Univariate Panel Data Models and GMM Estimators: An Exploration Using Real and Simulated Data

Bronwyn H. HALL

Nu±eld College, Oxford; UC Berkeley; IFS; and NBER

Jacques MAIRESSE

INSEE-CREST; EHESS; and NBER

Version of January 2000

ABSTRACT

This paper explores the time series properties of commonly used variables in ¯rm-level panels: sales (turnover), employment, R&D, investment, and cash °ow. We focus on two questions: 1) whether the behavior of these series is consistent with stationarity, and if so, 2) what order of autoregressive process best describes them. The answer to these two questions is of substantive interest for those who model the dynamic evolution of ¯rms and their behavior. In particular, we are interested in whether ¯rm data is trend stationary (exhibits regression to individual ¯rm means) or di®erence stationary (evolves as a random walk, possibly with a non-zero drift). We ¯nd that estimation of even these very simple processes using fairly large but short panels is fraught with di±culty and we explore the convergence rate of the GMM estimator using simulation methods. We also report the results of using a new class of tests proposed by Im, Pesaran, and Smith for discriminating between stationary and nonstationary processes in medium-sized panels.

The ¯rst author thanks Stephen Bond for very helpful conversations and INSEE-CREST, Paris, Nu±eld College, Oxford, and the Institute for Fiscal Studies, London for hospitality while an earlier version of this paper was being written.

Keywords: panel data, investment, unit roots, international comparisons.
JEL codes:
Corresponding author: Bronwyn H. Hall,
Department of Economics, 549 Evans Hall
UC Berkeley, Berkeley, CA 94720-3880
email: bhhall@econ.berkeley.edu

Univariate Panel Data Models and GMM Estimators:

an Exploration Using Real and Simulated Data

Bronwyn H. Hall and Jacques Mairesse

1 Introduction

This results reported in this paper arose from our exploration of a simple question: could we ¯nd a simple parsimonious model that was a good description of the time series behavior of the key observable variables that describe the behavior of individual ¯rms: sales, employment, investment, R&D, and cash °ow or pro¯ts. The class of models we were interested in were those where the heterogeneity across ¯rms was summarized by a single ¯xed e®ect or individual-speci¯c intercept. A complication in this endeavor is that for most of these variables, the within ¯rm behavior in a short time series can be described approximately either as an autoregressive process with a near-unit root or as an autoregressive process with an individual speci¯c e®ect and a small positive coe±cient; either version will be rather di±cult to estimate in a short panel, owing to the importance of initial conditions to the process.

Our exploration started with a fairly general autoregressive model that included ¯rm-speci¯c e®ects, but we soon observed behavior in our GMM estimators that suggested the presence of ¯nite sample bias in our coe±cient estimates, which implied that our testing procedure was biased. We thus turned to evaluating two separate but related questions:

1. Using very simple stylized processes to generate data (random walk vs. ¯xed e®ect with no autoregression), what is the magnitude of the ¯nite sample bias when we estimate an autoregressive model using these data?

2. Using the same data generating processes, can our tests distinguish between data with a unit root and data with individual specific effects and a root outside the unit circle?

Having learned from these explorations which methods work best, we then proceed to test our real time series for the presence of a unit root, and to construct the most appropriate autoregressive models for each. Our conclusion is that the process which describes each of our variables is more similar across countries than across variables, and that the variables can be clearly ranked by their long run "persistence": employment, R&D, sales, cash flow, and investment (whose behavior most resembles that of a stationary process).

2 A Univariate Autoregressive Model with Firm Effects

We began by considering the following univariate autoregressive model with fixed firm and year effects:[1]

$$y_{it} = \alpha_i + \delta_t + \sum_{s=1}^{m} \beta_{t,s} y_{i,t-s} + \varepsilon_{it} \qquad i = 1,\ldots,N \text{ firms}; \quad t = 1,\ldots,T \text{ years} \tag{1}$$

y_{it} is the logarithm of the variable of interest (sales, R&D, investment, employment, or cash flow). In our application, the number of years T is 12 and the number of firms N is approximately 200, so we treat the time effect δ_t as fixed and include a set of year dummies in all estimations. The α_i are the permanent unobserved differences across firms, additive in the logs.

Initially we explored three features of the specification in equation (1): the presence of (correlated) firm effects α_i; the stationarity of the process, and the length of the lag m. In the course of these explorations, we found evidence that some of our GMM estimates for this process had considerable small sample bias, and we therefore turned to Monte-Carlo methods to evaluate our

[1] Equation (1) is similar to the basic model considered by Holtz-Eakin, Newey, and Rosen (1988).

estimators. We found that distinguishing between a nonstationary process that has no firm effect α_i and a stationary process that does have firm-specific means is extremely difficult in these data. The difference between these two processes matters both for estimation and for the economic interpretation.

It makes a difference for the estimation strategy because when the lag coefficient is unity, ordinary least squares on the differences is a consistent estimator, whereas the instrumental variable estimator is not identified because there are no valid instruments for the differenced y. On the other hand, for a stationary process, ordinary least squares on the differences is inconsistent, whereas instrumental variables is consistent, although possibly biased in samples of our size. From the economic modeling perspective, the nonstationary case implies that firm data obeys a form of generalized Gibrat's Law, where growth is independent of current size, whereas a finding of stationarity means that the sales, employment, etc. for any given firm does tend to regress to that firm's own mean.

In our simulations we used a version of equation (1) that contains a simple time trend as the estimating model with data generated by three simpler univariate time series processes:

- Fixed effect plus deterministic trend

$$y_{it} = a_i + \delta t + \varepsilon_{it} \qquad \text{or} \qquad \Delta y_{it} = \delta + \varepsilon_{it} - \varepsilon_{i,t-1} \tag{2}$$

- Random walk plus drift

$$y_{it} = y_{it-1} + \delta + \varepsilon_{it} \qquad \text{or} \qquad \Delta y_{it} = \delta + \varepsilon_{it} \tag{3}$$

- AR(1) with $\rho = 0.99$ plus deterministic trend

$$y_{it} = \rho y_{it-1} + \delta t + \varepsilon_{it}. \qquad \text{or} \qquad \Delta y_{it} = \rho \Delta y_{i,t-1} + \delta + \varepsilon_{it} - \varepsilon_{i,t-1} \tag{4}$$

All of these processes are capable of qualitatively reproducing the patterns in our data, but they have very di®erent medium to long run implications for the way in which our data series are expected to evolve. Before reporting the results of our simulations using data generated by the three models, we present some descriptive statistics for our real data series; we will use these empirical moments to calibrate our data generating models.

3 Data and Descriptive Statistics

Our data is drawn from the three datasets that were constructed for the bivariate causality tests involving investment, R&D, cash °ow, and sales in Hall, Mairesse, Branstetter, and Crepon (1999). In that paper our goal was to produce similar samples of high-technology manufacturing ¯rms for each of our countries: France, the United States, and Japan. However, our sources of data were quite di®erent and this means that the samples will never be exactly comparable, although they are quite representative. The table in Appendix A gives the sources of our data, our de°ators, and some indication of the number of observations available to us, both before and after cleaning.

The primary di®erence in the data sources is between France and the other two countries: in France, we have access to a Census of Manufactures-type sample with R&D data collected in survey form by the government for the Ministère de la Recherche. This means the data tend to be at a level somewhat lower than that of consolidated accounts (the \group" level), and that it is not con¯ned to publicly traded ¯rms. For the other two countries, we have data based on the ¯lings of publicly-traded ¯rms with agencies charged with monitoring the ¯nancial markets. Although the Japanese data are somewhat less consolidated than those for the United States, in the sense that they are not at the \group" level, they are consolidated to a level roughly comparable to that in the United States. Also in the case of Japan, the R&D data has been augmented with data from another survey, because the quality of publicly reported R&D data is very uneven.[2]

The de°ators also di®er somewhat across countries. In all cases, we de°ate R&D, investment, and cash °ow by a de°ator that is common across all industries.[3] On the other hand, we have attempted to varying degrees to construct real output measures rather than sales by de°ating our sales ¯gures by at least a 2-digit-level de°ator. In the United States, we are using de°ators aggregated up to the 2-digit level from the NBER Productivity Database, which is at the 4-digit level (Bartelsman and Gray 1994). In France, we are using the N40 industry level de°ator (approximately 2-digit), which do not contain very much of the type of hedonic quality adjustment that is used in the United States. In Japan, we have constructed \¯rm-level" de°ators based on the 4-digit industry composition of the ¯rm's output.

Our choice of years (1978-1989) re°ects data availability, as well as a desire to have a fairly long time series available for each variable for use in instrumental variables estimation. Because these datasets are large, and in some cases fairly dirty, and because we want to focus on the common time series properties across ¯rms, rather than isolated reorganizations and other speci¯c disturbances, we apply cleaning rules to all the variables:

1. We require their growth rates to be between -90% and 900%.
2. In order to remove erroneous data values that might produce misleading autoregressive estimates, we remove ¯rms that have sequential growth rates that are large and alternate in sign. For sales and employment, large is de¯ned as below -50% or above 100%; for R&D, it is -67%, 200%; for investment and cash °ow, it is -80%, 400%.

[2] See Griliches and Mairesse (1990) for further discussion of this point.

[3] Thus we are implicitly assuming that the market for capital goods, and the market for R&D are common across all our ¯rms, so they face the same prices. In fact, even when measured carefully, there is little variation in the relative price of investment goods across industries, so our procedure is unlikely to produce much bias.

3. We work with the logarithms of all the variables, in order to minimize heteroskedasticity and problems with in°uential outliers.

After cleaning, and requiring a full 12 years of data for each ¯rm, we were left with 204 ¯rms for the United States, 156 for France, and 221 for Japan. Table 2 in the appendisx gives some indication of the typical size of these ¯rms, and their importance in their national economies. Each sample is a small but not insigni¯cant portion of its economy, and a larger fraction of that economy's private R&D activity. The Japanese sample has the largest coverage of both GDP and BERD (Business Enterprise R&D) and the French sample the smallest. Although the national R&D intensities are ranked Japan (2.2%), the U.S. (2.0%), and France (1.6%), the typical ¯rm in these samples is more R&D-intensive in the U.S. (with an R&D to sales ratio of 4.0%), followed by France (3.6%) , and then Japan (2.8%). This perhaps re°ects the somewhat greater selectivity of the U.S. and French samples, and the large integrated ¯rm structure typical of Japan. Because of this vertical integration, we suspect that our Japanese ¯rms are slightly less concentrated in the high-technology sector than the sample of ¯rms from the other two countries.

Tables 1 and 2 show the means, standard deviations, and autocorrelations for the two of the series: R&D in the United States, which the most highly autocorrelated series, and investment in France, which is the least. Table 1 shows the statistics for the levels (in logs) and Table 2 for the ¯rst di®erences (in logs, which is approximately equal to the growth rates). In general the means and standard deviations of our variables do not change much over time, although those for the United States do show a slight increase. Figures 1 and 2 show the autocorrelation plots for all the variables for the United States and France respectively. It is clear from these plots that the autocorrelations decay very slowly for all variables with the possible exception of investment and cash °ow in France, which suggests either that the time series process used to describe these data

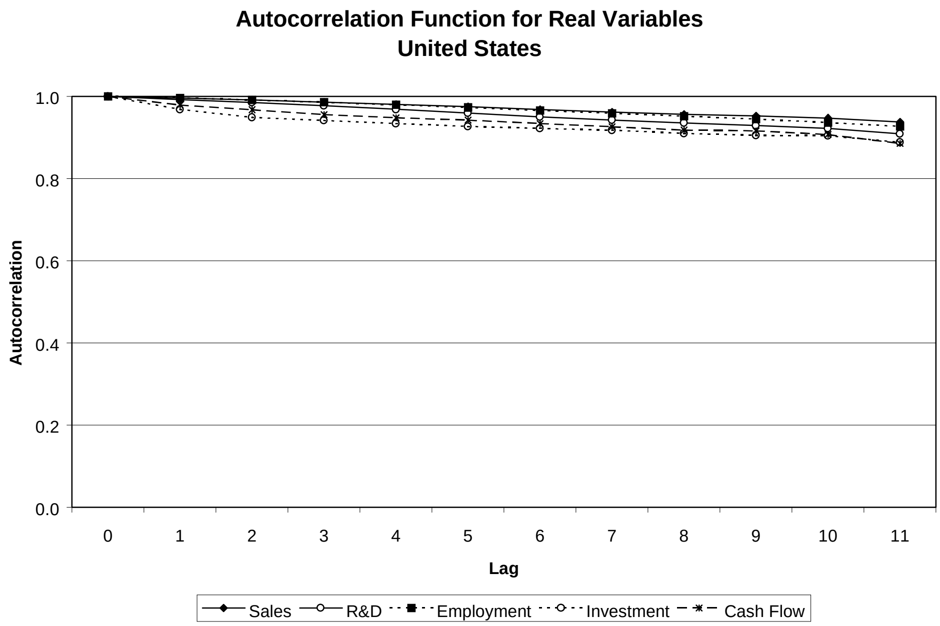

Figure 1: United States AC Plot

will have a root close to one, or that the series are dominated by the permanent di®erences in the level of the variables across ¯rms and low variance within. For this reason, we are interested in the ability of our estimation methods to distinguish the models described earlier.

4 Short Panels and Finite Sample Bias

This section reports the results of a set of simulations that explore the properties of ordinary least squares (OLS), instrumental variables (IV), and panel Generalized Method of Moments (GMM) estimation on the three simple univariate time series processes given in equations (2)-(4). All of these models are highly simpli¯ed versions of the model in (1) and it is well known that data generated by them can look very similar in short panels (and also that ¯rm level data is often well-described by such models).

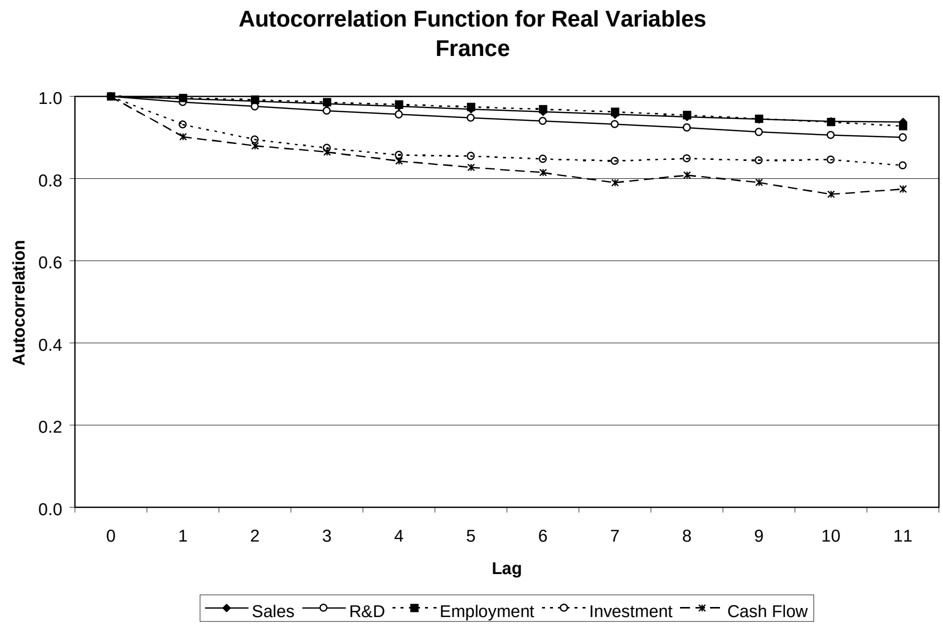

Figure 2: France AC Plot

The precise data generation models we used were chosen to mimic the distribution of the U.S. R&D series in our dataset. The deterministic trend model was the following:

$$y_{it} = a_i + \delta t + \varepsilon_{it} \tag{5}$$

$$\text{with} \quad \delta = 0.085; \; a_i \sim N(1.93; 2.18) \text{ and } \varepsilon_{it} \sim i.i.d.N(0; 0.1).$$

This model implies the estimating equation $\Delta y_{it} = \delta + \Delta\varepsilon_{it}$. But if the researcher does not know the process and includes lagged differences of y in the equation, the ordinary least squares estimates of their coefficients will be inconsistent (and typically very negative) because of the presence of $\varepsilon_{i,t-1}$ in the disturbance. When estimated by instrumental variables or GMM, $y_{i,t-2}$ will be an appropriate instrument for Δy_{it-1} when this latter variable is included in the regression.

The stochastic trend model was the following:

$$y_{it} = y_{it-1} + \delta + \varepsilon_{it} \tag{6}$$

$$\text{with} \quad \delta = 0.085; \; y_{i0} \sim N(1.93; 2.13) \text{ and } \varepsilon_{it} \sim i.i.d.N(0; 0.33).$$

For this model the estimating equation is $\Delta y_{it} = \delta + \varepsilon_{it}$ and ordinary least squares with lagged differenced y will produce consistent estimates (that is, coefficients of zero), but using $y_{i,t-2}$ as an instrument for the lagged first difference will not work well because the lagged differences of y are white noise (once the year means are removed). The IV estimator in this case is not identified.

The autoregressive model was the following:

$$y_{it} = \rho y_{it-1} + \delta t + \varepsilon_{it} \tag{7}$$

$$\text{with} \quad \delta = 0.01; \; y_{i0} \sim N(1.93; 2.20) \text{ and } \varepsilon_{it} \sim i.i.d.N(0; 0.186).$$

As in the case of the fixed effect model, for this model the estimating equation is $\Delta y_{it} = \rho \Delta y_{it-1} + \delta + \Delta\varepsilon_{it}$ and ordinary least squares estimates will be inconsistent due to the serial correlation in the error term and the presence of a lagged dependent variable. Instrumental variable or GMM estimation using $y_{i,t-2}$ as an instrument will be consistent. However, note that the estimated coefficient of the lagged dependent variable should be close to one, whereas for the fixed effect model it will be approximately zero, even though superficially the implied level behavior for series generated by the two models is quite similar.

Note also that the only difference between the data generating process in the two differenced models corresponding to equations (5) and (6) is that the error process for the first is first order moving average and that for the second is white noise. In all other respects they are identical.

The results of using OLS, IV, and GMM estimation on data generated by these two processes are shown in Tables 3 (the fixed effect model), 4 (the random walk model), and 5 (the autoregressive model). In all cases, the model estimated was the following:

$$\Delta y_{it} = \delta + \sum_{s=1}^{5} \beta_s \Delta y_{i,t-s} + u_{it} \qquad i = 1,\ldots,N \text{ firms}; \quad t = 1,\ldots,T \text{ years} \tag{8}$$

The IV and GMM estimates were obtained using the intercept and $y_{it-2},\ldots,y_{it-6}$ as instruments. Two choices of T and N were used: $T = 6$ or 20 and $N = 200$ or 2000. The first set of choices corresponds roughly to the order of magnitude in our data, but in some cases we found that only the large sample size yielded good finite sample behavior for the estimates. The OLS and IV estimates were obtained by stacking the model to have NT observations; in the case of the IV estimator, this implies that the projection in the first stage has the same coefficients in all the years. The GMM estimators are obtained with first stage projections that vary across the years. The estimator labelled "GMM 1-Step" is the estimator preferred by Blundell and Bond (1998), which is obtained using a weighting matrix based on the time series covariance predicted assuming white noise εs in the level fixed effect model (that is, first order moving average us). We will discuss the results in each of the 3 tables in turn.

Tables 3a and 3b contain the results for the model with fixed firm effects and no lagged dependent variable. The ordinary least squares estimates are biased downward as expected, and by essentially the same order of magnitude regardless of sample size. Thus even for samples of our size ($T = 6$ and $N = 200$) the results have converged to the asymptotic bias (which is quite large). In contrast, the IV estimates suffer severely from finite sample bias, which gets somewhat better as we increase T.[4] For samples of our size, the estimated lag coefficients using IV are similar to the (very inconsistent) ordinary least squares estimates and only get close to the true values when

[4] In an earlier version of this paper, we presented results that used all the available instruments rather than simply 5 per year. In this case, the finite sample bias gets worse as T increases rather than better, due to overfitting at the first stage. The current version of the paper uses only lags 2-6 as instruments for this reason. Other researchers have found similar results when estimating dynamic panel data models using GMM (cites?).

N = 20,000, which is 100 times our sample size. At a sample size of N = 2000, they still have a standard error that is so large that a coefficient of unity is amost as likely as a coefficient of zero. Also noteworthy is the fact that in this case the one-step GMM estimator appears to be better than the two-step estimator. The finite sample problem here is probably due to the fact that the efficient GMM estimator uses a very large number of orthogonality conditions relative to the amount of information in the instruments.

The conclusion from Table 3 is that for our sample size, no estimator does very well: only IV gets the right answer within a (very large) standard error and even the one-step GMM estimator appears to be biased downward. The specification tests are generally correct: the test for serial correlation of the disturbances accept the absence of second-order correlation but not of first, as would be expected. The omnibus chi-squared (J) test for overidentification accepts the validity of the instruments, although note that when the sample size get larger, this test comes close to rejection. We also include a Sargan test for fixed effects which is computed by adding the 6 additional orthognality conditions in levels that are implied by the absence of correlated firm effects; this test always rejects, as it should.

Tables 4a and 4b contain the results for the random walk plus drift model. The results show that, in spite of the lack of identification of this model, data generated by the random walk with drift model give reasonable answers using GMM; although the instruments are poor in this case, the IV and GMM 2-Step estimates were correct (within the very large standard errors). However, the GMM 1-Step estimates are far from the true values, reflecting perhaps the fact that the weighting matrix in this case is inappropriate.[5] Of course, OLS is the preferred estimator for data generated

[5]Recall that the GMM-1 estimator of Arellano and Bond is based on the assumption that the error term in the equation is MA(1). Another way to rewrite the differenced random walk model is the following:

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta \varepsilon_{it} \qquad \text{with } \rho = 1$$

The above way of writing the model contains a zero drift and a moving average error, but in fact is exactly the same by this model and those estimates do indeed appear to be the best, with by far the lowest standard errors.[6] Note that all of the test statistics accept that this is the correct specification, except that the Sargan test for fixed effects rejects, even though there are no fixed effects in the model. This is perhaps not surprising empirically, since the level version of the model has a lagged dependent variable with a coefficient of unity, but it is puzzling and confirms that nonstationarity in panel data can cause peculiar estimator behavior in the same way it does in ordinary time series.

Tables 5a and 5b contain the results for the autoregressive model with deterministic trend. Recall that in this case, unlike the preceding, ordinary least squares is inconsistent and IV should be consistent. This is clearly true in our simulations, although note that the inconsistency declines slightly as T increases. All of the instrumental variable estimates are pretty good when the sample size is 2000 or higher. However, for samples of our size, IV is somewhat imprecise, and the one-step GMM estimator is strongly preferred to the two-step, which appears to be biased downward. In this case, the Sargan test accepts the absence of fixed effects, as it should. Clearly the reason is because the level equation is essentially the same as the first-differenced equation, unlike the random walk case.

The results in Tables 3, 4, and 5 are quite discouraging, because it appears that there is no estimator among the set we looked at that is appropriate for all of our (admittedly) extreme models, at least for our sample size. Where one is consistent and/or unbiased, the other will be inconsistent or at least biased in finite samples. The preferred estimator is probably the one-step

model as the random walk

$$\Delta y_{it} = \varepsilon_{it}$$

. Since the GMM-1 estimator assumes an MA(1) error, it is perhaps not surprising that our estimates of the slope coefficient are biased upward substantially and our estimate of the drift coefficient downward toward zero, particularly when the sample size is large in the T dimension.

[6] In theory, GMM 2-Step is asymptotically as good as OLS for this model, of course, but we do not expect this estimator to converge as quickly (if at all) to the true value given the poor instruments.

GMM estimator, but this estimator is not identified if the true process is a random walk and suffers from finite sample bias in the other cases. This means that if we wish to investigate the univariate time series process in our data, we probably should not rely too heavily on these instrumental variable approaches. Two other approaches suggest themselves: estimation directly on the implied moments of the process without using instrumental variables and models that allow more completely for the heterogeneity in the data. In the next section we explore another method of distinguishing a deterministic from a stochastic trend in panel data that allows for some heterogeneity.

5 Testing for Unit Roots - the IPS Method

Recent work by Im, Pesaran, and Shin (1997, hereafter IPS) suggests another approach for distinguishing these two data generating processes that allows for more heterogeneity of behavior than we allowed for in Tables 3 to 5. They assume the following model:

$$y_{it} = (1 - \phi_i)\mu_i + \phi_i y_{i,t-1} + \varepsilon_{it}; \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{9}$$

where initial values y_{i0} are given, and they test for the null hypothesis that ϕ_i is unity for all observations versus an alternative that some of the ϕ_is are less than one. Under the null, there is no fixed effect, while under the alternative, each fixed effect α_i is equal to $(1 - \phi_i)\mu_i$. They propose tests based on the average over the individual units of a Lagrange-multiplier test of the hypothesis that ϕ_i as well as tests based on the average of the (augmented) Dickey-Fuller statistics, which they find to have somewhat better finite sample properties.

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

$$y_{it} = (1 - \phi_i)\mu_i + (1 - \phi_i)\delta_i t + \phi_i y_{i,t-1} + \varepsilon_{it}; \quad i = 1, \ldots, N; \quad t = 1, \ldots, T \tag{10}$$

We will use both these tests for our data, since there is reason to believe that trends do exist in the series. Note that an important di®erence between these models and the models considered in the previous section is that both the lag coe±cient and the trend coe±cient are allowed to di®er across ¯rms under the alternative hypothesis of stationarity.

We applied these tests to our simulated data, with the results shown in Table 6. 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 of the test, based on critical values given in the tables of the IPS paper. The precise statistic we used is computed as in their equation (5.3), which allows for the number of augmenting lags to di®er across the individuals. In practice, we found a p of either 2 or 3 was the preferred value given by the individual unit root tests, even though the data were in all cases generated from models where $p = 0$ was appropriate. In the tables, we present results for a model both with and without a ¯rm-speci¯c 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. Because our simulated data have a single time trend, we expect that removal of these means will make the two tests (with and without allowing for a time trend) equivalent.

The results using the simulated data con¯rm that this is the case; in fact, the test without a trend has more power to discriminate between ¯xed e®ects and a random walk than that with in this case. Focusing on the results using the second version of the IPS test, with a trend and an intercept, here we ¯nd that although the empirical size of the test does di®er depending on whether there is a lagged dependent variable or not, even in the ¯xed e®ect case we will fail to reject the null hypothesis 32 percent of the time, using the "best" augmenting lag. And there is no hope of distinguishing a stationary autoregression from one with a unit root if the autoregressive coe±cient is 0.99 in samples the size of ours: for this model the test statistic is essentially the same as the value for the random walk model.

We explored the e®ects of increasing our sample size in the bottom two panels of Table 6. Larger sample size in the N dimension reveals that our approximation to the IPS distribution for the short sample and p = 3 is not very good, so we focus on the p = 2 case. Here the results for discrimination between ¯xed e®ects and random walk are excellent, but there is still no di®erence between the random walk and AR(1) models. To discriminate between these two models, one needs much longer time series than we are likely to have available in this type of data (note that increasing T from 6 to 20 does not help much). .

Our conclusion from this exercise is that in a short panel it will be extremely di±cult to tell with con¯dence whether the process can be described by an autoregressive process with a unit root but no permanent di®erence across ¯rms, or by a ¯xed ¯rm e®ect plus perhaps an autoregressive process with much smaller coe±cients, especially when there is serial correlation in the disturbances (or a higher order autoregressive process for the series). However, if we restrict the process under consideration to a low order autoregression, the IPS unit root test appears to be able to distinguish a model with ¯xed e®ects and no autoregression from one with autoregression, even in a short panel.

6 Results of Unit Root Tests for Observed Data

[This section still under development]

Turning to the real data (Table 7) and using the Sargan test, only 3 data series are able to reject the absence of ¯rm e®ects: sales and R&D in the U.S. and R&D in France, although cash °ow in the U.S. and France nearly rejects. However, the IPS unit root test, which allows the alternative hypothesis to have both a ¯rm-speci¯c autoregressive coe±cient and a ¯rm-speci¯c intercept, accepts nonstationarity (no firm effects) for all the series except investment in both the U.S. and France and cash flow in France.[7] When we allow for a firm-specific trend in addition, nonstationarity is rejected only for investment and cash flow in France. We view these results as somewhat inconclusive, but as suggestive that future modeling for these firms should try to account for the heterogeneity in slope coefficients that appears to be causing the differences in these results.

....

7 Stationary with Firm Effects or Nonstationary?

[This section still under development.]

In a general panel data regression model, α_i may or may not be correlated with the regressors. In the case here, the question of correlation is not really very interesting; if α_i is present, it is certainly correlated with lagged y. The relevant question is whether the data are well-described by the same autoregressive model where firms differ only in their initial condition, or whether there is a tendency for them to regress toward individual means; that is, whether there are permanent unobserved differences across firms.

To answer this question, we estimated the model given by equation (1) for each of our data series using the Generalized Method of Moments with two sets of orthogonality conditions. The first uses only orthogonality conditions between first differences of the model and levels of the lagged variables and is consistent even if there are permanent correlated firm effects. The second uses orthogonality conditions between levels of the model and levels of the lagged dependent variable and is efficient if there are no firm effects, but inconsistent otherwise. We conducted tests for the validity of the additional restrictions implied by the equations in levels using a lag length (m) of 5.

[7] All the tests were conducted allowing for augmenting lags in the Dickey-Fuller regression; the order of the lag was always 2 or 3, and the test was computed as the average of the t-statistics standardized by the appropriate values from Table 2 of IPS.

Tables 5 (old) and 6 (old) show the results of this procedure, first for simulated data and then for our actual data. In Table 5 (old), we show the results of estimation using data generated by the following processes (in all cases ε_{it} is a white noise process):

$$y_{it} = y_{it-1} + .085 + \varepsilon_{it} \qquad \text{(Random walk with drift)} \tag{11}$$

$$y_{it} = \alpha_i + .085t + \varepsilon_{it} \qquad \text{(Fixed effect with trend)} \tag{12}$$

$$y_{it} = 0.99\, y_{it-1} + .011 + \varepsilon_{it} \qquad \text{(AR(1) with intercept)} \tag{13}$$

$$y_{it} = \alpha_i + 0.76\, y_{it-1} + .012t + \varepsilon_{it} \qquad \text{(AR(1) with fixed effect and trend)} \tag{14}$$

The parameters for these processes were chosen to mimic the empirical behavior of the log R&D series for the United States, which has a mean that increases from 1.9 to 2.9 between 1978 and 1989, a variance of approximately 4.6, and first order serial correlation of about 0.99. The fixed effect α_i and the initial conditions were generated from a normal distribution with mean 1.93 and variance 4.4 and 4.75 respectively. The sample size was 200 (the same order of magnitude as our actual sample size) and the number of time periods used for estimation was 7, with 5 additional time periods for the lagged instruments, for a total number of years equal to 12. The data generated by these processes were then fit to the model of equation (1) both in level equations (inconsistent in the presence of α_i) and in first differences, with the length of the lag chosen using conventional F-tests. In addition, a test for the presence of firm effects was conducted by testing for the significance of the overidentifying restrictions implied by equation (1) in levels.

The first 3 columns of Table 5 (old) show estimates of the first differenced model, the next two columns show estimates of the model in levels, and the final 2 columns display two tests for the presence of correlated firm effects.[8] The first differenced estimates are based on orthogonality

[8] Because of the evidence in Arellano and Bond (1991) that the efficient two-step GMM estimates have downward-

conditions (OCs) between the ¯rst di®erences of the model and level variables, lagged 2 to 5 times (yielding 6*(4+1) = 30 OCs). FD and level estimates are based on OCs between the model in levels and level variables lagged 1 to 5 times (yielding 7*(5+1) = 42 OCs). In both cases, the number of lags shown is the number arrived at under each model by conventional Wald tests at the one percent level. The sum of the lag coe±cients is the sum of the estimated lag coe±cients at the maximum number of lags (5) and heteroskedastic-consistent standard errors are shown in parentheses.

The Sargan test is a test for the validity of the additional orthogonality conditions implied by the levels model, which are invalid in the presence of correlated e®ects. The Hausman test is a speci¯cation test for the equality of the coe±cient estimates in the two sets of estimates. A puzzle, which we do not resolve here, is that even when the Sargan test suggests that the level estimates are valid, the coe±cient estimates are su±ciently di®erent between the two methods that the Hausman test rejects. The Sargan test correctly accepts the absence of ¯xed e®ects in the case of the random walk and the AR(1) model, but it also accepts in the case of the simple ¯xed e®ect model; the reason is clear when one looks at the level estimate of the lag coe±cients, which sum to nearly unity. However, an important di®erence between data generated by this model and data generated by the random walk model is the estimated length of the lag: in the former case, there is an equal loading of about 0.2 on each of the 5 lags, whereas in the latter, the ¯rst coe±cient is unity (as it should be) and the rest are zero.

The AR(1) model with no e®ects is very similar to the random walk model, except that in this case all the lag coe±cients enter when the model is estimated in levels; it is clear that the near

biased standard errors in samples of this size, the coe±cient estimates and the Hausman tests in the table are for the one-step estimator with the MA(1) covariance of the di®erenced residuals suggested by those authors. However, the Sargan test is based on the chi-squared from the two-step estimates, since this test requires the use of e±cient estimates if it is to have the correct distribution under the null.

non-stationarity of this model creates the same difficulties for estimation as the random walk model, due to the weakness of the instruments. The AR(1) model with fixed effects is the best-behaved of the four, clearly rejecting the absence of firm effects using both the Sargan and the Hausman tests. The (consistent) first differenced estimate of the sum of the lag coefficients is close to the true value, but the lag length is wrong (5 instead of 1).[9]

8 Estimating the AR Process for the Observed Data

[this section to be redone using results of previous sections.]

We performed series of investigations into the functional form of the AR model for the observed data.[10] These tests were performed using the full set of linear moment restrictions implied by the structure of the model; that is the number of orthogonality conditions used grows as T^2.

With 12 years of data available for each variable from 1978-1989, the maximum number of lags in the autoregressive process is $m = 10$; after differencing to eliminate the α_i. We choose the length of the lag m in each case below by finding the longest set of lags (up to $m = 10$) where the last $10 - (m + 1)$ enter with a p-value greater than .01, using a Wald or generalized F test. This means we are using a different number of equations for estimation, depending on the maximum order m. The models and tests whose estimates are shown in Tables A.1-A.3 are described below:

1. ONEST: Stationary coefficients.

$$y_{it} = \alpha_i + \delta_t + \sum_{s=1}^{m} \beta_s y_{i,t-s} + \varepsilon_{it} \qquad (15)$$

[9] One reason for these differences may be that the data generation processes use drift (in the case of random walk and AR(1) and trend (in the case of Fixed Effect and AR(1) with FE) to model the common time series behavior, whereas the estimation procedures use a model with year dummies. Since in practice, the researcher has no idea what the time series behavior is, our approach to modeling it seems conservative.

[10] These investigations were carried out on a somewhat larger sample of firms.

2. ONEME: Stationary with measurement error. The model is the same as (A.2), but we need lags from $m+1$ to $2m$ as instruments in this case. This reduces the length of lag that can be investigated to $m = 5$. For this model, we test the validity of the additional orthogonality conditions allowed if there is no measurement error (implying that lags 1 to m can be used as instruments).

3. ONELAM: Time-varying coefficients for the fixed effect, but constant lag coefficients.

$$y_{it} = \psi_t \alpha_i + \delta_t + \sum_{s=1}^{m} \beta_s y_{i,t-s} + \varepsilon_{it}. \tag{16}$$

We test whether $\psi_t = 1$ for all t (this reduces A.3 to A.2).

4. ONELAM2: Proportional lag coefficients, and a constant fixed effect

$$y_{it} = \alpha_i + \delta_t + \lambda_t \left[\sum_{s=1}^{m} \beta_s y_{i,t-s} \right] + \varepsilon_{it}. \tag{17}$$

We test whether $\lambda_t = 1$ for all t (this reduces A.4 to A.2).

5. ONENS:Time-varying lag coefficients, and a constant fixed effect.

$$y_{it} = \alpha_i + \delta_t + \sum_{s=1}^{m} \beta_{t,s} y_{i,t-s} + \varepsilon_{it}. \tag{18}$$

The test NS1 tests stationarity of the coefficients, that is, $\beta_{t,s} = \beta_s$ for all t, and the test NS2 tests proportionality of the β's over time, which is equivalent to the model in equation (A.4):

$$\frac{\beta_{t,s}}{\beta_{t-1,s}} = \frac{\beta_{t,r}}{\beta_{t-1,s}} \tag{19}$$

for all t and for all r and s.

Except for sales and cash °ow in Japan (where coe±cients were fully nonstationary), we tended to ¯nd that the data were well ¯t by either a fully stationary model or one with a single proportionality coe±cient across years, that is, by equations (15), (16) or (17). The data are typically unable to choose between models (16) and (17), which is not surprising given the presence of lagged endogenous variables on the right hand side. The optimal lag length m ranged from 2 to 8, but was usually 3 and 5.

9 Conclusions

to be written

References

[1] Arellano, Manuel. 1988. \An Alternative Transformation for Fixed E®ects Models with Predetermined Variables." Oxford University Applied Discussion Paper No. 57. Institute of Economics and Statistics, Oxford University.

[2] Arellano, Manuel, and Stephen Bond. 1991. \Some Tests of Speci¯cation for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." Review of Economic Studies 58: 277-297.

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

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

[5] Blundell, Richard S., and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," ....

[6] 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. Institute of Fiscal Studies, London: Photocopied.

[7] Crepon, Bruno, Francis Kramarz, and Alain Trognon. 1993. \Parameter of Interest, Nuisance Parameter, and Orthogonality Conditions: An Application to Autoregressive Error Components Models." Journal of Econometrics | : | | .

[8] 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 Scienti¯c Sector", in D. Audretsch and R. Thurik (eds.), Innovation, Industry Evolution and Employment, Cambridge: Cambridge University Press.

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

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

[11] 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.

[12] 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 Di®erences, Measurement, and Explanations. Amsterdam: Elsevier-North Holland.

[13] 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.

[14] OECD. 1991a. OECD Economic Outlook: Historical Statistics, 1960-1989. Paris: OECD.

[15] OECD. 1991b. Basic Science and Technology Statistics. Paris: OECD.

[16] Pesaran, M. H., Shin, Y., Smith, R. P. 1997. "Pooled Estimation of Long-Run Relationships in Dynamic Heterogeneous Panels," University of Cambridge, Cambridge, England: Photocopied.

[17] World Bank. 1993. World Tables 1993. Baltimore, Maryland: Johns Hopkins University Press.

Table 1
Empirical First and Second Moments

Variable = Real Log (R&D) for United States (N=424)

	Mean	St Dev	Correlation Matrix											
Year			1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
1978	1.925	2.195	1.000											
1979	2.057	2.174	0.991	1.000										
1980	2.175	2.166	0.980	0.993	1.000									
1981	2.304	2.120	0.973	0.985	0.994	1.000								
1982	2.400	2.135	0.962	0.976	0.988	0.995	1.000							
1983	2.505	2.130	0.953	0.966	0.981	0.988	0.994	1.000						
1984	2.610	2.116	0.939	0.954	0.971	0.980	0.987	0.993	1.000					
1985	2.687	2.107	0.928	0.941	0.960	0.969	0.975	0.981	0.991	1.000				
1986	2.740	2.117	0.923	0.935	0.955	0.962	0.969	0.974	0.983	0.991	1.000			
1987	2.824	2.142	0.916	0.929	0.950	0.956	0.962	0.969	0.979	0.988	0.995	1.000		
1988	2.865	2.138	0.920	0.932	0.949	0.954	0.960	0.964	0.974	0.981	0.990	0.993	1.000	
1989	2.861	2.191	0.909	0.924	0.939	0.943	0.947	0.951	0.961	0.970	0.978	0.981	0.991	1.000

Variable = Real Log (Investment) for France (N=156)

	Mean	St Dev	Correlation Matrix											
Year			1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
1978	8.951	1.767	1.000											
1979	9.011	1.738	0.933	1.000										
1980	9.111	1.674	0.908	0.932	1.000									
1981	9.027	1.663	0.888	0.897	0.941	1.000								
1982	9.022	1.588	0.842	0.848	0.872	0.929	1.000							
1983	9.054	1.594	0.842	0.843	0.871	0.895	0.936	1.000						
1984	9.273	1.616	0.856	0.877	0.874	0.884	0.879	0.921	1.000					
1985	9.370	1.659	0.855	0.853	0.873	0.890	0.877	0.906	0.935	1.000				
1986	9.430	1.600	0.850	0.841	0.872	0.871	0.837	0.867	0.889	0.933	1.000			
1987	9.439	1.569	0.849	0.856	0.880	0.874	0.838	0.869	0.896	0.908	0.926	1.000		
1988	9.520	1.527	0.859	0.844	0.868	0.856	0.815	0.846	0.863	0.868	0.898	0.935	1.000	
1989	9.610	1.593	0.832	0.834	0.841	0.824	0.780	0.814	0.835	0.844	0.876	0.903	0.928	1.000

Table 2
Empirical First and Second Moments - First Differenced Series

Variable = Real Log (R&D) for United States (N=424)													
	Mean	St Dev	Correlation Matrix										
Year			1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
1979	0.132	0.293	1.000										
1980	0.118	0.261	0.051	1.000									
1981	0.129	0.242	0.058	0.038	1.000								
1982	0.096	0.212	0.016	0.068	0.058	1.000							
1983	0.105	0.233	0.078	0.016	0.067	0.081	1.000						
1984	0.105	0.259	-0.013	0.069	0.039	0.062	0.112	1.000					
1985	0.077	0.278	-0.045	-0.015	0.097	0.029	0.086	0.099	1.000				
1986	0.053	0.286	-0.012	-0.055	0.010	0.100	0.059	0.092	0.131	1.000			
1987	0.084	0.210	-0.123	0.017	-0.049	0.017	0.098	0.125	0.086	0.146	1.000		
1988	0.041	0.259	-0.059	-0.130	0.037	-0.020	0.075	0.113	0.146	0.127	0.153	1.000	
1989	-0.004	0.300	0.188	-0.080	0.078	-0.056	-0.064	0.071	0.015	0.104	-0.013	0.212	1.000

Variable = Real Log (Investment) for France (N=156)													
	Mean	St Dev	Correlation Matrix										
Year			1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
1979	0.060	0.644	1.000										
1980	0.101	0.633	-0.230	1.000									
1981	-0.084	0.573	-0.130	-0.226	1.000								
1982	-0.004	0.619	-0.020	-0.139	-0.221	1.000							
1983	0.031	0.571	-0.057	-0.026	-0.171	-0.205	1.000						
1984	0.220	0.638	0.035	-0.070	-0.020	-0.160	-0.207	1.000					
1985	0.096	0.591	-0.024	0.037	-0.069	-0.029	-0.181	-0.181	1.000				
1986	0.061	0.600	-0.043	-0.037	0.046	-0.086	-0.036	-0.199	-0.188	1.000			
1987	0.009	0.610	0.026	-0.046	-0.042	0.061	-0.073	-0.026	-0.166	-0.182	1.000		
1988	0.081	0.560	-0.134	0.058	-0.049	-0.042	0.151	-0.132	0.009	-0.124	-0.251	1.000	
1989	0.090	0.595	0.114	-0.155	0.111	-0.048	-0.173	0.135	-0.042	-0.011	-0.115	-0.321	1.000

Table 3a
SIMULATION RESULTS - 50 Draws
Data Generation Model: Fixed Effect plus Linear Trend
Estimation Model: Fixed Effect, Linear Trend, and 5 Lags of Dependent Variable

# of time periods # of observations Estimation method	True Model	6 200 OLS	6 2000 OLS	6 20000 OLS	6 200 IV	6 2000 IV	6 20000 IV	6 200 GMM 1Step	6 2000 GMM 1Step	6 200 GMM 2Step	6 2000 GMM 2Step
DY (-1)	0	-.838 (.026)	-.834 (.010)	-.834 (.002)	-7.6 (48.9)	.073 (.185)	-.002 (.062)	-.388 (.197)	-.133 (.125)	-.790 (.289)	-.380 (.248)
DY (-2)	0	-.673 (.034)	-.669 (.010)	-.668 (.004)	-5.5 (35.4)	.057 (.147)	-.001 (.050)	-.313 (.158)	-.108 (.099)	-.632 (.227)	-.306 (.197)
DY (-3)	0	-.509 (.036)	-.501 (.008)	-.500 (.003)	-4.3 (27.6)	.043 (.108)	-.000 (.038)	-.236 (.117)	-.080 (.074)	-.478 (.167)	-.229 (.147)
DY (-4)	0	-.334 (.032)	-.332 (.009)	-.334 (.004)	-3.1 (19.8)	.030 (.073)	-.000 (.025)	-.155 (.086)	-.053 (.048)	-.311 (.120)	-.152 (.096)
DY (-5)	0	-.165 (.025)	-.167 (.009)	-.167 (.003)	-1.4 (9.0)	.013 (.038)	-.001 (.014)	-.079 (.044)	-.028 (.025)	-.156 (.058)	-.077 (.047)
Sum of coefficients	0	-2.52 (.12)	-2.50 (.03)	-2.50 (.01)	-21.9 (140.7)	.216 (.547)	-.006 (.189)	-1.17 (.58)	-.401 (.367)	-2.37 (.85)	-1.14 (.73)
Trend	0.085	.298 (.012)	.298 (.003)	.298 (.001)	1.9 (11.8)	.067 (.046)	.085 (.016)	.190 (.045)	.119 (.031)	.286 (.072)	.182 (.062)
Standard error	0.34	.356 (.007)	.358 (.002)	.358 (.001)	3.25 (17.65)	.489 (.044)	.469 (.015)				
R-squared		.421 (.016)	.417 (.006)	.417 (.002)	.395 (.056)	.373 (.074)	.363 (.093)				
Chi-squared (DF)	30.0									30.7 (6.5)	34.3 (7.5)
degrees of freedom	30									30	30
LM (1) Test	sig.							-10.66 (3.45)	-42.5 (2.3)	-1.96 (7.01)	-32.4 (12.7)
LM (2) Test	insig.							-0.20 (.80)	-0.22 (.68)	-1.19 (1.13)	-1.24 (1.48)
Sargan Test for FE	sig.									27.2 (10.1)	263.4 (31.3)
degrees of freedom	6									6	6

Table 3b
SIMULATION RESULTS - 50 Draws
Data Generation Model: Fixed Effect plus Linear Trend
Estimation Model: Fixed Effect, Linear Trend, and 5 Lags of Dependent Variable

# of time periods # of observations Estimation method	True Model	20 200 OLS	20 2000 OLS	20 10000 OLS	20 200 IV	20 2000 IV	20 10000 IV	20 200 GMM 1Step	20 2000 GMM 1Step	20 200 GMM 2Step	20 2000 GMM 2Step
DY (-1)	0	-.833 (.014)	-.834 (.005)	-.834 (.002)	-.039 (.178)	.003 (.045)	.005 (.022)	-.142 (.065)	.049 (.147)	-.735 (.121)	-.148 (.045)
DY (-2)	0	-.667 (.020)	-.667 (.006)	-.667 (.003)	-.034 (.140)	.002 (.039)	.004 (.019)	-.119 (.045)	.046 (.128)	-.590 (.091)	-.118 (.038)
DY (-3)	0	-.497 (.021)	-.500 (.006)	-.500 (.003)	-.021 (.102)	.003 (.030)	.004 (.022)	-.080 (.031)	.044 (.110)	-.433 (.068)	-.089 (.029)
DY (-4)	0	-.330 (.018)	-.333 (.006)	-.333 (.002)	-.015 (.074)	.001 (.020)	.003 (.009)	-.062 (.033)	.039 (.090)	-.289 (.049)	-.060 (.020)
DY (-5)	0	-.168 (.014)	-.167 (.015)	-.167 (.012)	-.012 (.042)	.001 (.011)	.001 (.004)	-.036 (.025)	.036 (.070)	-.150 (.029)	-.029 (.012)
Sum of coefficients	0	-2.49 (.07)	-2.50 (.02)	-2.50 (.01)	-.120 (.525)	.011 (.146)	.017 (.065)	-.438 (.185)	.213 (.543)	-2.20 (.34)	-.444 (.142)
Trend	0.085	.297 (.006)	.297 (.002)	.298 (.001)	.095 (.045)	.084 (.012)	.084 (.005)	.123 (.016)	.071 (.040)	.273 (.029)	.122 (.012)
Standard error	0.34	.358 (.005)	.358 (.001)	.358 (.001)	.462 (.040)	.470 (.011)	.471 (.005)				
R-squared		.417 (.009)	.417 (.003)	.417 (.001)	.314 (.133)	.321 (.108)	.306 (.117)				
Chi-squared (DF)	108.0									110.2 (10.8)	121.5 (14.0)
degrees of freedom	108									108	108
LM (1) Test	sig.							-25.95 (.72)	-60.1 (37.6)	-6.67 (6.25)	-81.4 (1.6)
LM (2) Test	insig.							.10 (.43)	-1.93 (3.29)	-1.98 (1.30)	-.13 (.49)
Sargan Test for FE	sig.									72.3 (11.2)	1484.5 (22.8)
degrees of freedom	6									6	6

The chi-squared shown is for the Sargan test of overidentification (instrument validity).
In all cases, a Hausman test for uncorrelated effects rejects.
The number in parentheses are the standard deviation of the estimate over the 50 simulation draws.
The columns with 10000 or 20000 observations are averages over 20 simulation draws.
Instruments are y(-2)-y(-6).

Table 4a
SIMULATION RESULTS - 50 Draws
Data Generation Model: Random Walk with Drift
Estimation Model: Fixed Effect, Linear Trend, and 5 Lags of Dependent Variable

# of time periods		6	6	6	6	6	6	6	6	6	6
# of observations	True	200	2000	20000	200	2000	20000	200	2000	200	2000
Estimation method	Model	OLS	OLS	OLS	IV	IV	IV	GMM 1Step	GMM 1Step	GMM 2Step	GMM 2Step
DY (-1)	0	-.002 (.030)	-.002 (.010)	-.001 (.003)	-1.62 (10.1)	2.37 (11.71)	.43 (4.02)	.383 (.185)	.436 (.175)	-.046 (.279)	.010 (.210)
DY (-2)	0	-.001(.033)	-.002 (.010)	-.000 (.003)	-.047 (.258)	-.010 (.080)	-.002 (.013)	-.003 (.037)	-.001 (.012)	.001 (.039)	-.002 (.010)
DY (-3)	0	-.002 (.028)	.000 (.007)	-.000 (.003)	.015 (.330)	-.021 (.159)	-.002 (.014)	.000 (.028)	.001 (.009)	-.001 (.029)	.000 (.007)
DY (-4)	0	.009 (.025)	.001 (.010)	-.000 (.003)	.004 (.168)	-.001 (.060)	.000 (.009)	.008 (.026)	.001 (.010)	.013 (.027)	-.001 (.010)
DY (-5)	0	.004 (.025)	-.002 (.009)	-.000 (.003)	.034 (.168)	-.001 (.082)	-.001 (.010)	.000 (.026)	-.002 (.010)	.000 (.029)	-.001 (.010)
Sum of coefficients	0	.008 (.071)	-.003 (.024)	-.001 (.007)	-1.61 (10.0)	2.33 (11.53)	.42 (4.00)	.389 (.187)	.436 (.173)	-.033 (.280)	.008 (.207)
Trend	0.09	.084 (.006)	.085 (.002)	.085 (.001)	.228 (.868)	-.114 (.972)	.049 (.340)	.052 (.015)	.048 (.015)	.087 (.024)	.084 (.018)
Standard error	0.100	.100 (.002)	.100 (.001)	.100 (.000)	.409 (.932)	.491 (1.102)	.249 (.332)				
R-squared	0.000	.004 (.002)	.000 (.000)	.000 (.000)	.001 (.001)	.000 (.000)	.000 (.000)				
Chi-squared (DF)	30									29.6 (7.0)	29.8 (6.7)
degrees of freedom	30									30	30
LM (1) Test	sig.							-9.38 (3.16)	-32.2 (9.8)	1.14 (6.93)	-1.25 (18.49)
LM (2) Test	insig.							.09 (.60)	-.07 (.62)	-.10 (.59)	-.05 (.58)
Sargan Test for FE	sig.									33.3 (10.4)	350.4 (29.9)
degrees of freedom	6									6	6

Table 4b
SIMULATION RESULTS - 50 Draws
Data Generation Model: Random Walk with Drift
Estimation Model: Fixed Effect, Linear Trend, and 5 Lags of Dependent Variable

# of time periods		20	20	20	20	20	20	20	20	20	20
# of observations	True	200	2000	10000	200	2000	10000	200	2000	200	2000
Estimation method	Model	OLS	OLS	OLS	IV	IV	IV	GMM 1Step	GMM 1Step	GMM 2Step	GMM 2Step
DY (-1)	0	.001 (.016)	-.000 (.005)	-.001 (.002)	3.10 (16.50)	-.02 (6.30)	.11 (3.44)	.679 (.190)	.753 (.203)	.009 (.106)	.026 (.150)
DY (-2)	0	.000 (.017)	-.001 (.005)	-.000 (.002)	-.017 (.173)	-.008 (.046)	-.001 (.007)	.001 (.019)	.005 (.012)	.001 (.022)	.001 (.006)
DY (-3)	0	.005 (.017)	.001 (.005)	-.000 (.002)	-.013 (.167)	-.001 (.029)	.001 (.006)	.008 (.017)	.007 (.011)	.007 (.018)	.000 (.006)
DY (-4)	0	.001 (.018)	-.000 (.004)	.000 (.002)	-.035 (.248)	-.001 (.016)	-.003 (.013)	-.002 (.024)	.001 (.006)	.001 (.023)	.000 (.005)
DY (-5)	0	-.003 (.016)	-.000 (.005)	-.001 (.002)	-.023 (.152)	.003 (.011)	-.000 (.005)	-.018 (.038)	-.021 (.041)	-.001 (.021)	.002 (.004)
Sum of coefficients	0	.004 (.045)	-.001 (.012)	-.002 (.004)	3.02 (15.78)	-.03 (6.36)	.11 (3.45)	.667 (.181)	.744 (.105)	.017 (.116)	.030 (.147)
Trend	0.085	.085 (.004)	.085 (.001)	.085 (.000)	-.17 (1.31)	.088 (.544)	.076 (.293)	.029 (.015)	.022 (.015)	.084 (.010)	.082 (.013)
Standard error	0.100	.100 (.001)	.100 (.000)	.100 (.000)	.406 (1.643)	.272 (.575)	.243 (.259)				
R-squared	0.000	.001 (.001)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)	.000 (.000)				
Chi-squared (DF)	108.0									110.0 (10.1)	107.4 (15.5)
degrees of freedom	108									108	108
LM (1) Test	sig.							-20.7 (8.7)	-57.3 (36.8)	-.53 (5.97)	-4.54 (26.60)
LM (2) Test	insig.							.004 (.685)	-.12 (.70)	.11 (1.00)	0.04 (0.47)
Sargan Test for FE	insig.									67.1 (10.2)	1356.5 (30.2)
degrees of freedom	6									6	6

The chi-squared shown is for the Sargan test of overidentification (instrument validity).
In all cases, a Hausman test for uncorrelated effects rejects.
The number in parentheses are the standard deviation of the estimate over the 50 simulation draws.
The columns with 10000 or 20000 observations are averages over 20 simulation draws.
Instruments are y(-2)-y(-6).

Table 5a
SIMULATION RESULTS - 50 Draws
Data Generation Model: AR1 with Deterministic Trend
Estimation Model: Fixed Effect, Linear Trend, and 5 Lags of Dependent Variable

# of time periods		6	6	6	6	6	6	6	6	6	6
# of observations	True	200	2000	20000	200	2000	20000	200	2000	200	2000
Estimation method	Model	OLS	OLS	OLS	IV	IV	IV	GMM 1Step	GMM 1Step	GMM 2Step	GMM 2Step
DY (-1)	0.99	.008 (.029)	.008 (.010)	.008 (.003)	1.03 (.27)	1.013 (.054)	.989 (.017)	.827 (.108)	.977 (.033)	.520 (.230)	.970 (.035)
DY (-2)	0	.009 (.033)	.009 (.010)	.009 (.003)	.001 (.042)	-.001 (.015)	.001 (.004)	.000 (.039)	-.001 (.015)	.002 (.040)	-.000 (.015)
DY (-3)	0	.008 (.028)	.011 (.007)	.010 (.003)	.000 (.036)	.002 (.012)	.001 (.005)	.004 (.033)	.002 (.012)	.004 (.031)	.001 (.012)
DY (-4)	0	.019 (.024)	.012 (.010)	.010 (.003)	.004 (.032)	.001 (.012)	-.000 (.004)	.006 (.027)	.001 (.011)	.013 (.028)	.001 (.011)
DY (-5)	0	.014 (.024)	.009 (.009)	.010 (.003)	-.002 (.035)	-.002 (.013)	-.000 (.004)	.002 (.030)	-.001 (.012)	.005 (.035)	-.001 (.012)
Sum of coefficients	0.99	.058 (.065)	.049 (.024)	.047 (.007)	1.032 (.251)	1.013 (.048)	.990 (.017)	.839 (.108)	.978 (.032)	.544 (.214)	.971 (.032)
Trend	0.01	.070 (.005)	.071 (.002)	.071 (.001)	.007 (.017)	.009 (.003)	.010 (.001)	.020 (.007)	.011 (.002)	.039 (.014)	.011 (.002)
Standard error	0.263	.188 (.004)	.188 (.001)	.188 (.000)	.270 (.039)	.269 (.007)	.263 (.002)				
R-squared	0.000	.005 (.002)	.001 (.001)	.001 (.000)	.001 (.001)	.000 (.000)	.000 (.000)				
Chi-squared (DF)	30									38.1 (9.2)	33.3 (8.5)
degrees of freedom	30									30	30
LM (1) Test	sig.							-13.9 (0.7)	-44.6 (0.5)	-10.9 (3.9)	-44.6 (0.5)
LM (2) Test	insig.							0.17 (0.69)	-.07 (.63)	0.07 (0.85)	-.10 (0.65)
Sargan Test for FE	sig.									-0.8 (7.3)	3.2 (5.2)
degrees of freedom	6									6	6

Table 5b
SIMULATION RESULTS - 50 Draws
Data Generation Model: AR(1) with Deterministic Trend
Estimation Model: Fixed Effect, Linear Trend, and 5 Lags of Dependent Variable

# of time periods		20	20	20	20	20	20	20	20	20	20
# of observations	True	200	2000	10000	200	2000	10000	200	2000	200	2000
Estimation method	Model	OLS	OLS	OLS	IV	IV	IV	GMM 1Step	GMM 1Step	GMM 2Step	GMM 2Step
DY (-1)	0.99	.043 (.016)	.042 (.006)	.042 (.002)	1.148 (.401)	.980 (.059)	.984 (.028)	.795 (.278)	.599 (.314)	.651 (.107)	.947 (.015)
DY (-2)	0	.043 (.016)	.042 (.004)	.043 (.002)	-.009 (.027)	.000 (.008)	.001 (.004)	-.033 (.286)	.059 (.253)	.007 (.030)	.003 (.010)
DY (-3)	0	.048 (.017)	.044 (.005)	.044 (.003)	-.003 (.031)	.002 (.007)	.000 (.004)	.031 (.345)	.080 (.436)	.014 (.028)	.002 (.011)
DY (-4)	0	.045 (.017)	.043 (.004)	.044 (.002)	-.011 (.038)	-.001 (.007)	.000 (.004)	-.024 (.349)	.023 (.251)	.010 (.028)	.003 (.011)
DY (-5)	0	.041 (.015)	.044 (.005)	.045 (.002)	-.012 (.032)	.001 (.007)	.000 (.003)	.187 (.245)	.224 (.159)	.010 (.029)	.004 (.011)
Sum of coefficients	0.99	.220 (.037)	.216 (.010)	.217 (.005)	1.114 (.319)	.982 (.049)	.985 (.023)	.957 (.053)	.985 (.027)	.692 (.091)	.958 (.013)
Trend	0.01	.199 (.009)	.200 (.002)	.199 (.001)	-.021 (.078)	.012 (.012)	.011 (.006)	.024 (.014)	.021 (.011)	.084 (.023)	.018 (.003)
Standard error	0.263	.191 (.002)	.191 (.001)	.191 (.000)	.288 (.062)	.262 (.008)	.263 (.004)				
R-squared		.013 (.004)	.012 (.001)	.012 (.001)	.003 (.002)	.003 (.001)	.003 (.000)				
Chi-squared (DF)	108.0									128.5 (19.0)	123.4 (13.7)
degrees of freedom	108									32	114
LM (1) Test	sig.							-17.1 (6.2)	-37.4 (23.9)	-17.1 (1.4)	-59.8 (0.5)
LM (2) Test	insig.							-.15 (10.17)	-1.4 (36.2)	-.12 (.46)	0.13 (0.39)
Sargan Test for FE	insig.									87.9 (12.8)	
degrees of freedom	6									6	6

The chi-squared shown is for the Sargan test of overidentification (instrument validity).
In all cases, a Hausman test for uncorrelated effects rejects.
The number in parentheses are the standard deviation of the estimate over the 50 simulation draws.
The columns with 10000 or 20000 observations are averages over 20 simulation draws.
Instruments are y(-2)-y(-6).

TABLE 6
Testing for Nonstationarity
Simulated Data - 50 Draws

	T=12; N=200											
	IPS Test; no trend						IPS Test with trend					
Data Generating Process *(truth)*	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size
No. of Augmenting Lags (p)	2		3		Best		2		3		Best	
Random Walk	-0.02	0.88	-2.00	0.48	-0.19	0.90	-0.36	0.78	-0.16	0.94	-0.21	0.92
(accept)	(1.31)		(2.60)		(1.06)		(1.57)		(0.97)		(1.11)	
Firm Fixed Effects	-5.86	0.00	-5.32	0.02	-4.97	0.00	-2.89	0.20	-1.72	0.46	-2.04	0.32
(reject)	(1.38)		(2.09)		(1.12)		(1.29)		(0.79)		(0.92)	
AR(1)	0.06	0.88	-1.96	0.44	-0.13	0.90	-0.36	0.80	-0.16	0.94	-0.21	0.92
(reject)	(1.30)		(1.93)		(1.05)		(1.61)		(0.99)		(1.14)	

	T=12; N=2000											
	IPS Test; no trend						IPS Test with trend					
Data Generating Process *(truth)*	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size
No. of Augmenting Lags (p)	2		3		Best		2		3		Best	
Random Walk	-0.67	0.68	-6.64	0.00	-1.10	0.64	-0.80	0.72	-0.29	0.96	-0.43	0.94
(accept)	(1.57)		(2.83)		(1.28)		(1.28)		(0.79)		(0.91)	
Firm Fixed Effects	-18.01	0.00	-17.82	0.00	-15.32	0.00	-8.31	0.00	-4.92	0.00	-5.85	0.00
(reject)	(1.22)		(2.29)		(0.99)		(1.15)		(0.71)		(0.81)	
AR(1)	-0.57	0.76	-6.77	0.00	-1.02	0.66	-0.80	0.68	-0.30	0.96	-0.43	0.96
(reject)	(1.49)		(2.69)		(1.22)		(1.24)		(0.76)		(0.88)	

	T=20; N=200;Number of Draws=20											
	IPS Test; no trend						IPS Test with trend					
Data Generating Process *(truth)*	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size	T-bar Statistic	Empirical Size
No. of Augmenting Lags (p)	2		3		Best		2		3		Best	
Random Walk *(accept)*	1.50 (1.05)	1.00	1.53 (1.53)	1.00	1.34 (1.34)	1.00	1.91 (1.17)	1.00	1.54 (1.35)	1.00	1.71 (1.12)	1.00
Firm Fixed Effects *(reject)*	-10.15 (1.08)	0.00	-7.45 (0.85)	0.00	-10.07 (1.06)	0.00	-5.23 (1.12)	0.00	-2.97 (1.07)	0.15	-5.19 (1.09)	0.00
AR(1) *(reject)*	1.37 (1.16)	1.00	1.35 (1.22)	1.00	1.2 (1.13)	1.00	1.90 (1.15)	1.00	1.52 (1.37)	1.00	1.70 (1.11)	1.00

The empirical size is the fraction of the time that the null of nonstationarity is rejected at the five percent level.

The numbers in parentheses are the standard deviation of the test statistics over the 50 draws.

The number of augmenting lags are the number used when computing the IPS test statistic. "Best" means 2 or 3, depending on what the individual data series prefer.