ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 55-56 – 1999

Spurious Stochastics in a Short Time-Series Panel Data

Clive W.J. GRANGER, Namwon HYUNG*

ABSTRACT. – This paper analyzes the effects of individual-specific size factors in a dynamic panel regression model. Theory and simulation show that an individual-specific size factor, with a fat-tailed distribution or a time-varying property, may cause spurious stochastics. If a pair of panel variables depends on size in some way, then they appear to find a strong relationship, if the size variable is not used in the regression, even if the variables are otherwise independent. Moreover, forecasts based on models that have omitted size-factors are affected seriously by the property of the size-factors. A pooling regression with very short time-series appears to fit well in sample, but forecasts poorly out-of-sample if the neglected individual-specific size-factor has a fat-tailed distribution.

Inférence spécieuse dans les modèles de panel avec une courte dimension temporelle

RÉSUMÉ. – Ce papier analyse les effets des facteurs de taille individuels dans les modèles de panel dynamiques. La théorie et les simulations montrent qu'un facteur de taille individuel avec une distribution de queue épaisse ou des propriétés d'évolution dans le temps peut entraîner des inférences spécieuses. Si une paire de variables dans un panel dépend d'une certaine facon d'un facteur de taille, alors la régression entre les deux variables fait apparaître une dépendance forte des deux variables, si la variable de taille n'est pas présente dans la régression et ce, même si les variables sont, hormis cet effet, indépendantes. De plus, les prévisions, basées sur des modèles dans lesquels les facteurs de taille ont été omis sont affectées sérieusement par cet effet. De même, une régression sans effet individuel avec des séries temporelles très courtes semble s'adapter correctement sur les données mais produit des prévisions de mauvaises qualités si les effets de taille individuels négligés ont une distribution à queue épaisse.

* C.W.J. GRANGER: University of California; N. HYUNG: University of California.
The authors thank P. SEVESTRE (the editor) and two anonymous referees for their valuable comments. The second author also thanks C. HAEFKE and D. MCCUBBIN.

1 Introduction

We will consider a panel that has a large cross-section dimension (N) but a small time dimension (T), possibly just a single digit. Suppose that the variables of interest vary with an individual-specific size variable which is either constant, such as area, or time-varying, such as population or total income. It is a property of many panels that the distribution of these size variables is wide-spread or fat-tailed, such as a panel of countries including Luxembourg and Cyprus as well as India and China. If X_{it}, Y_{it}, are a pair of variables measured from i-th region at time t, and if each depends on size in some way, then a regression of X_{it} on Y_{it} will appear to find a strong relationship, if the size variable is not used in the regression, even if the variables are otherwise independent. This would be an example of a "*spurious relationship*" occurring from mis-specification. There are, of course, a number of well-known ways of dealing with this problem, which we discuss below. However, the problem is more difficult if the size variable is time-varying or a model has dynamic terms. If one regresses X_{it} on X_{it-1} and the size variable is either constant or very slowly-changing, one will get a high correlation between X_{it} and its lag, suggesting, incorrectly, a unit root. This paper investigates this question of a "*spurious stochastic*" relationship, using some theory but mostly by simulations. We consider the effects of alternative classical resolutions of the difficulty and particularly consider the implications for long-run forecasting quality on the alternative approaches that are available.

Section 2 explains how spurious results could occur in a regression of miss-pecified models, and section 3 applies similar methods to dynamic panel regressions to illustrate spurious stochastics. Monte Carlo methods are adopted to investigate these problems in sections 4 and 5, using the results of a simulation study for estimation, unit root tests and forecast comparison, and the final section concludes.

2 Regressing Two Panel Series

Suppose that S_i is a size variable (such as area, population, total income) and that X_{it}, Y_{it} are a pair of variables, for region i and time t, which are given by

$$\begin{aligned} X_{it} &= f_i(S_i) + \tilde{X}_{it} \\ Y_{it} &= A f_i(S_i) + \tilde{Y}_{it} \end{aligned} \tag{1}$$

where $\tilde{X}$, $\tilde{Y}$ are not related to size. Here $f_i(S_i)$ is some function of the size.

If one runs the panel regression

$$Y_{it} = \alpha + \beta X_{it} + e_{it} \tag{2}$$

then if the size variable has a wide range [1] (*e.g.*, panel includes both very small and very large countries) one will get an apparent relationship. One could call this is spurious relationship if a regression of

$$\tilde{Y}_{it} = \alpha + \beta \tilde{X}_{it} + \tilde{e}_{it} \tag{3}$$

has $R^2 = 0$. The apparent relationship (2) occurs because of the missing variable S_i in the regression. However, terms such as $f_i(S_i)$ are actually (or effectively) fixed effects (because they either do not change with time, or they change very slowly).

PROPOSITION 1: Suppose (2) is estimated by least-squares with the following assumptions,
(a) $\tilde{X}_{it} = \varepsilon^x_{it}, \tilde{Y}_{it} = \varepsilon^y_{it}$ where $\varepsilon^x_{it}, \varepsilon^y_{it}$ are independent and i.i.d. with variances σ^2_x, σ^2_y.
(b) $f_i(S_{it}) \equiv S_i$, is drawn (independently) from a fat-tailed distribution with (μ_s, σ^2_s) then we obtain the following results, for a larger variance of size variable,
(a) $\beta \approx A$
(b) the *t*-statistics on β will be significant,
(c) R^2 large

PROOF: See Appendix A.

One could improve (2) by including fixed effects and then running the regression

$$Y_{it} = \alpha_i + \beta X_{it} + e_{it}. \tag{4}$$

If there are *N* regions, we now have to estimate $N + 1$ parameters to estimate in (4) (apart from variances) whereas in (2) there are only two parameters. If the data set is $i = 1,...,N, t = 1,...,T$, with *N* large, *T* small then the α_i's will be poorly estimated. The quality of forecast will be not good in this case. [2]

Suppose that one could "*size adjust*" each variable

$$\begin{aligned} X^{SA}_{it} &= X_{it} - \overbrace{f_i(S_i)} \\ Y^{SA}_{it} &= Y_{it} - \overbrace{Af_i(S_i)} \end{aligned} \tag{5}$$

where $\overbrace{\quad}$ indicates estimates. One can now run a regression like (3)

$$Y^{SA}_{it} = \widehat{\alpha} + \widehat{\beta} X^{SA}_{it} + \widehat{e}_{it} \tag{6}$$

1. This condition is equivalent to $\mathrm{corr}(X_{it}, f_i(S_i))$ close to one. Under this condition, $\beta \approx A$, since $E[\widehat{\beta}_{OLS}] \approx A\,\mathrm{Cov}(X_{it}, f_i(S_i))/\mathrm{Var}(X_{it} = A[\mathrm{corr}(X_{it}, f_i(S_i))]^2$.
2. In the table 5, the mean squared forecast error of (4) when $h = 1$, $T = 10$, is larger than that for other models.

and one would expect $\widehat{\beta} \approx 0,\ R^2 \approx 0$. In a sense, (2) is a spurious regression in that the inference can suggest that there is a true relationship whereas in fact there is none.

When the size variable is not a constant, the situation is more complicated. If terms such as $f_i(S_{it})$ change very slowly, there will be still an apparent spurious relationship in (4) because of the missing variable in this regression, which can not be captured by the fixed effects specification. As a related work, ENTORF [1996] investigates the nonsense regression between two independent panel data with cross-sections of long time series. He showed that if the least-squares dummy variable approach is applied to a mistakenly conjectured fixed-effects model when in fact the data generating process for two series are independent random walks with drift, one can get a spurious fixed-effects models. We investigate short panels with a large number of individuals and the DGP for two series have a common size factor which follows random walks with drift. In this case the fixed effect model (4) also has a misspecification problem. So this model cannot capture the time-varying size factor properly, rather it shows a spurious relation, $\beta \approx A$, between two series because of the size factor. This is summarized as follows,

PROPOSITION 2: Suppose (2) or (4) is estimated by least-squares panel regression when the size variable follows a data generating process: $f_i(S_{it}) \equiv S_{it} = \delta + S_{it-1} + \eta_{it}$, where $\eta_{it} \sim \text{iid}\,(0, \sigma_\eta^2)$ and $S_{io} = 0$ for all i for simplicity, then $\beta \approx A$ for each regression.

PROOF: See Appendix A.

Suppose the size variable is a trend-stationary process, then we would have a similar result as before. The proof is analogous to that of the proof of proposition 2 and is therefore skipped.

3 Spurious Stochastics in a Dynamic Panel Model

For a particular model suppose that the explanatory variables for Y_{it} are a vector $\underline{X}_{it}$ which can include lagged Y_{it}. In this section it will be assumed for reasons of representational clarity that $\underline{X}_{it}$ consists of only Y_{it-1}. Results do not change qualitatively for more general assumptions.

Let Y_{it} be generated as above, then clearly if one regressed Y_{it} on Y_{it-1}

$$Y_{it} = m + \rho\, Y_{it-1} + e_{it} \tag{7}$$

then one will get a spurious unit root.

PROPOSITION 3: Suppose (7) is estimated by least-squares with following assumptions; $Y_{it} = S_i + \tilde{Y}_{it}$, where $\tilde{Y}_{it} = \rho\tilde{Y}_{it-1} + \varepsilon_{it}, \varepsilon_{it} \sim$ iid $(0, \sigma_\varepsilon^2)$, $|p| < 1$ and S_i is drawn (independently) from a fat-tailed distribution with (μ_s, σ_s^2). Then one will get a spurious unit root, for a size variable with a larger variance,

(a) $\rho \approx 1$ [3]

(b) the t-statistics on $\rho = 1$ will be significant

(c) $m \approx 0$

(d) $R^2 \approx 1$

PROOF: See Appendix A.

When there is a spurious unit root in the system, the resulting model will fit the data well but may forecast badly out of sample. Note that if Y_{it-1} is replaced by Y_{it-k}, similar results are obtained but with different dynamics. Because the data does not contain a unit root, this size-produced phenomenon can be called "*spurious stochastics*".

If the size variables, $f_i(S_i)$, are fixed effects, one could improve (7) by including individual-specific fixed effects, like (4).

$$Y_{it} = m_i + \rho Y_{it-1} + e_{it} \tag{8}$$

As with (4), there are similar problems with large *N* and small *T*. There is an additional problem caused by short time series in a dynamic panel data with individual specific effects, *i.e.*, a downward bias [4] of ρ.

When the size variable is time-varying, there will still be an apparent spurious relationship in (8) because of neglected time-varying size variable, which the fixed effects can not capture. If the size variable is time-varying stochastically (*e.g.*, random walk with drift) and (8) is estimated by least squares regression, then $\rho \approx 1$. The proof of this follows directly from proposition 2.

4 Estimation and Unit Root Test of Panel Models: Simulation Study

In our simulation design, we consider two types of individual-specific factors, *i.e.*, constant and time-varying factors, as data generating processes. A

3. If $\text{corr}(X_{it}, f_i(S_i)) \approx 1$, then $\widehat{\rho}_{OLS} \approx 1$, since $E[\widehat{\rho}_{OLS}] \approx \rho + (1-\rho)[\text{corr}(X_{it}, f_i(S_i))]^2$.

4. See NICKELL [1981] for the detail of this. He calculated the bias for small *T*. Let the DGP for Y_{it}, be

$$Y_{it} = m_i + \rho Y_{it-1} + \varepsilon_{it} = \frac{m_i}{1-\rho} + \sum_{s=0}^{\infty} \rho^s \varepsilon_{it-s}.$$

Then the asymptotic bias of least squares estimator of ρ is

$$\underset{N\to\infty}{p\lim}(\widehat{\rho} - \rho) = -\frac{1+\rho}{T-1}\left(1 - \frac{1-\rho^T}{T(1-\rho)}\right)\left\{1 - \frac{2\rho}{(1-\rho)(T-1)}\left[1 - \frac{1-\rho^T}{T(1-\rho)}\right]\right\}^{-1}$$

size variable, like area, can be represented well by a constant individual-specific factor, whereas population or total income are classified as time-varying individual-specific factors. In this section, spurious estimation results are presented which arise from the misspecification caused by the omitted variable problem of individual specific effects. The data generating processes of $\{Y_{it}\}$ are designed as followings: DGP 1 characterizes panel data with constant individual-specific effects. To replicate serial dependency of economic variables, we assume size-free component, $\tilde{Y}_{it}$ follows autoregressive process. To avoid the problem of fixing Y_{i0}, 50 pre-samples are generated.

DGP 1 (Panel Data with Constant Individual-Specific Effects): $Y_{it} = S_{it} + \tilde{Y}_{it}$, where $\tilde{Y}_{it} = 0.7\tilde{Y}_{it-1} + \varepsilon_{it}$, $\varepsilon_{it} \sim$ i.i.d. $N(0,\sigma_\varepsilon^2)$, and $S_{it} = S_i \sim \chi^2_{(12)}$ or $\chi^2_{(24)}$, so that across regions the constant size variable is drawn from a χ^2 distribution with large variance.

DGP 2 (Panel Data with Time-varying Individual-Specific Effects): ***DGP 1*** with $S_{it} = \delta + S_{it-1} + \eta_{it}$, S_{i0} is drawn from a $\chi^2_{(6)}$ distribution. $\delta = 0.05$ or 0.5, and η_{it} is drawn from i.i.d. $N(0,\sigma_\eta^2)$ where $\sigma_\eta^2 = 0.1$ or 1.0.

Two regression models are considered: Pooling Panel Regression; $Y_{it} = m + \rho Y_{it-1} + e_{it}$ (Model 1) and Panel Regression with Individual-Specific Fixed Effects; $Y_{it} = m_i + \rho Y_{it-1} + e_{it}$, (Model 2). Panel series with several different size components are investigated in the following tables, which give averaged simulation values based on a sample size of $N = 50$, 100, 500 and $T = 5$, 10, 25, 50, with 1,000 replications. [5]

The estimates for models when the data has individual-specific fixed effects are given in table 1. The estimated value of $\widehat{\rho}$ in model 1 shows a clear tendency to unity (0.98 → 0.99) as the variance of the size factor (S_i) increases from 24 to 48. For the smaller number of cross-sections, *i.e.* N = 50 and N = 100, they have similar properties. Changes of time-series observation and cross-section do not have any significantly different effects on the estimation; only the variance of size factor does. When $T = 5$, as expected from the results of section 3, even though model 1 is misspecified, it fits very well in the terms of R^2, but forecasts badly out-of-sample (see Table 5). A misspecified model (model 1), ignoring constant individual-specific factors, has $\widehat{\rho}$ near unity. Here, a simple interpretation of $\widehat{\rho}$ clearly leads to spurious stochastics. To examine this possibility statistically, a power analysis of a unit root test is conducted in this section. From model 2, it is seen that the omitted variable problem can be removed by using fixed effects, although there still is a small-sample bias in $\widehat{\rho}$. For the very small T, even though it is a correct model, there is a severe downward bias as pointed out by NICKELL [1981] and very poor predictability as can be seen from Table 5 in section 5.

Panel regressions for panel data with time-varying individual-specific factors are considered in table 2. The choice of δ and σ_η covers several specific drift-noise ratios (0.05, 0.16, 0.50, 1.58). This choice of parameter

5. We do not report results of all of these combinations to save space. Thc simulation results of the rest can be requested from the authors. We need to note that the roles of T and N in the simulation. The ratio of T, and N does not have any significant effects on the estimates.

TABLE 1
Estimation of Panel Data with Individual-Specific Fixed Effects

Model 1		$T=5$	$T=10$	$T=25$	$T=50$
	m	0.2727	0.2753	0.2732	0.2737
CASE 1	ρ	0.9773	0.9771	0.9772	0.9771
Var(S_i) = 24	s.e.	0.0047	0.0032	0.0019	0.0014
	R^2	0 9551	0.9548	0.9549	0.9548
	m	0.2820	0.2859	0.2844	0.2839
CASE 2	ρ	0.9883	0.9881	0.9881	0.9882
Var(S_i) = 48	s.e.	0.0034	0.0023	0.0014	0.0010
	R^2	0.9765	0.9764	0.9764	0.9765
Model 2					
	$\overline{m}$	9.4341	6.2300	4.5225	4.0337
CASE I	ρ	0.2145	0.4805	0.6230	0.6634
Var(S_i) = 24	s.e.	0.0252	0.0139	0.0073	0.0048
	R^2	0.9768	0.9675	0.9632	0.9620
	$\overline{m}$	18.8429	12.4609	9.0474	8.0769
CASE 2	ρ	0.2145	0.4805	0.6230	0.6634
Var(S_i) = 48	s.e.	0.0252	0.0139	0.0073	0.0048
	R^2	0.9879	0.9831	0.9809	0.9804

Note: Individual-Specific Effects (S_i) for CASE 1 and CASE 2 are drawn from $\chi^2-(12)$ and $\chi^2_{(24)}$, respectively. The results are based on 500 cross-section and 1.000 replications. $\overline{m}$ denotes averaged value of m_i, for all individuals.

values represents certain economically typical DGP, since it could cover some range of the empirical findings of HYLLEBERG and MIZON [1989]. [6] The influence on ρ is arising from the drift and noise of the neglected size variable. From proof of proposition 2, we conjecture σ_η has less impact on ρ in regressions (7) and (8) than δ.[7] Even though the error (η_{it}) of the size-factor has a very small variance, say 0.1, it will dominate the estimation in large samples, and drive $\widehat{\rho}$ to one, since regression (8) also has an omitted variable problem. Without considering the time-varying size-factor properly, one cannot get the true dynamic structure of $\tilde{Y}_{it}$, which is not related to the size component. In a small sample, the estimated values have a slightly larger bias in every model than the estimated values of DGP 1, since DGP 2 contains a random walk component. However, since all of these models are misspecified, they have very different multi-step forecastability compared with the DGP I case as will be shown in section 5.

In tables 1 and 2, misspecified panel models for individual-specific effects may produce "*Spurious Stochastics*". To examine this statistically, we conduct

6. They estimated the magnitude of the ratio of drift-noise for some economic time series and found values between – 0.14 and 0.78. We thank one anonymous referee for suggesting this paper.
7. The speed of convergence to 1 for ρ in (7) depends on $T^{-1}\sigma^2_\varepsilon/\sigma^2_\eta$ and $T^2\sigma^2_\varepsilon/\delta^2$. The speed of convergence of ρ to 1 in (8) depends on $\sigma^2_\varepsilon/\sigma^2_\eta$ and $T^{-1}\sigma^2-\varepsilon/\delta^2$.

TABLE 2
Estimation of Panel with Time-varying Individual-Specific Effects

Model 1		$T = 5$	$T = 10$	$T = 25$	$T = 50$
	m	0.3065	0.3056	0.3056	0.3026
CASE 1	ρ	0.9579	0.9588	0.9611	0.9649
	s.e.	0.0067	0.0044	0.0026	0.0017
	R^2	0.9112	0.9129	0.9177	0.9255
	m	0.2858	0.2555	0.2028	0.1611
CASE 3	ρ	0.9613	0.9669	0.9768	0.9846
	s.e.	0.0084	0.0052	0.0027	0.0015
	R^2	0.8681	0.8856	0.9183	0.9449
Model 2					
	$\overline{m}$	4.6979	3.0698	2.0598	1.5461
CASE 1	ρ	0.2350	0.5134	0.6945	0.7920
	s.e.	0.0251	0.0136	0.0067	0.0039
	R^2	0.9533	0.9358	0.9296	0.9324
	$\overline{m}$	4.0768	2.1803	1.0330	0.7178
CASE 2	ρ	0.4702	0.7902	0.9547	0.9879
	s.e.	0.0228	0.0097	0.0028	0.0010
	R^2	0.9456	0.9320	0.9544	0.9810

Note: CASE 1 $\delta = 0.05$, $\sigma_\eta^2 = 0.1$; CASE 2 $\delta = 0.5$, $\sigma_\eta^2 = 0.1$; CASE 3 $\delta = 0.05$, $\sigma_\eta^2 = 1.0$. The results are based on 500 cross-section and 1.000 replications.

TABLE 3
Power Analysis of Panel Unit Root Test under H_a; $\rho = 0.7$
Panel Data with Individual-Specific Fixed Effects

Model 1	$T = 5$	$T = 10$	$T = 25$	$T = 50$
$N = 50$	37.8	79.1	100.0	100.0
	20.8	41.7	97.0	100.0
$N = 100$	71.4	99.4	100.0	100.0
	42.0	82.8	100.0	100.0
$N = 500$	99.9	100.0	100.0	100.0
	98.0	100.0	100.0	100.0
Model 2				
$N = 50$	34.3	98.1	100.0	100.0
	34.3	98.1	100.0	100.0
$N = 100$	38.2	99.9	100.0	100.0
	38.2	99.9	100.0	100.0
$N = 500$	100.0	100.0	100.0	100.0
	100.0	100.0	100.0	100.0

Note: The first numbers equals the power when $S_i \sim \chi^2_{(12)}$ and the number in the second rows is the power when $S_i \sim \chi^2_{(24)}$ with 5% critical values from LEVIN and LIN [1992]. The results are based on 1.000 replications.

a power analysis of the panel unit root tests suggested by LEVIN and LIN [1992]. Initially the power of the test is considered in the presence of stationary errors where critical values for the tests are available from LEVIN and LIN [1992].

When data has the individual-specific fixed effects, the panel unit root test of model 1 has low power for a short panel and has proper power for a large *T*.

The low power of the panel unit root test in a short panel suggests that stationary panel series could be incorrectly specified as unit root processes with non-zero probability. The problem can persist with larger *T*. If the variance of the size-factor is increasing as *T* increases, say $\chi^2_{(12)} \rightarrow \chi^2_{(24)}$, then "*spurious stochastics*" will be found more frequently as suggested by table 3, *i.e.*, our results indicate less power for larger variance of size-factor. Model 2, in spite of a severe downward bias, has proper power for a moderate size of *T*. The power of model 2 is not affected by the variance of the size-factor, since the individual-specific terms can remove the effect of the size-factor.

In table 4, the panel unit root test for the data with time-varying individual specific factors is reported. The power is lower than in the case of table 3 because the time-varying size-factor includes a mild unit root component. As the variance of the size component increases from 0.1 to 1.0, the power of the

TABLE 4

Power Analysis of Panel Unit Root Test under H_a; ρ = 0.7
Panel Data with Individual-Specific Time-Varying Effects Case

Model 1	$T = 5$	$T = 10$	$T = 25$	$T = 50$
$N = 50$	62.9	95.4	100.0	100.0
	35.6	64.6	88.6	95.2
$N = 100$	89.2	100.0	100.0	100.0
	62.9	91.3	99.4	100.0
$N = 500$	100.0	100.0	100.0	100.0
	100.0	100.0	100.0	100.0
Model 2				
$N = 50$	25.7	85.3	100.0	100.0
	9.4	26.7	76.3	91.8
$N = 100$	28.1	97.2	100.0	100.0
	7.5	35.3	94.4	99.3
$N = 500$	100.0	100.0	100.0	100.0
	100.0	100.0	100.0	100.0

Note: The first numbers equals the power when $\delta = 0.05$, $\sigma^2_\eta = 0.1$. The number in the second rows is the power when $\delta = 0.05$, $\sigma^2_\eta = 1.0$ with 5% critical values from LEVIN and LIN [1992]. The results are based on 1.000 replications.

test is lower. [8] When T is large, the constant term in the size-component equation dominates. In that case, individual series show a clear time trend, so preference is given to a model which contains a time trend.

5 Forecast Comparison of Panel Models

One clear way to judge the relevance of a model is to ask how well it forecasts compared to other models. In this section the relative forecasting performance of three models is compared. We explain briefly the evaluation techniques for panel model forecasts examined in GRANGER and HUANG [1997] and then compare three panel models using simulations.

5.1. Evaluation of Forecasts of Panel Models

Consider any pair of models which produce forecasts and hence forecast errors. Let $e_{it}^{(1)}, e_{it}^{(2)}, i = 1,\ldots,N, t = 1,\ldots,T$, denote the period-$t$ forecast errors from models (1) and (2), respectively. If the loss function is quadratic and the forecast errors are unbiased and normally distributed, then the null hypothesis of equal forecast accuracy corresponds to equal forecast error variances:

$$H_0 : E[(e_{it}^{(1)})^2] = E[(e_{it}^{(2)})^2]$$

Define $x_{it} = e_{it}^{(1)} + e_{it}^{(2)}$ and $z_{it} = e_{it}^{(1)} - e_{it}^{(2)}$. A test of the previous null hypothesis is easily conducted using x_{it} and z_{it}, as $E[(e_{it}^{(1)})^2] = E[(e_{it}^{(2)})^2]$ only when $\mathrm{Cov}(x_{it}, z_{it})$ is zero.

$$H_0 : \mathrm{Cov}(x_{it}, z_{it}) = 0$$

There is a well-known test ("*sign test*") which is based on p_1 and p_2 where p_1 is the number of times $e_{it}^{(1)}$ is greater than $e_{it}^{(2)}$ and p_2 is the number of times $e_{it}^{(2)}$ is greater than $e_{it}^{(1)}$. The total number of comparisons is $Q =$

8. Since the values of $T^{-1}\sigma_\varepsilon^2/\sigma_\eta^2$ in the model 1 and $\sigma_\varepsilon^2/\sigma_\eta^2$ in the model 2 are getting smaller as σ_η^2 increases, estimated values of ρ of both models are closer to 1 as expected from footnote 7. And, of course, the loss of power of model 1 is much larger than model 2 since the convergence speed of model 1 is faster than model 2 as σ_η^2 increases.

$p_1 + p_2$. Under the null, the test statistic for the sign test is:

$$\left(\frac{p_1}{Q} - \frac{1}{2}\right) \Big/ \sqrt{1/4Q} \sim N(0,1) \tag{11}$$

Another test of the same null hypothesis is an extension of the sum-difference test GRANGER and NEWBOLD [1986] p. 279). Under the assumption that the individual forecast is unbiased, the forecast errors are not autocorrelated and independent across individuals, the sum-difference test can be extended to panel models. The test statistics are

$$\frac{\sqrt{NT-1}}{2} \log\left(\frac{1+r}{1-r}\right) \sim N(0,1) \tag{12}$$

where

$$r = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(e_{it}^{(1)} + e_{it}^{(2)})(e_{it}^{(1)} - e_{it}^{(2)})}{\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(e_{it}^{(1)} + e_{it}^{(2)})^2 \sum_{i=1}^{N}\sum_{t=1}^{T}(e_{it}^{(1)} - e_{it}^{(2)})^2\right]^{1/2}}$$

For multiple (h-step-ahead) forecast horizons, a sequence of forecast errors will in general follow a moving average (MA) process of order $(h-1)$, which violates the assumption that the forecast errors are serially uncorrelated as used in the sum-difference test. The no-serial-correlation assumption may be relaxed by using the MEESE-ROGOFF [1988] test. The MEESE-ROGOFF test can be extended to panel models by positing independence across observations (For details see Appendix B):

$$\frac{\gamma_{xz}}{\sqrt{\Sigma/NT}} \sim N(0,1) \tag{13}$$

where

$$\widehat{\gamma}_{xz} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}z_{it}\,,$$

$$\widehat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{\tau=-h+1}^{h-1}\left(1 - \frac{|\tau|}{T}\right)\{\widehat{\gamma}_{x_i x_i}(\tau)\widehat{\gamma}_{z_i z_i}(\tau) + \widehat{\gamma}_{x_i z_i}(\tau)\widehat{\gamma}_{z_i x_i}(\tau)\}\right],$$

$$\widehat{\gamma}_{x_i x_i}(\tau) = \frac{1}{T}\sum_{t=1}^{T-\tau} x_{it}x_{it+\tau},\ \widehat{\gamma}_{x_i z_i}(\tau) = \frac{1}{T}\sum_{t=1}^{T-\tau} x_{it}z_{it+\tau},$$

and $\widehat{\gamma}_{z_i z_i}(\tau) = \frac{1}{T}\sum_{t=1}^{T-\tau} z_{it}z_{it+\tau}$

All of these tests have a standard normal distribution asymptotically, but in small samples, they will have different properties. In the univariate case, DIEBOLD and MARIANO [1995] show that the MEESE-ROGOFF test is robust to serial correlation in large samples, but it is oversized in small samples in the presence of serial correlation. The sum-difference test will be affected by serial correlation which pushes the empirical size above the nominal size; however, in small samples, this is better than the MEESE-ROGOFF test. We report the results of the MEESE-ROGOFF test. [9]

5.2. Forecasts of Panel Models: Simulation Study

To investigate the effects of the variance and the constant term in the size component equation, several types of fixed and time-varying individual-specific factors were simulated. The results in tables 5 and 6 show the mean

TABLE 5

Forecasting Comparison of the Panel Regressions of Panel Data with Individual-Specific Fixed Effects

h-step	Mean Squared Forecast Errors			1 vs.2	1 vs. RW	2 vs. RW
obs	Model 1	Model 2	RW	MR	MR	MR
$h = 1$			**CASE 1**			
$T = 10$	1.1627	1.3459	1.1758	98.9*(0.1+)	98.2(0)	0.1(97.6)
$T = 50$	1.1626	1.0277	1.1764	0(100)	100(0)	100(0)
$h = 5$						
$T = 10$	3.2373	2.8364	3.2648	0(71.0)	14.5(0.3)	74.5(0)
$T = 50$	3.2117	2.0757	3.2620	0(100)	2.3(0)	100(0)
$h = 20$						
$T = 10$	6.2018	2.9150	3.9184	0(100)	0(98.5)	98.9(0)
$T = 50$	6.0192	2.1839	3.9255	0(100)	0(99.8)	100(0)
$h = l$			**CASE 2**			
$T = 10$	1.1701	1.3499	1.1767	97.6(0)	82.5(0)	0(96.7)
$T = 50$	1.1690	1.0273	1.1761	0(100)	97.5(0)	100(0)
$h = 5$						
$T = 10$	3.2653	2.8426	3.2678	0(73.5)	9.3(1.8)	73.0(0)
$T = 50$	3.2383	2.0744	3.2609	0(100)	0.2(0)	100(0)
$h = 20$						
$T = 10$	5.5715	2.9230	3.9278	0(99.9)	0(96.8)	98.6(0)
$T = 50$	5.3469	2.1801	3.9153	0(100)	0(100)	100(0)

Note: *h* denotes *h*-step-ahead forecasting horizons. MR means MEESE-ROGOFF test for panel regression. * (+) denotes the % of Model 1 (2)'s superiority to Model 2 (1) at the 5% level. RW means random walk with drift. Individual-Specific Effects (S_i) for CASE I and CASE 2 are drawn from $\chi^2_{(12)}$ and $\chi^2_{(24)}$, respectively. The results are based on 100 cross-section, 50 Out-of-Sample and 1,000 replications.

9. The results of the sign test and the sum-difference test are similar and available from the authors upon request.

squared forecast errors of the two panel regression models, when the panel data contains fixed or time-varying individual-specific effects. We compared these models with a random walk with drift as a benchmark model. The panel data with individual-specific fixed effects is discussed first.

The results may be summarized as follows: (1) Model 1 is the best model at the one-period-ahead forecast horizon, but it is the worst model for multi-step forecasts; (2) Model 2 is the best model except for the short forecast horizon in small sample. It is the worst model at $h = 1$ when the sample size is $T = 5$ or $T = 10$. Although model 2 suffers from small-sample bias, the model seems to be useful for multi-step forecasts.

In table 6, we consider time-varying effects with varying levels of δ and $\mathrm{Var}(\eta_{it})$; small variance and mild trend (Case 1: $\sigma_\eta^2 = 0.1$, $\delta = 0.05$); small variance and clear trend (Case 2: $\sigma_\eta^2 = 0.1$, $\delta = 0.5$); large variance and mild trend (Case 3: $\sigma_\eta^2 = 1.0$, $\delta = 0.05$). The results are as follows: (1) The DGP of case 2 (includes a large drift term) has a clear common trend component across individual and relatively small individual-specific time-varying size factor. Since the DGP of case 2 can be approximated well by model 1 in large sample, this model can perform the best multi-step forecast, regardless of the variance of the size-factor. However with the short forecasting horizon ($h = 1$), model 2 performs poorly. (2) When the time-varying size factor includes very distinct individual-specific time-varying components (case 3),

TABLE 6

Forecasting Comparison of the Panel Regressions of Panel Data with Time-Varying Individual-Specific Effects

h-step obs	Mean Squared Forecast Errors Model 1	Model 2	RW	Mean Squared Forecast Errors Model 1	Model 2	RW
$h = 1$		**CASE 1**			**CASE 3**	
$T = 5$	1.7634	73.2867	1.2783	2.2232	16.8001	2.1812
$T = 10$	1.6902	14.2860	1.2759	2.2025	6.7861	2.1772
$T = 50$	1.3227	1.3681	1.2767	2.1826	2.5342	2.1763
$T = 100$	1.2829	1.2875	1.2754	2.1767	2.2869	2.1739
$h = 5$						
$T = 5$	15.3291	282.6274	7.7758	9.5694	37.4978	8.2922
$T = 10$	13.9702	160.7196	7.7709	9.1813	33.0556	8.3085
$T = 50$	5.0826	6.2584	7.7587	8.6095	14.8645	8.3081
$T = 100$	3.9943	4.1128	7.7594	8.4392	11.0548	8.2953
$h = 20$						
$T = 5$	154.9523	591.6183	96.2264	37.5940	55.7706	24.8123
$T = 10$	148.2044	576.4762	96.2352	35.6242	56.1410	24.8811
$T = 50$	31.5963	51.2217	96.0909	30.4803	57.9604	24.8946
$T = 100$	10.7783	12.8425	96.1480	27.9416	50.3628	24.8237

Note: h denotes h-step-ahead forecasting horizons. RW means random walk with drift. Individual-Specific Effects (S_{it}) are generated with $\delta = 0.5$, $\sigma_\eta^2 = 0.1$ (Case 2), $\delta = 0.05$, $\sigma_\eta^2 = 1.0$ (Case 3). The results are based on 100 cross-section, 50 Out-of-Sample and 1,000 replications.

random walk with drift produce a better forecast spuriously than other models do. (3) For the short forecasting horizon ($h = 1$) model 1 might produce better forecast spuriously as expected from proposition 3. But it would not be true for the multi-step forecast or larger time-series panel data.

From the simulation study of forecasts, a regression model using fixed individual specific effects (*i.e.*, model 2) predicts very well for the long horizon if the size-factor is not time-varying. However, when the size factors vary, even very slowly, then this model forecasts poorly. Surprisingly, even though model 1 is misspecified, it can sometimes produce better forecasts. When the size factors have a clear drift term that is common for all individual series a pooling regression (*i.e.*, model 1) can capture this factor very well. If the size factors have a small trend component but still follow a random walk, then a random walk with drift can forecast better than model 1 or model 2.

6 Summary and Conclusions

This paper considers the effects of individual-specific factors in a dynamic panel regression model. Theoretical results show that size-factor, with a fat-tailed distribution or that has a time-varying property, may cause spurious stochastics in a regression. Moreover, the forecasts based on models which have omitted size-factors are affected seriously by the underlying property of size-factors. A pooling model appears to fit well in sample, but forecasts poorly out-of-sample if the individual-specific size-factor has a fat-tailed distribution. A panel model with individual-specific effect could be problematic if the panel series has a very short time-dimension. Since individual constant terms are estimated poorly, the forecasts based on them are poor. These problems may be more serious if the individual-specific factor is not constant but time-varying.

• References

DIEBOLD F.X., MARIANO R.S. (1995). – “Comparing Predictive Accuracy”, *Journal of Business & Economic Statistics*, 113, pp. 253-263.

ENTORF H. (1997). – “Random Walks with Drifts: Nonsense Regressions and Spurious Fixed-Effects Estimation”, *Journal of Econometrics,* 80, pp. 287-296.

GRANGER C.W.J., HUANG L.-L. (1997). – “Evaluation of Panel Data Models: Some Suggestions from Times Series”, *UCSD working paper*, pp. 97-10.

GRANGER C.W.J., NEWBOLD P. (1986). – *Forecasting Economic Time Series*, Academic Press, Orlando.

HYLLEBERG S., MIZON G.E. (1989). – “A Note on the Distribution of the Least Squares Estimator of a Random Walk with Drift”, *Economics Letters*, 29, pp. 225-230.

LEVIN A., LIN C.-F. (1992). – “Unit Root Test in Panel Data: Asymptotic and Finite-Sample Properties”, *UCSD working paper*, pp. 92-23.

MEESE R., ROGOFF K., (1988). – “Was it Real ? The Exchange Rate-Interest Differential Relation over the Modern Floating-Rate Period”, *Journal of Finance*, 43, pp. 933-948.

NICKELL S. (1981). – “Biases in Dynamic Models with Fixed Effects”, *Econometrica*, 49, pp. 1417-1426.

PRIESTLEY M.B. (1981). – *Spectral Analysis and Time Series*, New York: Academic Press.

APPENDIX A

Proof of Proposition 1:

With assumption of fixed individual-specific effects, the data generating process (1) can be written as, $X_{it} = S_i + \varepsilon_{it}^x$ and $Y_{it} = AS_i + \varepsilon_{it}^y$. Compute the asymptotic distribution by taking probability limits as $N \to \infty$ for given T. Since ε_{it}^x, ε_{it}^y are iid. for each t, when t is fixed we can replace *plim* as $N \to \infty$ by expectations across i, E_i, and thus obtain

$$\text{(A.1)} \qquad \underset{N\to\infty}{p\lim} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\overline{X})^2 = E_i\left(\frac{1}{T}\sum_{t=1}^{T}X_{it}^2\right) - \left(E_i\left(\frac{1}{T}\sum_{t=1}^{T}X_{it}\right)\right)^2 = \sigma_x^2 + \sigma_s^2$$

$$\text{(A.2)} \qquad \underset{N\to\infty}{p\lim} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\overline{X})(Y_{it}-\overline{Y}) = A\sigma_s^2,$$

where $$\overline{X} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}, \; \overline{Y} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Y_{it}$$

(a) For the given T and large N, the OLS estimator of $\widehat{\beta}$ in (2) can be written as,

$$\text{(A.3)} \qquad \underset{N\to\infty}{p\lim}\,\widehat{\beta} = \frac{\underset{N\to\infty}{p\lim} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\overline{X})(Y_{it}-\overline{Y})}{\underset{N\to\infty}{p\lim} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\overline{X})^2} = \frac{A\sigma_s^2}{\sigma_s^2+\sigma_x^2}$$

so that $\underset{N\to\infty}{p\lim}\,\widehat{\beta} \approx A$ when the variance of size (σ_s^2) is much larger than σ_x^2, $\sigma_s^2 \gg \sigma_x^2$.

(b) Now the t-statistic is

$$\text{(A.4)} \qquad \underset{N\to\infty}{p\lim} \frac{1}{\sqrt{N}}t_\beta = \underset{N\to\infty}{p\lim} \frac{\widehat{\beta}}{s\left[\sum_{i=1}^{N}\sum_{t=1}^{T}(X_{it}-\overline{X})^2\right]^{-1/2}} = \frac{A\sigma_s^2\sqrt{T}}{[A^2\sigma_s^2\sigma_x^2+\sigma_s^2\sigma_y^2+\sigma_x^2\sigma_y^2]^{1/2}}$$

where $\underset{N\to\infty}{p\lim}\, s^2 = \underset{N\to\infty}{p\lim} \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it}-\widehat{\alpha}-\widehat{\beta}X_{it})^2 = \frac{A\sigma_s^2\sigma_x^2}{\sigma_s^2+\sigma_x^2}+\sigma_y^2.$

Conventional t-statistics will diverge, *i.e.*, $t_\beta = O_p(N^{1/2})$ for given T.
(c) Next consider the coefficient of determination

(A.5)
$$\underset{N\to\infty}{p\lim}\, R^2 = \underset{N\to\infty}{p\lim} \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(\widehat{\alpha} + \widehat{\beta}X_{it} - \overline{Y})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(Y_{it} - \overline{Y})^2} = \frac{A^2\sigma_s^4}{(A^2\sigma_s^2 + \sigma_y^2)(\sigma_s^2 + \sigma_x^2)}$$

Then $\underset{N\to\infty}{p\lim}\, R^2 \approx 1$ as when the variance of size (σ_s^2) dominates σ_x^2 and σ_y^2. ■

Proof of Proposition 2:

(a) For given T, the OLS estimator of $\widehat{\beta}$ in (2) can be written as,

$$\underset{N\to\infty}{p\lim}\, \widehat{\beta} = A\frac{\Delta}{\Delta + \sigma_x^2 T^{-2}} \approx A$$

where $\Delta = \frac{1}{12}\delta^2 + \frac{1}{2}\sigma_\eta^2 T^{-1} + \left(-\frac{1}{12}\delta^2 + \frac{1}{2}\sigma_\eta^2\right) T^{-2}$.
(b) For given T, the OLS estimator of $\widehat{\beta}$ in (4) can be written as,

$$\underset{N\to\infty}{p\lim}\, \widehat{\beta} = A\frac{\Delta}{\Delta + \sigma_x^2 T^{-1}} \approx A$$

where $\Delta = \frac{1}{12}\delta^2 + \frac{1}{12}\sigma_\eta^2 T^{-1} + O(T^{-2})$ ■

Proof of Proposition 3:

From assumptions of Proposition 3, the DGP of Y_{it}, can be rewritten as, $Y_{it} = (1-\rho)S_i + \rho Y_{it-1} + \varepsilon_{it}$. Then (a), (b), and (d) follow immediately from Proposition 1. The proof of (c) is given below,
(c) $\widehat{m} = \overline{Y} - \widehat{\rho}\overline{Y}_{-1} = (1-\rho)\overline{S} + \rho\overline{Y}_{-1} + \overline{\varepsilon} - \widehat{\rho}\overline{Y}_{-1}$

where $\overline{Y}_{-1} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it-1}$

and $\overline{S} = \frac{1}{N}\sum_{i=1}^{N} S_i$, then $\underset{N\to\infty}{p\lim}\, \widehat{m} = \frac{(1-\rho)\mu_S}{1 + (1-\rho^2)\sigma_S^2/\sigma_\varepsilon^2}$

so $\widehat{m}$ approaches zero as the variance of size (σ_s^2) donimates. ■

APPENDIX B

From PRIESTLEY [1981], pp. 692-693) we obtain the distribution of the sample cross covariance function under the assumption of independence across individuals:

$$
\begin{aligned}
\mathrm{Var}(\widehat{\gamma}_{xz}) \\
&= E[\widehat{\gamma}_{xz}^2] - (E[\widehat{\gamma}_{xz}])^2 \\
&= \frac{1}{(NT)^2} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \sum_{i=1}^{N} E[x_{it} z_{it} x_{i\tau} z_{i\tau}] - \left(\frac{1}{NT} \sum_{t=1}^{T} \sum_{i=1}^{N} E[x_{it} z_{it}] \right)^2 \\
&= \frac{1}{(NT)^2} \sum_{t=1}^{T} \sum_{\tau=1}^{T} \sum_{i=1}^{N} \{E[x_{it} z_{it} x_{i\tau} z_{i\tau}] - E[x_{it} z_{it}] E[x_{i\tau} z_{i\tau}]\} \\
&= \frac{1}{(NT)^2} \sum_{i=1}^{N} \left[\sum_{t=1}^{T} \sum_{\tau=1}^{T} \{E[x_{it} z_{it} x_{i\tau} z_{i\tau}] - E[x_{it} z_{it}] E[x_{i\tau} z_{i\tau}]\} \right] \\
&= \frac{1}{(NT)^2} \sum_{i=1}^{N} \left[\sum_{t=1}^{T} \sum_{\tau=1}^{T} \{E[x_{it} x_{i\tau}] E[z_{it} z_{i\tau}] + E[x_{it} z_{i\tau}] E[z_{it} x_{i\tau}]\} \right]
\end{aligned}
$$

where we assume multivariate normality, so the fourth joint cumulant of the distribution of $[x_{it}, x_{i\tau}, z_{it}, z_{i\tau}]$ is zero. For MA$(h-1)$ processes, $\{x_{it}\}$ and $\{z_{it}\}$,

$$
\Sigma \equiv Var(\widehat{\gamma}_{xz}) = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{T} \sum_{\tau=-h+1}^{h-1} \{\gamma_{x_i x_i}(\tau) \gamma_{z_i z_i}(\tau) + \gamma_{x_i z_i}(\tau) \gamma_{z_i x_i}(\tau)\} \right]
$$

where

$$
\gamma_{x_i x_i}(\tau) = \mathrm{Cov}(x_{it}, x_{it+\tau}), \gamma_{x_i z_i}(\tau) = \mathrm{Cov}(x_{it}, z_{it+\tau}),\ \gamma_{z_i x_i}(\tau) = \mathrm{Cov}(z_{it}, x_{it+\tau}).
$$

A consistent estimator of Σ is

$$
\widehat{\Sigma} = \frac{1}{N} \sum_{i=1}^{N} \left[\sum_{\tau=-h+1}^{h-1} \left(1 - \frac{|\tau|}{T} \right) \{\widehat{\gamma}_{S_i S_i}(\tau) \widehat{\gamma}_{D_i D_i}(\tau) + \widehat{\gamma}_{S_i D_i}(\tau) \widehat{\gamma}_{D_i S_i}(\tau)\} \right]
$$

as showed by MEESE and ROGOFF [1988]. ■