E C O N O M I C S   B U L L E T I N

Efficiency measurement with nonstationary variables: an application of panel cointegration techniques

Efthymios Tsionas
Athens University of Economics and Business

Dimitris Christopoulos
Panteion University

Abstract

In this paper, we apply panel cointegration tests and estimation techniques to obtain efficiency measures when it is uncertain whether the underlying technological relationship is structural or spurious due to possible non−stationarity of the data. We illustrate the dangers of efficiency measurement with panel data when integration and cointegration are not taken into account. We apply these techniques to efficiency measurement in U.S. airlines and find striking differences compared to results obtained with the traditional approach.

Acknowledgment: We would like to thank the Associate Editor, Chung−Ming Kuan, and Clive Richardson for many useful comments.
Citation: Tsionas, Efthymios and Dimitris Christopoulos, (2001) "Efficiency measurement with nonstationary variables: an application of panel cointegration techniques." *Economics Bulletin,* Vol. 3, No. 14 pp. 1−7
Submitted: October 10, 2001. **Accepted:** October 11, 2001.
URL: http://www.economicsbulletin.com/2001/volume3/EB−01C20005A.pdf

Efficiency measurement with non-stationary variables: An application of panel cointegration techniques

Efthymios G. Tsionas
Department of Economics, Athens University of Economics and Business,
Athens, Greece

Dimitris K. Christopoulos
Department of Economic and Regional Development, Panteion University,
Athens, Greece

Abstract

In this paper, we apply panel cointegration tests and estimation techniques to obtain efficiency measures when it is uncertain whether the underlying technological relationship is structural or spurious due to possible non-stationarity of the data. We illustrate the dangers of efficiency measurement with panel data when integration and cointegration are not taken into account. We apply these techniques to efficiency measurement in U.S. airlines and find striking differences compared to results obtained with the traditional approach.

Key words: Cointegration; panel data; efficiency measurement; fully modified estimation.

JEL codes: C23, D24.

Acknowledgment: We would like to thank the Associate Editor, Chung-Ming Kuan, and Clive Richardson for many useful comments.

Corresponding author: E. G. Tsionas, Department of Economics, Athens University of Economics and Business, 76 Patission Street, 104 34 Athens, Greece.
Tel.: (++301) 8203 392, e-mail: tsionas@aueb.gr

1. Introduction

The measurement of efficiency is important in nearly every field of applied economics. In the long run, only efficient firms can survive competition, and a reasonable business strategy to achieve success is to imitate existing best practices in the industry as far as possible. The typical efficiency study starts from a production or cost function and uses estimated residuals or estimated fixed effects to produce efficiency measures. However, recent advances in time series analysis have brought to attention an important problem. If, for each unit, the time series are integrated, then the issue of false correlation arises. If the underlying variables are not cointegrated (in other words, if the estimated production or cost function is spurious) efficiency measurement is meaningless.

The present paper applies panel cointegration techniques to ensure that the estimated technological relationship is structural as opposed to spurious, and uses fully modified OLS to obtain parameter estimates and efficiency measures. We compare our results with those obtained by measuring efficiency along traditional lines, and find quantitatively important differences. In addition, we conduct a small Monte Carlo experiment to show that independent random walks, when combined in a panel data context, will produce apparently reasonable efficiency estimates, whereas in fact such measures do not even exist because the underlying technology is spurious.

The paper is organized as follows. The model is presented in Section 2. The small experiment is presented in Section 3. In Section 4, we apply the techniques to estimate the efficiency of U.S. airlines. A summary and concluding remarks are contained in the final section.

2. The model

Consider the model

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + v_{it} - u_{it}, \quad i = 1,..,N, \; t = 1,..,T \tag{1}$$

where y_{it} is the observation in year t for the i*th* dependent variable (typically the log of production level), $\mathbf{x}_{it}$ is a $K \times 1$ vector of observations of the explanatory variables (typically logs of inputs) in year t and $\boldsymbol{\beta}$ is a $K \times 1$ vector of coefficients. Measurement errors v_{it} are assumed to be *i.i.d.* $\mathrm{N}(0,\sigma^2)$. The non-negative disturbance u_{it} reflects an inefficiency component that forces production to be below the frontier. If $u_{it} = 0$, then the firm is fully efficient. When all variables are integrated of order one, i.e. $I(1)$, if (1) represents a cointegrating relationship, then least squares provides super-consistent estimates. Otherwise, (1) is a spurious relationship, and efficiency measurement is seriously misguided. Additional complications arise if (1) is not balanced, that is, when different variables have different orders of integration; for example, some are stationary but others contain unit roots.

3. An experiment with artificial data

We have conducted the following experiment. We set $K = 2$ so that we have a linear model with an intercept and a single explanatory variable x_{it}, with $N = 25$ firms and $T = 10$ years. We assume that y_{it} and x_{it} are independent random walks with drift:

$$y_{it} = c + y_{it} + \xi_{it} \tag{3a}$$

$$x_{it} = d + x_{it-1} + \zeta_{it} \tag{3b}$$

$$i = 1,.., N\ ,\ t = 1,.., T$$

with $c = d = 0.1$, where ξ_{it}, ζ_{it} are mutually independent $N(0,0.1)$ errors. 5,000 different data sets were generated and each time the model (1) was estimated using OLS with fixed effects. For each data set, efficiency was estimated using the usual approach. The density of average efficiency estimates is presented in Figure 1. It is clearly skewed to the left, and may produce a false sense of security that these data are compatible with inefficiency levels near 14%. These results look quite plausible despite the fact that there is no underlying production function. Therefore, one must be extremely careful in interpreting efficiency from time series cross-section models.

4. Empirical application

In this section we consider nine (N=9) US airlines over the period 1971-1985 (T=15). In this application, total cost (C) is a function of output (Y) and the prices of capital (P_K), labour (P_L), materials (P_M) and fuel (P_F). For the sources and the structure of the variables, see Appendix A of Baltagi, Griffin and Vadali (1998)[1]. We assume that the parametric form of our cost function is Cobb-Douglas and can be written as follows

$$\ln C_{it} = \alpha_i + \sum_{k=1}^{m} \beta_k \ln P_{itk} + Y_{it} + e_{it} \tag{4}$$

$$e_{it} = v_{it} + u_{it}$$

P_{itk} is the k^th^ input price (k=K,L,M,F) of the i^th^ firm (i=1,2,..9) in time period t (t=1971,1972,…1985), Y_{it} is the output of the i^th^ firm in time period t, u_{it} is a non-negative error representing technical inefficiency, and v_{it} is the usual statistical noise.

[1] The whole data set consists of annual observations on the domestic operations of twenty-four airline firms over the period 1971-1986 (panel data). A balanced panel was constructed by including the nine firms that operated throughout the period 1971-1985.

Following Jondrow *et al*. (1982), the conditional distribution of u_{it} given e_{it} is

$$E(u_{it} \mid e_{it}) = s_{it}[e_{it}\lambda_{it} / \sigma_{it} + \phi(e_{it}\lambda_{it} / \sigma_{it})\Phi(e_{it}\lambda_{it} / \sigma_{it})^{-1}] \tag{5}$$

where $s_{it} = \sigma_v \sigma_{u,it} / \sigma_{it}$, $\lambda = \sigma_u / \sigma_v$, $\sigma^2 = \sigma_u^2 + \sigma_v^2$, ϕ and Φ denote the standard normal density function and the standard normal distribution function respectively evaluated at $(e_{it}\lambda)/\sigma$.

Next, we use panel unit root tests to see if our series are $I(1)$. To test for the existence of a unit root in a panel data setting, we have used tests due to Harris and Tzavalis (HT) (1999), Im, Pesaran and Shin (IPS) (1997) and Maddala and Wu (MW) (1999). In each test, the null hypothesis is that of a unit root. The results are reported in Table 1. All series contain a unit root with the exception of prices of materials ($\ln P_M$) where the IPS statistic indicates that this series is stationary, and the cost ($\ln C$) and output ($\ln Y$) series where the HT test rejects a unit root. However, according to IPS and MW tests this is not the case. Therefore, it seems reasonable to proceed on the working hypothesis that all series are $I(1)$.

The results of the MW and HT panel cointegration tests are reported in Table 2 and can be used to test for the existence of a cointegrating cost function. Panel cointegration tests are used in order to draw sharper inferences since the time spans of our economic time series are very short. Since the null hypothesis of no cointegration is rejected at the 10% level of significance, these results suggest that there is a cointegrating cost function. Having established that the dependent variable ($\ln C$) is structurally related to the explanatory variables, p_{it} (i=K,L,M,F) and y_{it}, we proceed to estimate the cost function using fully modified OLS (FMOLS) for heterogeneous cointegrated panels (Pedroni, 2000). This allows consistent and efficient estimation of cointegrating vectors. It is well known that OLS estimation is biased due to the endogeneity of the $I(1)$ regressors. More specifically, we consider the following cointegrated system for panel data

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it} \tag{6}$$

$$x_{it} = x_{i,t-1} + e_{it} \tag{7}$$

where $\xi_{it} = [u_{it}, e'_{it}]$ is stationary with covariance matrix Ω_i. Following Phillips and Hansen (1990), a semi-parametric correction can be made to the OLS estimator in order to eliminate the second order bias caused by the fact that the regressors are endogenous. Pedroni (2000) follows the same principle in the panel data context, and allows for heterogeneity in the short run dynamics and the fixed effects. Pedroni's estimator is

$$\hat{\beta}_{FM} - \beta = \left(\sum_{i=1}^{N} \hat{\Omega}_{22i}^{-2} \sum_{t=1}^{T} (x_{it} - \bar{x}_t)^2 \right)^{-1} \cdot \sum_{i=1}^{N} \hat{\Omega}_{11i}^{-1} \hat{\Omega}_{22i}^{-1} \left(\sum_{t=1}^{T} (x_{it} - \bar{x}_t) u_{it}^{*} - T\hat{\gamma}_i \right) \tag{8}$$

$$\hat{u}_{it}^{*} = u_{it} - \hat{\Omega}_{22i}^{-1}\hat{\Omega}_{21i}\,,\ \hat{\gamma}_i = \hat{\Gamma}_{21i} + \hat{\Omega}_{21i}^{0} - \hat{\Omega}_{22i}^{-1}\hat{\Omega}_{21i}(\hat{\Gamma}_{22i} + \hat{\Omega}_{22i}^{0}) \qquad (9)$$

where the covariance matrix can be decomposed as $\Omega_i = \Omega_i^0 + \Gamma_i + \Gamma_i$ where Ω_i^0 is the contemporaneous covariance matrix, and Γ_i is a weighted sum of autocovariances. Also, $\hat{\Omega}_i^0$ denotes an appropriate estimator of Ω_i^0.

FMOLS parameter estimates of the cost equation together with traditional stochastic frontier ML estimates are reported in Table 3. It must be noted that the estimated cost function includes fixed effects when estimated by either ML or FMOLS. The results in Table 3 indicate that the majority of coefficients are statistically different from zero at conventional levels of statistical significance, as expected. The value of λ is 4.27 and 1.018 for the ML and FMOLS estimation techniques, respectively. These results imply that the one-sided error term u dominates the symmetric error v in ML, whereas they have about the same magnitude in FMOLS. In other words, breaking up the residuals into noise and inefficiency is very sensitive to the method of estimation. Also, the inefficiency effects based on ML estimates are greater compared to those derived from the FMOLS results.

Efficiency for U.S. airlines and associated efficiency rankings for 1985 are presented in Table 4. Efficiency is 93.9% according to ML and 89.7% according to FMOLS. The standard deviation is close to 0.04, and the extreme values seem to be about the same. However, important differences arise when efficiency rankings are considered. According to FMOLS, American was most efficient in 1985 whereas it is fourth according to ML. Braniff is the second most efficient firm according to ML but it appears less efficient based on FMOLS. We can examine the relationship between efficiency measures further for each airline, and each year as well. The correlation coefficient between ML and FMOLS efficiency measures is only 0.122 (0.113 in logs) and, from Figure 2 (where efficiency estimates are considered together along with a non-parametric regression fit), it is apparent that the two sets of estimates bear little resemblance to each other. In other words, although first and second moments of efficiency estimates match well, the distributions are not the same. Therefore, the choice between using ML or FMOLS is important when variables are non-stationary, since one may end up with very different efficiency estimates.

Conclusions

We have argued in this paper that the estimation of efficiency from panel data has to face the spurious regression problem raised in the recent time series analysis literature. In a Monte Carlo experiment, we have shown that the effects of integrated but not cointegrated time series can lead to misleading efficiency measures. We have applied panel cointegration techniques to efficiency measurement in U.S. airlines, estimating the model using fully modified OLS. This gives efficiency measures that are radically different in terms of firm-specific efficiency measures from those obtained by the usual estimation procedure.

References

Baltagi, H.B., Griffin, M.J. and R. Vadali, 1998, Excess Capacity: A Permanent Characteristic of U.S. Airlines, *Journal of Applied Econometrics* , 13, 645-657.

Harris, R.D.F. and Tzavalis, E., 1999, Inference for Unit roots in Dynamic Panels where the Time Dimension is Fixed, *Journal of Econometrics*, 91, 201-226.

Im, S.K., Pesaran, H. M. and Shin, Y., 1997, Testing for Unit Roots in Heterogeneous Panel, Department of Applied Econometrics, University of Cambridge.

Jondrow, J., Lovell, C.A.K., Materov, I., and Schmidt, P., (1982). On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Model, *Journal of Econometrics*, 19, 233-238.

Maddala, G. S. and Wu, S., 1999, A Comparative Study of Unit Root tests with Panel Data and a New Simple Test, *Oxford Bulletin of Economics and Statistics*, 61, 631-652.

Pedroni, P., 2000, Fully Modified OLS for Heterogeneous Cointegrated Panels, *in Non-Stationary Panels, Panel Cointegration and Dynamic Panels*, Vol. 15, pp.93-130, Elsevier Science Inc.

Phillips, P.C.B. and Hansen, B.E., 1990, Statistical Inference in Individual Variables Regression with I(1) Process, *Review of Economic Studies*, 57, 99-125.

Table 1. Panel Unit Root Tests

	Levels			First differences		
Variables	IPS	HT	MW	IPS	HT	MW
$\ln C$	**1.44**	2.15	**15.2**	-2.96	-7.38	90.68
$\ln P_K$	**-0.85**	**1.21**	**12.8**	-2.20	-9.25	39.57
$\ln P_L$	**-0.23**	**1.20**	**18.6**	-4.59	-12.07	78.22
$\ln P_M$	-3.01	**1.52**	**16.9**	-5.44	-8.00	69.25
$\ln P_F$	**-1.46**	**1.36**	**18.2**	-3.50	-6.28	43.93
$\ln Y$	**3.23**	2.77	**17.1**	-6.21	-14.24	50.15

Notes: IPS, HT and MW are respectively the Im, Pesaran and Shin, Harris and Tzavalis, and Maddala and Wu tests for a unit root in the model. Bold face values denote sampling evidence in favour of unit roots. The critical value for the MW test is 25.99 at the 10% statistical level. All tests agree that first differences are stationary for all series.

Table 2. Panel Cointegration Tests

MW	HT
34.07*	-6.69*

Notes: A * signifies rejection of the null hypothesis of no-cointegration at the 10% significance level. The critical value for the MW test is 25.99 at the 10% level of significance.

Table 3. Maximum likelihood and fully modified OLS estimates

	Maximum Likelihood		Fully modified OLS	
Variables	*Coefficients*	*t-statistics*	*Coefficients*	*t-statistics*
$\ln P_K$	0.04	1.39	-0.04	0.04
$\ln P_L$	0.15*	3.89	0.36*	6.65
$\ln P_F$	0.31*	8.50	0.35*	12.62
$\ln P_M$	0.11	1.13	0.01	0.60
$\ln Y$	0.81*	23.60	0.60*	14.76
$\lambda = \sigma_u / \sigma_v$	4.27*	2.82	1.018	---

Notes*:* Firm-specific dummies are not reported to save space but are available from the authors upon request. A * indicates statistical significance at the 10% level of significance.

Table 4. Comparison of efficiency measures and rankings for 1985

Firms	ML		FMOLS
1. Midway	0.96	1. American*	0.93
2. Braniff	0.95	2. North Central/ Republic	0.93
3. Air Cal	0.95	3. Midway	0.93
4. American*	0.94	4. Delta	0.92
5. National/PanAm*	0.94	5. National/PanAm*	0.89
6. North Central/ Republic	0.94	6. Braniff	0.88
7. Southwest	0.94	7. Eastern*	0.87
8. Eastern*	0.93	8. Air Cal	0.87
9. Delta	0.91	9. Southwest	0.82
Mean efficiency	0.939		0.897
Stand. Dev. of efficiency	0.044		0.043
Minimum efficiency	0.759		0.738
Maximum efficiency	0.993		0.951

* Trunk airlines.

Figure 1. Density of average efficiency (N=25, T=10)

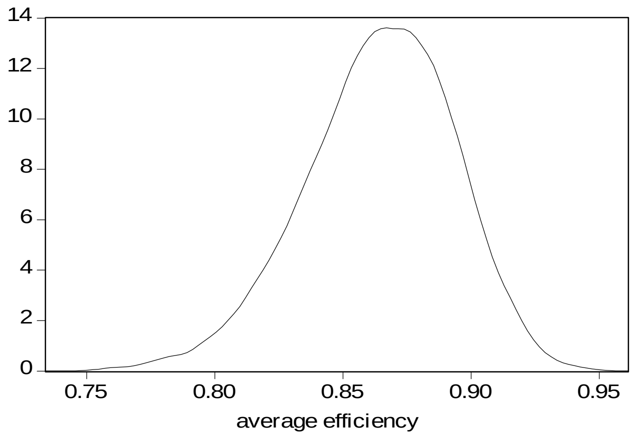

Figure 2. Comparison of efficiency estimates from OLS and FMOLS (nonparametric regression estimate is also provided)