Recent Development in Panel Data Econometrics

by

Yoshitsugu Kitazawa

Faculty of Economics, Kyushu Sangyo University

E-mail: kitazawa@ip.kyusan-u.ac.jp

1. Introduction

The development of panel data econometrics in recent years has lead to the expansion of the range of economic and financial models where the panel data model is applicable. The development has been mainly achieved in the panel data model dealing with the case of a large number of individuals and a small number of time-series observations. In addition, the development has predominantly grown out of introducing the GMM (Generalized Method of Moments) proposed by Hansen (1982). The GMM, including the IV (Instrumental Variable) method, allows the implementation of consistent estimations for the panel data model based on conditional expectations and the dynamic panel data model, which we cannot estimate consistently in the framework of traditional estimation techniques such as the OLS (Ordinary Least Squares) and the LSDV (Least Squares Dummy Variable) estimators. Furthermore, it is recognized that the GMM is applicable to accomplish consistent estimations for the count panel data model and the structure of the variance in the panel data model. Until now, a great deal of empirical studies in economics and finance was implemented in parallel with the development of the estimation techniques.

Explanations on the estimation methods and the empirical studies are conducted in section 2 of this paper. Section 3 concludes this survey paper.

2. Review on the Panel Data Econometrics

In this section, we epitomize the history on the recent development of the panel data econometrics coping with the panel data from a large number of individuals and a small number of time-series observations.

Through the thesis, indices for the individual and the time-series observation are i and t, respectively. Ranges of i and t are respectively $i = 1,\ldots,N$ and $t = 1,\ldots,T$ unless the ranges are specified otherwise. As a presumption, we suppose that variables on individual i are independent of variables on individual j, where $i \neq j$. The assumption in this thesis is that $N \to \infty$ and T is fixed. Accordingly, the asymptotic on estimators relies on $N \to \infty$.

Firstly, the static panel data model is introduced as a traditional panel data model and then two traditional estimation methods for the model are explained. In addition, a comparatively new method in econometrics is explained. Secondly, the dynamic panel data model is introduced and then a series of recent developments for estimating the model is outlined in detail. Thirdly, one of the panel data models with multiplicative individual effects and an estimation method for the model are illustrated.

Static Panel Data Model

The traditional model for panel data is the static panel data model.[1] For simplicity, we

[1] The literature review on the static panel data model are not conducted in this thesis. For details on the model, the estimators, and the literature review, see standard textbooks on econometrics as follows: Johnston (1997) and Hayashi (2000).

will discuss the case of the model with an explanatory variable as follows:

$$y_{it} = \beta x_{it} + u_{it} \tag{1}$$

and

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

where y_{it} is the dependent variable, β is the parameter of interest to be estimated, and x_{it} is the explanatory variable. We consider the *one-way error component* model as is indicated in (2). In *the one-way error component* model, the error term u_{it} is decomposed into η_i and v_{it}, where η_i is the individual-specific effect (that captures the individual heterogeneity) and v_{it} is the disturbance.[2] In the static panel data model, we assume on the error term for $t = 1, \ldots, T$ as follows:

$$E[\eta_i] = 0,$$

$$E[v_{it}] = 0,$$

$$E[\eta_i v_{it}] = 0,$$

$$E[\eta_i^2] = \sigma_\eta^2,$$

$$E[v_{it}^2] = \sigma_v^2,$$

and

$$E[v_{it} v_{is}] = 0, \quad \text{for } t \neq s.$$

Under the assumption on the error term plus assumptions

$$E[\eta_i x_{it}] = 0$$

and

$$E[v_{it} x_{is}] = 0, \quad \text{for } t \neq s \text{ and } t = s,$$

---

[2] In addition, specifying $u_{it} = \eta_i + \mu_t + v_{it}$ instead of (2) implies the *two-way error component* model, where μ_t is the time effect.

applying OLS method to (1) makes us obtain the consistent estimator for β.

However, under the assumptions on the error term plus assumptions

$$E[\eta_i \, x_{it}] \neq 0$$

and

$$E[v_{it} \, x_{is}] = 0, \quad \text{for } t \neq s \text{ and } t = s,$$

we cannot obtain the consistent estimator for β by applying the OLS method to (1). In this case, a traditional method for consistently estimating β is as follows. In the first step, we transform (1) into the form

$$y_{it} - \bar{y}_{i\bullet} = \beta(x_{it} - \bar{x}_{i\bullet}) + (v_{it} - \bar{v}_{i\bullet}), \tag{3}$$

where defined are $\bar{y}_{i\bullet} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{x}_{i\bullet} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, and $\bar{v}_{i\bullet} = \frac{1}{T}\sum_{t=1}^{T} v_{it}$. In the second step, we apply the OLS method to (3) in order to obtain the consistent estimator for β. This is because

$$E[(x_{it} - \bar{x}_{i\bullet})(v_{it} - \bar{v}_{i\bullet})] = 0.$$

The estimator based on this procedure is called the LSDV estimator.

Further, under the assumption on the error term plus assumptions

$$E[\eta_i \, x_{it}] \neq 0$$

and

$$E[v_{it} \, x_{is}] \neq 0, \quad \text{for at least an arbitrary } s \text{ in } s = 1, \ldots, T,$$

we cannot apply the LSDV estimator to (1) for consistently estimating β. This is because

$$E[(x_{it} - \bar{x}_{i\bullet})(v_{it} - \bar{v}_{i\bullet})] \neq 0.$$

A typical case for this is the case that x_{it} is endogenous (i.e. $E[v_{it} \, x_{it}] \neq 0$). In this case, if

$$E_{t-1}[v_{it}] = 0, \tag{4}$$

where $E_{t-1}[\bullet]$ implies the expectation conditional on the information set up to time $t-1$, we can estimate β consistently by employing the GMM using a set of unconditional moment restrictions obtained from equation (4) as follows:

$$E[\Delta v_{it}\, x_{is}] = 0, \quad \text{for} \quad s = 1,\ldots,t-2 \quad \text{and} \quad t = 3,\ldots,T, \qquad (5)$$

where Δ is the first-differencing operator.[3] The moment restrictions (5) imply that in the first step, we take the first-difference of (1) to eliminate the individual effect η_i as follows:

$$\Delta y_{it} = \beta\, \Delta x_{it} + \Delta v_{it}, \qquad \text{for} \quad t = 3,\ldots,T, \qquad (6)$$

and in the second step, we use lagged explanatory variables dated $t-2$ and before as instruments for equation (6). For the most part, the panel data model based on conditional expectations has been handled in verifying the permanent income hypothesis.[4] A representative paper is Runkle (1991). This type of approach is comparatively new one.

Dynamic Panel Data Model

The characteristic of the dynamic panel data model is that the lagged dependent variable is included in regressors of the equation to be estimated. For simplicity, we discuss the case that the dependent variable is denoted by y_{it} and the lagged dependent variable dated $t-1$ (that is, $y_{i,t-1}$) only is the regressor. The model is as follows:

---

[3] The IV estimator is a non-optimal case in GMM estimators. Hayashi (2000) explains in detail on GMM estimators.

[4] This hypothesis was proposed by Hall (1978). According to the hypothesis, the consumption follows a random walk.

$$y_{it} = \alpha\, y_{i,t-1} + u_{it}, \qquad \text{for } t = 2,\ldots,T, \tag{7}$$

and

$$u_{it} = \eta_i + v_{it}, \qquad \text{for } t = 2,\ldots,T, \tag{8}$$

where α is the parameter of interest to be estimated. Equation (8) implies the *one-way error component* model, where the error term u_{it} is decomposed into the individual effect η_i and the disturbance v_{it}. In the dynamic panel data model, we assume on the error term for $t = 2,\ldots,T$ as follows:

$$E[\eta_i] = 0,$$

$$E[v_{it}] = 0,$$

$$E[\eta_i\, v_{it}] = 0,$$

$$E[\eta_i^2] = \sigma_\eta^2,$$

$$E[v_{it}^2] = \sigma_v^2,$$

and

$$E[v_{it}\, v_{is}] = 0, \qquad \text{for } t \neq s.$$

In the case of the dynamic panel data model, we also assume that

$$E[v_{it}\, y_{i1}] = 0, \qquad \text{for } t = 2,\ldots,T,$$

where y_{i1} is called the *initial condition* for the dynamic panel data model.

In the dynamic panel data model, we use neither the OLS estimator nor the LSDV estimator for the purpose of estimating α consistently. The reason for the case of the OLS estimator is because

$$E[\eta_i\, y_{i,t-1}] \neq 0, \quad \text{for at least } t = 3,\ldots,T.$$

That is, the explanatory variable $y_{i,t-1}$ is correlated with the individual effect η_i. The reason for the case of the LSDV estimators is as follows. According to the traditional practice on the LSDV estimator, in the first step, we try to transform (7) as follows:

$$y_{it} - \bar{y}_{i\bullet} = \alpha (y_{i,t-1} - \bar{y}_{i,(-1)\bullet}) + (v_{it} - \bar{v}_{i\bullet}), \quad \text{for } t = 2,\ldots,T, \quad (9)$$

where $\bar{y}_{i\bullet} = \frac{1}{T-1}\sum_{t=2}^{T} y_{it}$, $\bar{y}_{i,(-1)\bullet} = \frac{1}{T-1}\sum_{t=1}^{T-1} y_{it}$, and $\bar{v}_{i\bullet} = \frac{1}{T-1}\sum_{t=2}^{T} v_{it}$. In the second step, we cannot obtain the consistent estimator for α applying the OLS to (9). This is because

$$E[(y_{i,t-1} - \bar{y}_{i,(-1)\bullet})(v_{it} - \bar{v}_{i\bullet})] \neq 0, \quad \text{for } t = 2,\ldots,T.$$

For this sort of reason, we cannot consistently estimate the dynamic panel data model (7) in the framework of the traditional estimation techniques such as the OLS and the LSDV.[5]

Anderson and Hsiao (1982) initiated the consistent estimation of the dynamic panel data model. They used the IV method for consistently estimating the dynamic panel data model. After that, Holtz-Eakin et al. (1988) and Arellano and Bond (1991) improved their method by utilizing the instruments efficiently and introducing the optimal GMM. With the aim of using the GMM estimator, the two papers propose the following moment restrictions to estimate α consistently:

$$E[\Delta v_{it}\, y_{is}] = 0, \quad \text{for } s = 1,\ldots,t-2 \text{ and } t = 3,\ldots,T. \quad (10)$$

Equation (10) is called the *standard* moment restrictions. The *standard* moment restrictions imply that in the first step, we take the first-difference of (7) to eliminate the individual effect η_i as follows:

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta v_{it}, \quad \text{for } t = 3,\ldots,T, \quad (11)$$

5 On these issues, see Hsiao (1986).

and in the second step, we use lagged dependent variables dated $t-2$ and before as the instruments for equation (11).[6]

However, Ahn (1990), Ahn and Schmidt (1995), and Ahn and Schmidt (1999) pointed out that there are some sets of the (non-linear) moment restrictions (to estimate α consistently) overlooked in Holtz-Eakin et al. (1988) and Arellano and Bond (1991). These are

$$E[u_{it}\,\Delta v_{i,t-1}]=0, \qquad \text{for } t=4,\ldots,T, \tag{12}$$

[7]

and

$$E[u_{it}^2-u_{i,t-1}^2]=0, \qquad \text{for } t=3,\ldots,T. \tag{13}$$

[8]

Under the assumption that y_{it} is mean-stationary, Ahn and Schmidt (1995) and Ahn and Schmidt (1999) showed that the moment restrictions (12) and (13) are written in the linear form as

$$E[u_{it}\,\Delta y_{i,t-1}]=0, \qquad \text{for } t=3,\ldots,T, \tag{14}$$

[9]

---

6 On this type of estimator, there are easy-to-understand explanations in Veerbeek (2000).

7 The original form of equation (12) in Ahn and Schmidt (1995) is $E[u_{iT}\,\Delta v_{i,t-1}]=0$ for $t=4,\ldots,T$. However, Blundell and Bond (1998) rewrote it as (12) for convenience. Both are equivalent.

8 The moment restrictions (12) can be rewritten as $E[\Delta v_{it}\,y_{i,t-1}-\Delta v_{i,t-1}\,y_{i,t-2}]=0$ for $t=4,\ldots,T$ in the linear form, exploiting (13) as an assumption. See Ahn (1990).

9 The original form in Ahn and Schmidt (1995) is $E[u_{iT}\,\Delta y_{i,t-1}]=0$ for $t=3,\ldots,T$. However, Blundell and Bond (1998) rewrote it as (14) for convenience. Both are equivalent. The moment restrictions (14) are proposed in Arellano and Bover (1995). We can consider $y_{i1}=\eta_i/(1-\alpha)+w_{i1}$ as an initial condition for making y_{it} mean-stationary, where assumptions on the disturbance w_{i1} are $E[\eta_i\,w_{i1}]=0$ and $E[w_{i1}\,v_{it}]=0$ for $t=2,\ldots,T$.

and

$$E[u_{it}y_{it} - u_{i,t-1}y_{i,t-1}] = 0\,, \quad \text{for} \quad t = 3, \ldots, T\,, \tag{15}$$

respectively. Equation (14) is called the *stationarity* moment restrictions.

It is recognized that the GMM estimator using alone the moment restrictions (10) suffers a downward bias when α is near to unity and/or $\sigma_\eta^2 / \sigma_v^2$ is large. Blundell and Bond (1998) revealed that the GMM estimator using jointly both moment restrictions (10) and (14) ameliorates the downward bias, in their theoretical illustrations and Monte Carlo experiments. On these issues, see also Chapter 2.

There is a body of empirical papers using the dynamic panel data model and the GMM estimators. As the representative papers in the study on the investment behavior of firms, Blundell et al. (1991) and Bond and Meghir (1994) are recognized. In the field of labor economics, Wadhwani and Wall (1991), Konings and Walsh (1994), and Bentolila and Saint-Paul (1992) are some of the representative papers. In the estimation of the production function using panel data sets, Griffith (1999) and Blundell and Bond (2000) employed the GMM estimator incorporating the *stationarity* moment restrictions to obtain satisfactory results.

Panel Data Model with Multiplicative Individual Effects

So far, we have discussed on the panel data model with additive individual effects. Here, we will discuss on one of the panel data model with multiplicative individual effects. On discussions of the panel data model with multiplicative individual effects, see also Arellano and Honoré (2001).

A representative example of the panel data model with multiplicative individual effects is the count panel data model. An illustrative specification of the count panel data model is as follows:

$$y_{it} \sim i.i.d.Po(\lambda_{it}) , \tag{16}$$

and

$$\lambda_{it} = \exp(\gamma\, x_{it} + \eta_i) , \tag{17}$$

where y_{it} is the dependent variable, γ is the parameter of interest to be estimated, x_{it} is the explanatory variable, and η_i is the multiplicative individual effect. Expression (16) implies that y_{it} is independent and identically Poisson-distributed with the mean (and the variance) being λ_{it}. In this case, the dependent variable y_{it} is inevitably a non-negative integer value.

From (16) and (17), we can construct the following conditional moment restrictions:

$$E[y_{it} \mid x_i^t, \eta_i] = \exp(\gamma\, x_{it} + \eta_i) , \tag{18}$$

where $x_i^t = (x_{i1}, \ldots, x_{it})$. In this case, we assume that x_{it} depends on x_i^{t-1}.

From the conditional moment restrictions (18), we can construct the moment restrictions independent of the individual effect η_i as follows:

$$E[\ \{\ y_{it} \exp(-\gamma\, \Delta x_{it}\ ) - y_{i,t-1} \}\ \mid\ x_i^{t-1}] = 0 . \tag{19}$$

These are conditional moment restrictions on the basis of the quasi-difference transformation proposed by Wooldridge (1997) and Chamberlain (1992). Using the conditional moment restrictions (19), we can obtain a set of unconditional moment restrictions as follows:

$$E[\ \{\ y_{it} \exp(-\gamma\, \Delta x_{it}\ ) - y_{i,t-1} \}\ x_{is}] = 0 ,$$

$$\text{for } s = 1, \ldots, t-1 \text{ and } t = 2, \ldots, T . \tag{20}$$

We can estimate consistently the parameter of interest γ using the GMM estimator based on the unconditional moment restrictions (20).

There are some empirical researches where the GMM estimator is used based on the unconditional moment restrictions similar to (19) under the specification of the count panel data model. Montalvo (1997) analyzed the technology transfer and the relationship between patents and R&D using the data collected from Japanese firms.[10]

The method for estimating the structure of the variance in the panel data model is explained in Chapter 4.

[10] There are other papers that examine the relationship between patents and R&D using this type of approach: Crépon and Duguet (1997) and Cincera (1997).

3. Conclusion

This paper has surveyed the development of the panel data econometrics. At present, the panel data econometrics is an econometric subject that has the room for discussion.

Acknowledgement

Revised version of this paper will be published in *EKONOMIKUSU*, the bulletin of Economics Association in Kyushu Sangyô University, (Vol. 6, No. 1, September 2001) in Japanese. The title is "Recent development in panel data econometrics", ("Paneru dêta keiryô-keizaigaku no saikin no dôkô", in Japanese). I would like to thank the association for publishing the paper. Also, this paper was presented at the conference of the Western Economic Association in Japan in Kyushu Sangyo University, Fukuoka on October 27, 2001. I would like to thank Prof. Mitsuo Takase for his helpful suggestions. This paper is depicted on the web under the permission by the association.

References

Ahn, S. C. (1990) Chapter 3: Efficient estimation of models for dynamic panel data, in "Three essays on share contracts, labor supply, and the estimation of models for dynamic panel data", *Ph.D. dissertation* (Michigan State University, East Lansing, MI)

Ahn, S. C., and Schmidt, P. (1995) Efficient estimation of models for dynamic panel data, *Journal of Econometrics*, **68**, 5-28.

Ahn, S. C., and Schmidt, P. (1999) Chapter 8: Estimation of linear panel data models using GMM, in "Generalized Method of Moments Estimation", edited by L. Mátyás.

Anderson, T. W., and Hsiao, C. (1982) Formulation and estimation of dynamic models using panel data, *Journal of Econometrics*, **18**, 47-82.

Arellano, M., and Bond, S. (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies*, **58**, 277-297.

Arellano, M., and Bover, O. (1995) Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics*, **68**, 29-52.

Arellano, M., and Honoré, B. (2001) Panel Data Models: Some Recent Developments, *Handbook of Econometrics*, Volume 5.

Bentolila, S., and Saint-Paul, G. (1992) The macroeconomic impact of flexible labor contracts, with an application to Spain, *European Economic Review*, **36**, 1013-1053.

Blundell, R., and Bond, S. (1998) Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics*, **87**, 115-143.

Blundell, R., and Bond, S. (2000) GMM estimation with persistent panel data: an application to production functions, *Econometric Reviews*, **19**, 321-340.

Blundell, R., Bond, S., Devereux, M., and Schiantarelli, F. (1992) Investment and Tobin's Q: evidence from company panel data, *Journal of Econometrics*, **51**, 233-257.

Bond, S., and Meghir, C. (1994) Dynamic investment models and the firm's financial policy, *Review of Economic Studies*, **61**, 197-222.

Chamberlain, G. (1992) Comment: sequential moment restrictions in panel data, *Journal of Business & Economic Statistics*, **10**, 20-26.

Cincera, M. (1997) Patents, R&D, and technological spillovers at the firm level: some evidence from econometric count models for panel data, *Journal of Applied Econometrics*, **12**, 265-280.

Crépon, B., and Duguet, E. (1997) Estimating the innovation function from patent numbers: GMM on count panel data, *Journal of Applied Econometrics*, **12**, 243-263.

Griffith, R. (1999) Using the ARD establishment level data to look at foreign ownership and productivity in the United Kingdom, *Economic Journal*, **109**, 416-442.

Hall, R. E. (1978) Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence, *Journal of Political Economy*, **86**, 971-987.

Hansen, L. P. (1982) Large sample properties of generalized method of moments estimators, *Econometrica*, **50**, 1029-1054.

Hayashi, F. (2000) Econometrics, Princeton University Press, United States of America.

Holtz-Eakin, D., Newey, W., and Rosen, H. S. (1988) Estimating vector autoregressions with panel data, *Econometrica*, **56**, 1371-1395.

Hsiao, C. (1986) Analysis of Panel Data, Cambridge University Press, United States of America.

Johnston, J. (1997) Econometric Methods, McGraw-Hill, Singapore.

Kitazawa, Y (2000) Estimating the leverage effect using panel data with a large number of stock issues over a short-run daily period: focus on the Tokyo Stock Exchange, *Journal of Financial Management and Analysis*, **13**, 21-27.

Kitazawa, Y (2001) Exponential regression of dynamic panel data models, *Economics Letters*, **73**, 7-13.

Konings, J., and Walsh, P. P. (1994) Evidence of efficiency wage payments in UK firm level panel data, *Economic Journal*, **104**, 542-555.

Meghir, C., and Windmeijer, F. (1999) Moment conditions for dynamic panel data models with multiplicative individual effects in the conditional variance, *Annales d'Économie et de Statistique*, **55/56**, 317-330.

Montalvo, J. G. (1997) GMM estimation of count-panel-data models with fixed effects and predetermined instruments, *Journal of Business & Economic Statistics*, **15**, 82-89.

Runkle, D. E. (1991) Liquidity constraints and the permanent-income hypothesis: evidence from panel data, *Journal of Monetary Economics*, **27**, 73-98.

Veerbeek, M. (2000) A Guide to Modern Econometrics, John Wiley & Sons, Ltd., England.

Wadhwani, S. B., and Wall, M. (1991) A direct test of the efficiency wage model using UK micro-data, *Oxford Economic Papers*, **43**, 529-548.

Wooldridge, J. M. (1997) Multiplicative panel data models without the strict exogeneity assumption, *Econometric Theory*, **13**, 667-678.