood Ratio Test	Empirical probability of rejection (#=1.0)**	Estimated rho (s.d.)	Average t-statistic on rho (s.d.)	Empirical probability of rejection (#=1.0)**	Average Likelihood Ratio Test	Empirical probability of rejection (#=1.0)**
Random walk	0.992 (.012)	-0.53 (0.65)	0.060 (0.23)	0.53 (1.17)	0.028 (0.23)	0.984 (.023)	-0.81 (1.05)	0.14 (0.24)	1.29 (2.95)	0.10
AR(1) with r=0.3	0.298 (.025)	-29.2 (1.0)	1.000	563.3 (33.5)	1.000	0.264 (.043)	-29.1 (1.0)	1.00	560.7 (24.2)	1.00
AR(1) with r=0.9	0.899 (.021)	-4.82 (0.95)	0.999	22.3 (8.1)	1.000	0.900 (.022)	-4.83 (1.00)	1.00	22.5 (34.9)	1.00
AR(1) with r=0.99	0.986 (.016)	-0.83 (0.78)	0.145 (0.12)	1.13 (1.88)	0.090 (0.12)	0.984 (.017)	-0.82 (0.86)	0.16 (0.07)	1.39 (2.50)	0.10
Fixed effect	-0.002 (.024)	-42.6 (1.2)	1.000	1031.6 (45.1)	1.000	-0.002 (.039)	-42.6 (2.0)	1.00	1031.5 (72.1)	1.00
AR(1) FE, r=0.3	0.300 (.024)	-29.1 (1.0)	1.000	560.8 (32.5)	1.000	0.305 (.046)	-28.9 (2.0)	1.00	554.0 (63.4)	1.00
AR(1) FE, r=0.9	0.900 (.021)	-4.79 (0.94)	1.000	22.6 (8.2)	1.000	0.890 (.039)	-5.24 (1.69)	0.96	27.0 (14.8)	0.96
AR(1) FE, r=0.99	0.986 (.014)	-0.78 (0.72)	0.123 (0.11)	1.00 (1.57)	0.063 (0.11)	0.979 (.025)	-1.29 (1.26)	0.28 (0.18)	2.57 (4.38)	0.23

In the first row, the column labeled size or power is the size of a one-tailed t-test for rho<1 with nominal size 0.05.

In the other rows, it is the empirical probability of rejection by such a test.

The model estimated is y(i,t) = a(i) + a(t) + rho y(i,t-1) + e(i,t).

The method of estimation is Conditional Maximum Likelihood (fixed effects conditioned out).

**The fraction that converged to exactly rho=1.00 is given in parentheses, when it is nonzero.

TABLE 2b

Testing for Nonstationarity using CML-HS Estimator - Simulated Data

T=12; N=200

Estimation Method DGP	Conditional ML with Heteroskedasticity Homoskedastic variance (100 draws)					Conditional ML with Heteroskedasticity Heteroskedastic variance (25 draws)				
Time series process for DGP	Estimated rho (s.d.)	Average t-statistic on rho (s.d.)	Empirical probability of rejection (#=1.0)**	Average Likelihood Ratio Test	Empirical probability of rejection (#=1.0)**	Estimated rho (s.d.)	Average t-statistic on rho (s.d.)	Empirical probability of rejection (#=1.0)**	Average Likelihood Ratio Test	Empirical probability of rejection (#=1.0)**
Random walk	0.987 (.016)	-1.15 (0.73)	0.120	6.65 (5.28)	0.720	0.989 (.009)	-0.97 (0.55)	0.16	7.76 (3.93)	0.88
AR(1) with r=0.3	0.299 (.025)	-26.4 (1.0)	1.000	515.4 (30.3)	1.000	0.290 (.023)	-26.8 (1.0)	1.00	525.8 (29.9)	1.00
AR(1) with r=0.9	0.925 (.031)	-2.94 (1.05)	0.960	20.6 (9.5)	0.960	0.921 (.031)	-3.33 (1.24)	0.88	21.1 (8.0)	1.00
AR(1) with r=0.99	0.982 (.016)	-1.33 (1.74)	0.320	6.33 (3.87)	0.720	0.986 (.011)	-1.03 (0.54)	0.12	5.21 (5.09)	0.52
Fixed effect	-0.006 (.025)	-39.3 (1.2)	1.000	954.1 (46.3)	1.000	-0.003 (.028)	-39.1 (1.4)	1.00	950.2 (49.8)	1.00
AR(1) FE, r=0.3	0.307 (.032)	-26.2 (1.3)	1.000	508.2 (39.6)	1.000	0.399 (.027)	-26.5 (1.2)	1.00	517.3 (33.0)	1.00
AR(1) FE, r=0.9	0.908 (.030)	-3.75 (1.25)	0.960	19.2 (8.9)	0.960	0.912 (.036)	-3.54 (1.39)	0.88	20.7 (8.0)	0.96
AR(1) FE, r=0.99	0.984 (.013)	-1.07 (0.56)	0.200	4.59 (6.00)	0.560	0.985 (.018)	-1.12 (0.67)	0.16	7.62 (5.14)	0.84

In the first row, the column labeled size or power is the size of a one-tailed t-test for rho<1 with nominal size 0.05.

In the other rows, it is the empirical probability of rejection by such a test.

The model estimated is y(i,t) = a(i) + a(t) + rho y(i,t-1) + e(i,t).

The method of estimation is Conditional Maximum Likelihood (fixed effects and variances conditioned out).

**The fraction that converged to exactly rho=1.00 is given in parentheses, when it is nonzero.

TABLE 3

Testing for Nonstationarity - Simulated Data

T=12; N=200; 1000 Draws per Simulation

DGP	Harris-Tzavalis Method** Homoskedastic errors		Harris-Tzavalis Method** Heteroskedastic errors		Pooled OLS without Fixed Effects*** Homoskedastic errors			Pooled OLS w/o Fixed Effects*** Heteroskedastic errors		
Time series process for DGP	Average t-statistic on rho (s.d.)	Empirical probability of rejection	Average t-statistic on rho (s.d.)	Empirical probability of rejection	Estimate of rho (s.d.)	Average t-statistic on rho (s.d.)	Empirical probability of rejection	Estimate of rho (s.d.)	Average t-statistic on rho (s.d.)	Empirical probability of rejection
Random walk	-0.004 (1.03)	0.055	-0.16 (1.87)	0.210	1.000 (.003)	0.002 (1.09)	0.060	1.000 (.005)	-0.01 (1.04)	0.056
AR(1) with r=0.3	-32.0 (1.3)	1.000	-31.98 (2.18)	1.000	0.298 (.022)	-35.9 (2.3)	1.000	0.298 (.022)	-35.8 (2.3)	1.000
AR(1) with r=0.9	-4.00 (1.07)	0.987	-4.00 (1.04)	0.990	0.900 (.010)	-11.2 (1.2)	1.000	0.899 (.010)	-11.3 (1.3)	1.000
AR(1) with r=0.99	-0.45 (1.02)	0.113	-0.43 (1.00)	0.111	0.990 (.003)	-3.41 (1.08)	0.963	0.990 (.003)	-3.49 (1.13)	0.953
Fixed effect	-47.0 (1.2)	1.000	-46.87 (2.21)	1.000	0.999 (.001)	1.82 (1.11)	0.558	0.982 (.006)	-6.63 (0.71)	1.000
AR(1) FE, r=0.3	-31.9 (1.2)	1.000	-31.949 (2.29)	1.000	0.999 (.001)	-1.69 (1.11)	0.516	0.989 (.003)	-5.21 (0.90)	0.999
AR(1) FE, r=0.9	-3.97 (1.07)	0.990	-4.12 (1.88)	0.909	0.995 (.002)	-2.41 (1.09)	0.771	0.997 (.002)	-1.94 (1.04)	0.610
AR(1) FE, r=0.99	-0.40 (1.01)	0.106	-0.46 (1.79)	0.240	0.993 (.0003)	-2.93 (1.07)	0.891	1.000 (.001)	-0.77 (1.06)	0.207

In the first row, the column labeled size or power is the size of a one-tailed t-test for rho<1 with nominal size 0.05.

In the other rows, it is the empirical probability of rejection by such a test.

**The model estimated is y(i,t) = a(i) + a(t) + rho y(i,t-1) + e(i,t).

The method of estimation is ordinary least squares (within or LSDV).

***The model estimated is y(i,t) = a(t) + rho y(i,t-1) + e(i,t).

The method of estimation is seemingly unrelated regression, allowing for correlation across time.

TABLE 4

Testing for Nonstationarity - Simulated Data

T=12; N=200; 100 Draws per Simulation

	IPS Test (no trend, augmenting lags=0)						IPS Test (trend, augmenting lags=0)					
DGP disturbances	**Homoskedastic**			**Heteroskedastic**			**Homoskedastic**			**Heteroskedastic**		
Data Generating Process	**Average t-statistic on rho**	**Std. dev. of t-statistic**	**Empirical probability of rejection**	**Average t-statistic on rho**	**Std. dev. of t-statistic**	**Empirical probability of rejection**	**Average t-statistic on rho**	**Std. dev. of t-statistic**	**Empirical probability of rejection**	**Average t-statistic on rho**	**Std. dev. of t-statistic**	**Empirical probability of rejection**
Random walk	-0.09	1.05	0.05	-0.22	1.06	0.07	-0.10	1.12	0.11	-0.12	1.15	0.09
AR(1) with r=0.3	-41.15	1.13	1.00	-41.15	1.17	1.00	-23.97	1.21	1.00	-23.96	1.24	1.00
AR(1) with r=0.9	-4.90	1.10	1.00	-4.86	1.22	0.99	-0.77	1.06	0.21	-0.92	1.11	0.23
AR(1) with r=0.99	-0.68	1.14	0.19	-0.73	1.17	0.19	-0.39	1.30	0.18	-0.22	1.23	0.12
Fixed effect	-67.84	1.40	1.00	-67.74	1.26	1.00	-46.46	1.47	1.00	-46.35	-1.38	1.00
AR(1) FE, r=0.3	-41.00	1.06	1.00	-41.03	1.16	1.00	-23.88	1.20	1.00	-23.77	1.19	1.00
AR(1) FE, r=0.9	-4.74	1.02	1.00	-4.97	1.12	1.00	-1.04	1.20	0.27	-0.88	0.99	0.24
AR(1) FE, r=0.99	-0.49	1.07	0.14	-0.62	1.05	0.19	-0.36	1.10	0.10	-0.31	1.13	0.10

In the first row, the column labeled size or power is the size of a one-tailed t-test for rho<1 with nominal size 0.05.

In the other rows, it is the empirical probability of rejection by such a test.

TABLE 5

Empirical Probability of Rejection - Panel Data Unit Root Tests

	Invalid under Heteroskedasticity			Allows for Heteroskedasticity				
Time series process for DGP	CMLE t-test	CMLE LR-test	H-T Test	CML-HS t-test	CML-HS LR-test	Pooled OLS Test	IPS Test (no trend)	IPS Test (trend)
				Homoskedastic Data				
Random walk	0.06	0.04	0.06	**0.12**	**0.72**	0.06	0.05	**0.11**
AR(1) with r=0.3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR(1) with r=0.9	0.99	1.00	0.99	0.96	0.96	1.00	1.00	**0.21**
AR(1) with r=0.99	**0.15**	**0.09**	**0.11**	**0.32**	0.72	0.96	**0.19**	**0.18**
Fixed effect	1.00	1.00	1.00	1.00	1.00	0.56	1.00	1.00
AR(1) FE, r=0.3	1.00	1.00	1.00	1.00	1.00	0.52	1.00	1.00
AR(1) FE, r=0.9	1.00	1.00	0.99	0.96	0.96	0.77	1.00	**0.27**
AR(1) FE, r=0.99	**0.12**	**0.07**	**0.11**	**0.32**	0.56	0.89	**0.14**	**0.10**
				Heteroskedastic Data				
Random walk	**0.14**	**0.10**	**0.21**	**0.16**	**0.88**	0.06	0.07	**0.09**
AR(1) with r=0.3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR(1) with r=0.9	1.00	1.00	0.99	0.88	1.00	1.00	0.99	**0.23**
AR(1) with r=0.99	**0.16**	**0.10**	**0.11**	**0.12**	0.52	0.95	**0.19**	**0.12**
Fixed effect	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR(1) FE, r=0.3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR(1) FE, r=0.9	0.96	0.96	0.91	0.88	0.96	0.61	1.00	**0.24**
AR(1) FE, r=0.99	**0.28**	**0.23**	**0.24**	**0.16**	0.84	**0.21**	**0.19**	**0.10**

Figures in bold deviate from the correct size by 0.05 or more, or have power less than 0.50.

TABLE 6
Estimation and Testing with Normality and Homoskedasticity
Scientific Sector Firms

	Harris-Tzavalis Test		IPS Tests		Pooled OLS Estimates	
Column	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	AR (1) Coeff.	**Normal Test**	**No trend (p-value)**	**With trend (p-value)**	**AR (1) Coeff**	**T-test (p-value)**
Sales						
U.S.	.819 (.012)	.069 (.018)	1.31 (.904)	2.73 (.997)	**.9925 (.0020)**	**-3.69 (.000)****
France	.758 (.017)	.008 (.020)	-.10 (.460)	2.10 (.982)	**.9923 (.0024)**	**-3.19 (.001)****
Japan	.811 (.013)	.061 (.012)	2.17 (.985)	3.13 (.999)	.9956 (.0025)	-1.76 (.040)
R&D						
U.S.	.749 (.014)	-.001 (.018)	5.27 (.999)	6.68 (.999)	**.9916 (.0028)**	**-2.98 (.002)****
France	**.648 (.019)**	**-.102 (.020)****	**-1.93 (.027)****	0.33 (.629)	**.9903 (.0031)**	**-3.13 (.001)****
Japan	**.644 (.015)**	**-.106 (.012)****	**-1.72 (.043)****	2.25 (.988)	**.9869 (.0029)**	**-4.49 (.000)****
Investment						
U.S.	**.454 (.020)**	**-.296 (.018)****	**-5.96 (.000)*****	**-2.81 (.002)*****	**.9877 (.0029)**	**-4.22 (.000)****
France	**.405 (.023)**	**-.245 (.020)****	**-5.24 (.000)*****	**-3.99 (.000)*****	**.9519 (.0055)**	**-8.79 (.000)****
Japan	**.344 (.020)**	**-.406 (.012)****	**-9.13 (.000)*****	**-4.49 (.000)*****	**.9682 (.0049)**	**-6.55 (.000)****
Employment						
U.S.	.828 (.012)	.078 (.018)	4.16 (.999)	2.25 (.988)	**.9903 (.0023)**	**-4.15 (.000)****
France	.890 (.014)	.140 (.020)	6.89 (.999)	1.81 (.965)	**.9849 (.0024)**	**-6.32 (.000)****
Japan	.863 (.010)	.113 (.012)	4.59 (.999)	7.94 (.999)	.9983 (.0015)	-1.09 (.138)
Cash flow						
U.S.	**.576 (.020)**	**-.174 (.018)****	-.86 (.194)	-.18 (.427)	**.9874 (.0028)**	**-4.39 (.000)****
France	**.235 (.024)**	**-.515 (.020)****	**-9.92 (.000)*****	**-8.41 (.000)*****	**.9480 (.0088)**	**-5.89 (.000)****
Japan	**.468 (.018)**	**-.282 (.012)****	**-2.57 (.005)*****	3.12 (.999)	**.9896 (.0030)**	**-3.43 (.000)****

(1) the estimated coefficient of the lag dependent variable in a regression with fixed effects. These estimates contain "Nickell" bias.

(2) the same coefficient corrected for bias and standardized by the SEE under the null (see Harris and Tzavalis for details).

*,** denote lower tail significance at the 0.05, 0.01 level respectively.

(3) IPS tests are conducted with year means removed from the data, using a zero augmenting lag.

(5) the AR1 coefficient estimated using SUR; standard error robust to HS and serial correlation.

(6) the SUR t-statistic for rho=0, robust to heteroskedasticity and serial correlation.

** denotes lower tail significance at the .01 level.

Bold cells are those that reject a unit root at the 5% level of significance (conventional, not adjusted for true size).

TABLE 7
CML Estimation and Testing with Homoskedasticity and Heteroskedasticity
Scientific Sector Firms

	CMLE Estimates (Kruiniger 1999)				CMLE Estimates with Heteroskedastic errors				Test for Heteroskedasticity	
Column	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Variable (Country)	**AR (1) Coeff**	**LR Test (p-value)**	**Variance Estimate**	**Log Likelihood**	**AR (1) Coeff**	**LR Test (p-value)**	**Mean of Variance Est.**	**Log Likelihood**	**Log Likelihood**	**P-value (DF)**
Sales										
U.S.	>=1.000	0.0 (1.000)	.0272 (.0008)	861.7	**0.9969 (.0007)****	**50.4 (.000)**	0.0316	1603.5	1483.5	0.000 (214)
France	.998 (.020)	0.01 (0.938)	.0201 (.0007)	917.8	**0.9700 (.0114)****	**---**	0.0235	1414.8	994.1	0.000 (166)
Japan	>=1.000	0.0 (1.000)	.0152 (.0005)	1481.0	**0.9984 (.0006)****	**80.3 (.000)**	0.0167	2199.8	1437.6	0.000 (231)
R&D										
U.S.	>=1.000	0.0 (1.000)	.0668 (.0020)	-148.3	**0.9951 (.0023)****	**74.7 (.000)**	0.0777	866.9	2030.3	0.000 (214)
France	**.903 (.025)****	**14.2 (.000)**	.0658 (.0022)	-92.9	**0.9641 (.0152)****	**9.9 (.002)**	0.0697	498.2	1182.4	0.000 (166)
Japan	**.901 (.021)****	**21.5 (.000)**	.0152 (.0034)	-720.7	**0.9611 (.0081)****	**41.93 (.000)**	0.1291	237.3	1916.0	0.000 (231)
Investment										
U.S.	**.643 (.024)****	**173.7 (.000)**	.313 (.009)	-1794.1	**0.720 (.028)****	**94.2 (.000)**	0.3239	-1294.9	998.6	0.000 (214)
France	**.574 (.027)****	**186.2 (.000)**	.362 (.012)	-1470.1	**0.581 (.031)****	**157.8 (.000)**	0.3417	-1197.0	546.1	0.000 (166)
Japan	**.516 (.024)****	**279.2 (.000)**	.287 (.009)	-1608.7	**0.506 (.028)****	**251.0 (.000)**	0.2549	-1224.4	768.6	0.000 (231)
Employment										
U.S.	>=1.000	0.0 (1.000)	.0265 (.0008)	888.6	0.9973 (.0025)	**48.7 (.000)**	0.0298	1696.3	1615.3	0.000 (214)
France	>=1.000	0.0 (1.000)	.0113 (.0004)	1413.6	**0.9982 (.0004)****	**107.0 (.000)**	0.0075	2281.6	1736.0	0.000 (166)
Japan	>=1.000	0.0 (1.000)	.0031 (.0001)	3227.2	**0.9990 (.0003)****	**210.6 (.000)**	0.0034	4222.5	1990.6	0.000 (231)
Cash flow										
U.S.	**.863 (.026)****	**25.9 (.000)**	.1178 (.0038)	-656.1	0.997 (.004)	**8.3 (.004)**	0.1003	112.0	1536.2	0.000 (184)
France	**.453 (.034)****	**180.7 (.000)**	.3055 (.0123)	-927.4	**0.517 (.037)****	**104.5 (.000)**	0.5441	-476.1	902.6	0.000 (114)
Japan	**.784 (.023)****	**73.2 (.000)**	.0769 (.0023)	-262.6	0.995 (.004)	**78.3 (.000)**	0.0753	486.7	1498.5	0.000 (210)

Notes to Table 7

(1) the AR1 coefficient estimated using ML conditioned on the effects (individual means removed).

(3) the estimated variance of the disturbance corresponding to the estimate in (1).

(5) the AR1 coefficient estimated using ML conditioned on the effects and with Var(e) = sig(I)sig(t).

(7) the average of the estimated variances across the firms, with sig(t=1) normalized at unity.

(9) the likelihood ratio test for the heteroskedastic variances with degrees of freedom = (N+T-2).

Bold cells are tests that reject a unit root at the 5% level of significance (conventional, not adjusted for true size).

Note that the standard errors for the CML-HS estimates are conventional, and therefore incorrect.

Appendix A: CML Estimation with heteroskedasticity

In this appendix we describe the computational implementation of the conditional maximum likelihood estimation with heteroskedasticity.[32] The likelihood function we wish to maximize is the following:

$$\log L(\rho,\{\sigma_i^2\},\{\sigma_t^2\};\{y_{it}\}) = -\frac{(T-1)}{2}\log(2\pi) - \frac{(T-1)}{2}\sum_{i=1}^{N}\log(\sigma_i^2) - \frac{N}{2}\log|\Phi| - \sum_{i=1}^{N}\frac{1}{2\sigma_i^2}(Dy_i)'\Phi^{-1}Dy_i$$

where Φ is a *T-1* by *T-1* matrix that contains powers of ρ and the parameters given by σ_t^2, $t=2,\ldots,T$:[33]

$$\Phi = DPV_\rho PD' \quad \text{and} \quad P = diag(\sigma_t)$$

Thus evaluating the likelihood involves manipulation of matrices of order of the number of time periods. To do this easily, we make use of the MLPROC procedure in TSP Version 4.5. MLPROC takes a procedure that defines a log likelihood function as the output of a sequence of commands and maximizes the value returned by the procedure with respect to the chosen parameters, via repeated calling of the procedure to evaluate the function and its derivatives (numerically).

To simplify the computation of the likelihood as much as possible, we make use of the fact that estimators for σ_i^2, $i = 1,\ldots,N$ can be obtained from the first order condition given values for ρ and the σ_t^2 and concentrate these parameters out of the likelihood function:

---

[32] The homoskedastic version is an obvious simplification of the algorithm described here.

[33] Note that the parameterization requires one normalization on the σ_i^2 or σ_t^2 in much the same way that including a second set of dummies in an equation requires an additional exclusion restriction. Our normalization is $\sigma_t^2 = 1$ for $t = 1$.

$$\sigma_i^2 = \frac{1}{N(T-1)}\sum_{i=1}^{N}(Dy_i)'\Phi^{-1}Dy_i = \frac{1}{N(T-1}trace[\Phi^{-1}Z_i]$$

where

$$Z_i = (Dy_i)'Dy_i \qquad \text{and} \qquad Z = \sum_{i=1}^{N} Z_i = (Dy)'Dy$$

is the covariance matrix of the first differenced *y*s. As is well-known, when it is possible to concentrate the likelihood, the standard error estimates for the remaining parameters (ρ and σ_t^2, $t=2,\ldots,T$ in this case) are not affected by this procedure.

The algorithm is therefore the following:

1. Given values for ρ and σ_t^2, $t = 1,\ldots,T$, compute estimates of σ_i^2.

2. Use these estimates to compute the value of the likelihood using the following expression:

$$\log L(\rho,\{\sigma_t^2\};\{y_{it}\}) = -\frac{N(T-1)}{2}\log(2\pi+1) - \frac{(T-1)}{2}\sum_{i=1}^{N}\log(\sigma_i^2[\rho,\{\sigma_t^2\}]) - \frac{N}{2}\log|\Phi[\rho,\{\sigma_t^2\}]|$$

3. Iterate on 1 and 2 in the usual way, using a gradient method to maximize the likelihood with respect to ρ and σ_t^2.

Appendix B: Autocorrelograms of the data

Figure B1

Autocorrelation Functions for Logs of Real Variables

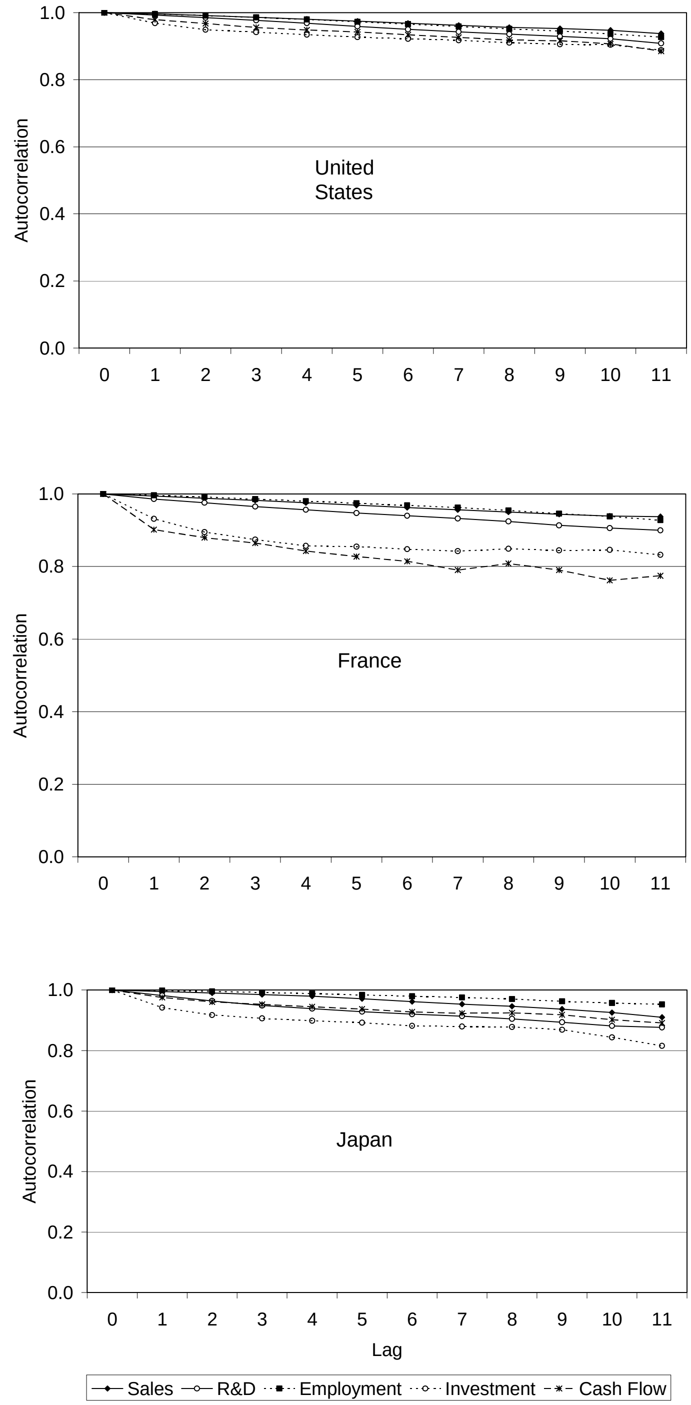
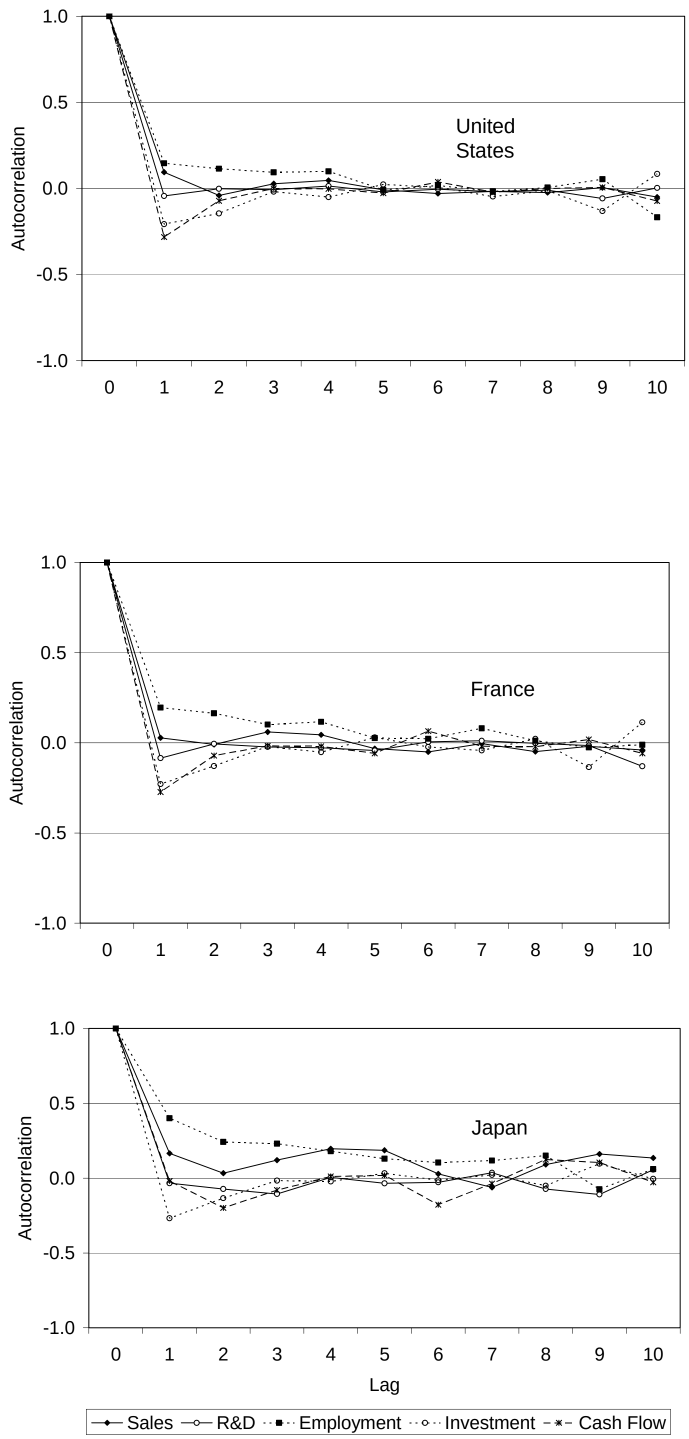


Figure B2

Autocorrelation Functions for Differenced Logs of Real Variables



Appendix C: Model estimates using simulated data

TABLE C

OLS and IV Estimates for Simulated Data

T=12; N=200; 1000 Draws

Estimation Method		Ordinary Least Squares Levels			Ordinary Least Squares Within			Ordinary Least Squares Differences			Instrumental Variables Differences		
Time series process for DGP	**True value of rho**	**Average estimate of rho**	**Std. deviation of estimated rho**	**Average estimated std. error on rho**	**Average estimate of rho**	**Std. deviation of estimated rho**	**Average estimated std. error on rho**	**Average estimate of rho**	**Std. deviation of estimated rho**	**Average estimated std. error on rho**	**Average estimate of rho**	**Std. deviation of estimated rho**	**Average estimated std. error on rho**
Random walk	1	1.000	0.003	0.003	**0.749**	**0.018**	**0.015**	**-0.001**	**0.026**	**0.025**	0.318	0.718	0.991
AR(1) with r=0.3	0.3	0.300	0.007	0.006	0.288	0.007	0.007	**-0.327**	**0.021**	**0.023**	0.294	0.052	0.058
AR(1) with r=0.9	0.9	0.900	0.004	0.004	**0.822**	**0.012**	**0.010**	**0.125**	**0.029**	**0.024**	0.896	0.026	0.066
AR(1) with r=0.99	0.99	0.990	0.003	0.003	**0.744**	**0.018**	**0.015**	**-0.006**	**0.024**	**0.025**	0.832	0.244	0.405
Fixed effect	0	**0.993**	**0.001**	**0.003**	-0.093	0.021	0.022	**-0.500**	**0.018**	**0.022**	-0.001	0.045	0.043
AR(1), FE, and r=0.3	0.3	**0.993**	**0.001**	**0.003**	**0.182**	**0.021**	**0.022**	**-0.349**	**0.021**	**0.023**	0.300	0.070	0.068
AR(1), FE, and r=0.9	0.9	**0.992**	**0.002**	**0.003**	**0.681**	**0.019**	**0.016**	**-0.046**	**0.024**	**0.025**	0.781	0.278	0.400
AR(1), FE,and r=0.99	0.99	0.993	0.003	0.003	**0.742**	**0.018**	**0.015**	**-0.004**	**0.025**	**0.025**	0.746	0.374	0.562

The data are simulated using parameters derived from the first and second moments

of log R&D in the United States, E(y)=2.5; V(y)=4.59; E(dy)=.085; V(dy)=.0672.

The model estimated in panel 1 is y(i,t) = a0 + delta*t + rho*y(i,t-1) + e(i,t).

The model estimated in panel 2 is y(i,t) = a(i) + delta*t + rho*y(i,t-1) + e(i,t).

The model estimated in panels 3 and 4 is dy(i,t) = delta + rho*dy(i,t-1) + e(i,t).

The instruments in panel 4 are dy(i,t-2), dy(I,t-3) and dy(I,t-4)