–学位論文最終要旨–
SOME DEVELOPMENTS IN DYNAMIC PANEL DATA ANALYSIS

Hosung Jung,
Graduate School of Economics
Hitotsubashi University

1 ABSTRACT

As many economic relationships are dynamic in nature, various econometric procedures for analyzing these dynamics have been studied. Among these procedures, this dissertation is particularly concerned with (i) specification tests in dynamic panel models (ii) asymptotic results in cointegrating regression and unit root test in nonstationary panel models. The purposes of this dissertation, therefore, are to propose autocorrelation and cross sectional correlation tests in dynamic panel models and to develop a unit root test that accounts for cross-sectional dependency in nonstationary panel models.. Hence, focus is on the impacts of serial correlations in either time or cross-section dimensions of panels and on the impacts of small or fixed time series dimensions. Additionally, some problems on the estimation of both stationary and nonstationary panel data models are considered.

Chapters 2–4 deal with generalized method of moment (GMM) estimations in dynamic panel models. Chapter 2 finds an alternative system GMM estimation by changing the weight matrix that accounts for the variance ratio between individual effects and idiosyncratic disturbances. Since an initial optimal weight matrix is not known, we suggest the use of *a suboptimal weight matrix*, which reduces the finite sample bias while increasing the efficiency in the system estimation. We also investigate the potential efficiency gain based on the Kantorobich Inequality (KI) and the small sample properties of the suboptimal system estimator. Chapter 3 suggests an autocorrelation test in dynamic panel models. It is well known that the standard Arellano and Bond (1991) GMM estimator loses its consistency when the errors are serially correlated. They suggests several specification tests to detect the existence of

serial correlation, such as the m_2 and Sargan tests. We, therefore, compare our test with the m_2 and Sargan tests where the error terms follow either AR(1) or MA(1) process. In Chapter 4, we shift our attention to correlation in the cross-section dimensions. Using a time-varying, individual-specific effects model, suggested by Holtz-Eakin (1988), we suggest Hausman and Sargan tests for sectional correlation and compare their small-sample properties with existing tests: the LM (Breusch and Pagan, 1980), CD (Pesaran, 2004) and Sargan-difference tests (Sarafidis, Yamagata and Robertson, 2006).

Chapters 5 and 6 develop the analysis of nonstationary panel data models. Chapter 5 studies the limit theory in a nonstationary panel while the time series dimension, T, is assumed to be fixed. The asymptotic distributions of the Least Square Dummy Variables (LSDV) and the Fully Modified (FM) OLS estimators are derived in both spurious and cointegrated regression. This chapter also demonstrates how the order of a limit sequence affects its asymptotic results. To this end, we compare the limiting behavior of these estimators with that of the sequential limit designed by Phillips and Moon (1999).

In Chapter 6, we develop a panel unit root test that accounts for cross-sectional correlation. The conventional unit root and cointegration tests assume cross-sectional independence in nonstationary panel data models. However, this rather convenient assumption is rarely satisfied in practice, especially in a macro-economic context. To remedy this, we develop a modified DF type unit root that uses the Levin and Lin (1992) test followed by the quasi-difference defactoring procedure.

2 Dynamic Panel Data Models: Chapters 2–4

Since the pioneering work of Balestra and Nerlove (1966), the analysis of dynamic panel data models has been a major topic of econometrics. These dynamic relationships are characterized by the inclusion of a lagged dependent variable among the regressors; for example,

$$y_{it} = \delta y_{i,t-1} + x_{it}'\beta + u_{it} \qquad i = 1, \ldots, N;\ t = 1, \ldots, T. \tag{1}$$

We may assume that u_{it} follow a simple one-way error component model,

$$u_{it} = \mu_i + v_{it}, \tag{2}$$

where $\mu_i \sim iid(0, \sigma_\mu^2)$ and $v_{it} \sim iid(0, \sigma_v^2)$ are independent of each other.[1] Alternatively, we may want to model a more complicated relationship, such as serial correlation in either the time series dimension or a cross-section dimension. For

[1] A two-way error component model, $u_{it} = \mu_i + f_t + v_{it}$, can also be assumed, where μ_i, f_t and v_{it} are independent of each other.

example, as in Holtz-Eakin (1988), we assume that u_{it} has time-varying individual specific effects,

$$u_{it} = \phi_i f_t + v_{it}, \tag{3}$$

so that there exist sectional correlations: $E(u_{it}u_{jt}) \neq 0$ for $i \neq j$.[2]

In this dissertation, we first consider the dynamic panel data regression described in (1) with a simple one-way error component model, specifically the efficient system GMM estimation, the autocorrelation test and the Sargan test for sectional correlation with a time-varying individual-effect model. In the context of the dynamic panel—i.e., a fixed effects model with a one-way error component structure—we find that the LSDV estimator can be severely downward biased when the time dimension is short, regardless of the cross-sectional size of the panel. On the other hand, Anderson and Hsiao (1981, 1982) propose two IV estimator (AH-IV) that make use of the dependent variable lagged two periods, $y_{i,t-2}$ or its first differencing, $\Delta y_{i,t-2} = y_{i,t-2} - y_{i,t-3}$ as instruments for their models in first differences. These instruments are not correlated with Δu_{it}, as long as u_{it} themselves are not serially correlated. Although the AH-IV estimators are consistent only for $N \to \infty$, they are inefficient because they do not exploit all valid instruments and do not take into account the MA(1) structure of the disturbance term. To remedy this, many GMM-based estimations have been suggested.

Chapter 2 examines the efficiency gain from a suboptimal weight matrix in the system GMM estimation. It is generally known that using many instruments can improve the efficiency of various IV and the GMM estimators (Arellano and Bover, 1995; Blundell and Bond, 1998). The system GMM estimator in dynamic panel data models, which combines two moment conditions (i.e., for the differenced equation and for the model in levels), is, therefore, more efficient than the standard GMM estimator. It is also well understood that an asymptotically efficient estimator can be obtained through the two-step procedure in standard GMM estimation. In that case, the two-step system GMM estimator must be the most preferred one in dynamic panel model. However, as in standard GMM estimation, the estimated standard error of the two-step system estimator can have severe downward bias for moderate sample sizes, N (Windmeijer, 2004). In practice, therefore, we often rely on inference based on the less-efficient one-step estimator, which is much more reliable than the two-step estimator.

The main object is, then, how we choose the weight matrix in the first step of the estimation, especially in small samples in time series dimension. Relating this, a previous study (Windmeijer, 1998) shows that the efficiency of the system GMM estimator is greatly affected by the choice of the weight matrix.

[2] We can also set ϕ_i as factor loadings to be estimated while f_t is unobserved common factor that induce sectional correlations. See Pesaran (2004).

We therefore suggest a suboptimal weight matrix that accounts for the variance ratio (the ratio of the variance of individual effects to the variance of the idiosyncratic error terms, say, $\rho = \frac{\sigma_\mu^2}{\sigma_v^2}$) instead of using the conventional weight matrix, i,e., the identity matrix, corresponds to the level estimation in the system GMM estimation. As this suboptimal weight matrix requires a consistent estimate of the variance ratio, the suboptimal system GMM estimator is categorized as a two-step system GMM estimator. However, as it does not directly use residuals from the first-step estimation, the downward-bias problem of the estimated standard error is expected to be negligible.

Chapters 3 and 4 study two independent but closely related specification tests in dynamic panel data models: an autocorrelation test for the time series dimension and a serial correlation test for the cross-section dimension. Arellano and Bond (1991) propose a standard GMM estimator that uses all the available lags at each period as instruments for the first-difference equations to enhance efficiency. In addition, they provide some specification tests (e.g., m_2 and Sargan tests) for the hypothesis that there is no second-order serial correlation for the disturbances of the first-differenced equation. These tests are fundamental because the consistency of the GMM estimators relies upon there being no correlation between the first-differenced and two lagged variables.

However, there are some shortcomings in that these tests are limited to uncorrelated disturbances under the null but only moving-average (MA) errors under the alternative hypothesis. Now, let us suppose that the disturbances have autocorrelation (say, AR(1)) structures. As in the case of MA(1) disturbances, the usual approach of using lagged values of the dependent variables as instruments in the differenced equations is no longer valid. Furthermore, a GMM estimator that uses lags as instruments under the assumption of white noise errors is inconsistent. This implies that the m_2 and Sargan tests—by using the inconsistently estimated residuals from the first-differenced GMM estimation, which also uses invalid instruments—may not be able to distinguish between AR and MA disturbances. The main purpose of these chapters is, therefore, to propose a test of serial correlation based on the consistently estimated residuals under either the null or the alternative hypothesis, and to compare its performances with the m_2 and Sargan tests.

A serial correlation in the time series dimension is not the standard GMM estimator's only source of the inconsistency. The impact of cross-sectional correlation—which may arise due to the presence of common shocks and unobserved components that become part of the error term, spatial dependence, or idiosyncratic pair-wise dependence in the disturbances with no particular pattern or common component—is comparatively more severe in a dynamic panel estimator. (Robertson and Symons, 2002; Pesaran, 2004; Anselin, 2001; etc.). If there is sufficient cross-sectional dependence in the data, and this is ignored in the estimation, the pooled least squares estimator provides little gain over the single OLS equation. Phillips and Sul (2003) show that the efficiency gains by pooling a population of cross-sections may largely diminish if the cross-sectional dependence is ignored.

In Chapter 4, we suggest that if the disturbances are serially uncorrelated in time, the conventional Sargan test can be used to detect the sectional correlation. This implicitly shows that the Sargan test for misspecification is very unstable if sectional correlation exists.

3 Nonstationary Panel Data Models: Chapters 5– 6

From the usual asymptotics of micro panels with large N (e.g., number of individuals) and small T (time series dimension), the focus of panel data econometrics has shifted towards studying the asymptotics of macro panels with large N and large T due to the growing use of cross-country data over time to study Purchasing Power Parity (PPP), international R&D spillover and growth convergence. These panels have different characteristics and implications for theoretical and empirical analysis from the large N abd small T panels that have been the traditional object of study in panel data analysis. For example, in order to properly analyze large N and large T panel data, the concept of multi-index asymptotics must be adopted.

There are two important streams in the literature on the use of large N and large T nonstationary panels that are closely related the dissertation: (i) estimation and inferences in cointegrating models and (ii) unit root and cointegration tests. Let us consider the following panel regression

$$y_{it} = x_{it}'\beta + z_{it}'\gamma + u_{it}, \tag{4}$$

where z_{it} are the deterministic components (i.e., zero, one, and the fixed effects, μ_i) or fixed effects as well as a time trend, t, and u_{it} are the stationary disturbance terms. We assume that x_{it} are a vector of integrated processes of order one for all i:

$$x_{it} = x_{i,t-1} + \epsilon_{it}, \tag{5}$$

where ϵ_{it} are i.i.d. disturbances. On the estimation and inference in the above panel cointegration models, Kao and Chiang (2000), Phillips and Moon (1999), and Pedroni (2000) have shown that the asymptotic properties of the LSE's of the regression coefficient and the associated test statistics are different from those of the time-series cointegration regression models. In particular, Kao and Chiang (2000) study the limiting distributions for the fully modified (FM) and dynamic ordinary least square (DOLS) estimators in a cointegrated regression and showed they are asymptotically normal. The above estimators can be seen as generalizations of Phillips and Hansen (1990) and Saikkonen (1991), respectively. Phillips and Moon (1999) and Pedroni (2000) also present similar results for these estimators while the latter considers the FM estimators in heterogeneous panel data models.

In the spirit of Kao and Chiang (2000), Chapter 5 derives the asymptotic distributions of the LSDV and FM-OLS estimators in both spurious and cointegrated panels. The asymptotic distributions of the LSDV and FM-OLS estimators are investigated based only on large N asymptotics; the time series dimension, T, is assumed to be fixed. Much past panel data research has focused on estimating effects from nonstationary panels with a large time series, T, and large individuals N, though the practical sample size in the time dimension is moderate. This chapter introduces a regression limit theory for nonstationary panel data with large numbers of cross-sectional observations but moderate time series observations. For such cases a new nonstationary panel limit theory which allows for large N and fixed T may be useful. We will, therefore, introduce a fixed-T limit theory[3] in nonstationary panel models with homoscedastic disturbance terms in both spurious and cointegrated cases.

The last chapter studies panel unit root tests that account for cross section dependence with a one-factor residual model. Consider the conventional model with no factor structure,

$$y_{it} = \rho_i y_{i,t-1} + z'_{it}\gamma + u_{it}, \tag{6}$$

where u_{it} is a stationary process.

Earlier works in the context of panel unit root tests include Levin and Lin (1992, 1993); Maddala and Wu (1999); Choi (2001a); Im, Pesaran and Shin (1997); and Harris and Tzavalis (1999), to mention a few. In their seminal work, Levin and Lin (1992, 1993) develop the asymptotic properties of unit root tests on panel data as both N and T grow arbitrarily large. They also show how an augmented DF test statistic for each individual time series can be used to construct a test for pooled LSE's regressions. The so-called IPS test allows for a heterogeneity of regression coefficients and offers an alternative testing procedure based on averaging individual unit root test statistics. Harris and Tzavalis (1999) derive unit root tests for pooled LSE regression when the time dimension of the panel T is fixed and show that Levin and Lin's (1992) tests can be substantially undersize when T is small.

Although there is much research not named here, almost all of the aforementioned research has in common in that they assume cross-sectional independence; that is, $E(u_{it}u_{jt}) = 0$. The transition from theory to a testable form inevitably involves the use of so-called *simplifying assumptions*. Cross-section independence in panel unit root tests has been this kind of assumption that is rarely satisfied in practice. In

[3] Harris and Tzavalis (1999), though it studies panel unit root tests, analytically derives the limiting properties of LSE's when T is fixed.

cross-country analysis there might be unobserved common influences on all members. For example, the EU countries are coordinated through their many common economic policies. Due to the strong links across markets and the use of a numeraire country in defining real exchange rates, as argued by O'Connell (1998), real exchange rates between countries should have high cross-correlation in both the short and long run. Furthermore, there is little evidence of cross-sectional independence within the context of micro economy such as regional productivity.

However, attention has been drawn recently to the assumption of cross section independence on which the asymptotic results of the conventional unit root test procedure rely (O'Connell, 1998; Maddala and Wu, 1999). Using Monte Carlo simulations, O'Connell (1998) shows that the Levin, Lin and Chu (2002) test suffers from sectional correlation in terms of increased size and reduced power. Maddala and Wu (1999) also report similar results from bootstrap experiments. This deficiency is easily explained by the cross-sectional correlations: The variance of the numerator of, say, the DF-t test increases with N. In that case, the Central Limit Theorem, which requires a finite variance cannot be applied. Hence, no convergence result can be stated for this general form of cross section dependence as N grows to infinity.

To remedy this, unit root tests that account for cross-sectional correlation have been proposed recently. Among them are attempts to model sectional correlation using common-factors representations of the data:

$$u_{it} = \lambda_i f_t + v_{it}, \tag{7}$$

where λ_i is a nonstochastic factor loading and f_t is an unobserved common factor independent of the idiosyncratic disturbances, v_{it}. Research into this criteria include Pesaran (2004), Moon and Perron (2004), Bai and Ng (2004), and Phillips and Sul (2003), to name a few.

The main purpose of Chapter 6 is to propose new panel unit root test (Modified Dickey and Fuller (MDF) tests) for one-factor residual models and to compare their small-sample performance with that of Pesaran's (2004) and Moon and Perron's (2004) tests. These two tests are sufficiently close to our tests in model specifications—while, at the same time, the defactoring procedures differ in important ways—to make it interesting to compare their properties. While Pesaran (2004) and Moon and Perron (2004) use cross-sectional augmentation and principal-component analysis to remove the sectional dependence, respectively, the MDF test directly removes the factor from the variables through the Quasi-difference procedure. As this defactoring procedure does not affect the limiting behavior of the original variables, the standard DF tests are applied to the defactored series. The asymptotic results reveal that the tests have limiting normal distributions under the null of a unit root. Monte Carlo studies also demonstrate that the tests have reasonable small-sample performance in terms of size and power.