



Discussion Paper Series

No.217

**An Alternative System GMM Estimation
in Dynamic Panel Models**

Hosung Jung
Hyeog Ug Kwon

July 2007

**Hitotsubashi University Research Unit
for Statistical Analysis in Social Sciences**

A 21st-Century COE Program

Institute of Economic Research
Hitotsubashi University
Kunitachi, Tokyo, 186-8603 Japan
<http://hi-stat.ier.hit-u.ac.jp/>

An Alternative System GMM Estimation in Dynamic Panel Models

Hosung Jung* & Hyeog Ug Kwon†

July 22, 2007

Abstract

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 known to be more efficient than the first-difference GMM estimator. However, an initial optimal weight matrix is not known for the system estimation procedure. Therefore, we suggest the use of ‘a suboptimal weight matrix’ which may reduce the finite sample bias whilst increasing its efficiency. Using the Kantorovich inequality, we find that the potential efficiency gain becomes large when the variance of individual effects increases compared to the variance of the idiosyncratic errors. (Our Monte Carlo experiments show that the small sample properties of the suboptimal system estimator are shown to be much more reliable than any other conventional system GMM estimator in terms of bias and efficiency.

Keywords: Dynamic panel data; sub-optimal weighting matrix; KI upper bound

1 Introduction

It is generally known that using many instruments can improve the efficiency of various IV and GMM estimators (Arellano and Bover, 1995; Blundell and Bond, 1998; Ahn and Schmidt, 1995; etc.). Therefore, the system GMM estimator in dynamic panel data models is more efficient than the first-difference GMM estimator.¹ Despite the substantial efficiency gain, using many instruments has two important drawbacks: increased bias and unreliable inference (Newey and Smith, 2004; Hayakawa, 2005). In this paper, we investigate how to decrease bias while increasing efficiency in the system GMM estimation. Instead of adjusting the number of instrumental variables, we suggest an alternative way of improving efficiency.

In general, an asymptotically efficient estimator can be obtained through the two-step procedure in the standard GMM estimation. However, the estimated standard error can be biased downwards quite severely for moderate sample sizes, N (Windmeijer, 1998). It is obvious that the same problem persists even in the case of the two-step system GMM estimation. In practice, therefore, we often rely on an inference based on the less efficient one-step estimator, whose inference is much more reliable than that of the two-step estimator. Under this constraint, it becomes important to choose the weight matrix in the first step, especially in small samples. Unfortunately, the optimal weight matrix for the system estimator is only available when the variance of individual effects is

*Samsung Economic Research Institute, e-mail:hosungj.jung@samsung.com

†Department of Economics, Nihon University, e-mail:kwon@eco.nihon-u.ac.jp

¹See Blundell and Bond (1998) for details.

zero. Hence, we suggest using a suboptimal weight matrix which contains the estimated variance ratio of the individual effects to that of the idiosyncratic error term. This yields the suboptimal system GMM (SYS_{sub} hereafter) estimation.²

To investigate the magnitude of the efficiency gain, KI upper bounds based on the Kantorovich inequality (Windmeijer, 1998) are applied. We find that the efficiency gain can potentially be large when the variance of individual effects increases. In addition, we conduct Monte Carlo studies to confirm the efficiency gain from using the SYS_{sub} estimator when compared to the conventional system GMM estimation in Blundell and Bond (1998). While the small-sample properties of the conventional system estimators are heavily affected by the increase of the variance ratio, the SYS_{sub} estimator is relatively reliable. As an empirical example, we estimate the Cobb-Douglas production function of a balanced panel of 1,002 Japanese manufacturing companies for the period 1991-2001.

The remainder of this paper is organized as follows. The next section presents the model and reviews the conventional system GMM estimation. In Section 2, we propose the SYS_{sub} estimation and consider the efficiency gain against using the identity matrix as an initial weight matrix. Section 3 reports the simulation results, while Section 4 present an empirical application to a production function. Section 5 concludes. The notation is fairly standard and self-explanatory: ‘ \rightarrow ’ denotes convergence in probability while ‘ \sim ’ or ‘ \Rightarrow ’ is used for convergence in distribution. The nonstochastic limit of a sequence is also denoted by ‘ \rightarrow ’ when the context makes the usage clear.

2 Models and the System GMM Estimator

To analyze the properties of the parameter estimators in the system GMM estimation, we consider a simple dynamic panel model with an autoregressive specification and a one-way error component, u_{it} :

$$\begin{aligned} y_{it} &= \alpha y_{it-1} + u_{it}, & |\alpha| < 1. \\ u_{it} &= \mu_i + v_{it} & (i = 1, \dots, N; t = 2, \dots, T), \end{aligned} \quad (1)$$

where $\mu_i \sim iid(0, \sigma_\mu^2)$ and $v_{it} \sim iid(0, \sigma_v^2)$. To begin with, we assume that μ_i and v_{it} have the familiar error component structure in which

$$E(\mu_i) = E(v_{it}) = E(\mu_i v_{it}) = 0 \quad \forall \quad i, t \quad (2)$$

and

$$E(v_{it} v_{is}) = 0. \quad \forall \quad i, t \neq s \quad (3)$$

The y_{it} series are assumed to be stationary and the series can alternatively be written as

$$y_{it} = \frac{\mu_i}{1 - \alpha} + \sum_{j=0}^{\infty} \alpha^j v_{i,t-j} \quad (4)$$

We define the variance ratio, $\rho_i = \frac{\sigma_\mu^2}{\sigma_v^2}$, for later use. The system GMM estimator combines moment conditions for the differenced equation with moment conditions for the model in levels. Adopting the standard assumptions concerning the error components (i.e., the white-noise error v_{it}), Blundell and Bond (1998) noted the validity of the following $m_s = (T + 1)(T - 2)/2$ linear moment restrictions for each i ,

$$E[y_{i,t-j} \Delta u_{it}] = 0 \quad \text{for } (j = 2, \dots, t - 1; t = 3, \dots, T) \quad (5)$$

$$E[\Delta y_{i,t-1} u_{it}] = 0 \quad \text{for } (t = 3, \dots, T). \quad (6)$$

²As we need the first-step estimation to obtain the variance ratio, this estimation can be categorized as a two-step GMM estimation. However, unlike the conventional two-step GMM estimation, we show that SYS_{sub} does not suffer from a downward bias of its estimated standard error.

For convenience, the moment restrictions can be expressed more compactly as

$$E[f_i(\alpha_0)] = E(Z'_{si}q_i) = \mathbf{0} \quad (7)$$

where

$$q_i = \begin{bmatrix} \Delta u_i \\ u_i \end{bmatrix}. \quad (8)$$

Z_{si} is a $2(T-2) \times m$ block diagonal matrix given by

$$Z_{si} = \begin{bmatrix} Z_{di} & \mathbf{0} \\ \mathbf{0} & Z_{li} \end{bmatrix} \quad (9)$$

where Z_{di} and Z_{li} refer to the instruments in the first-differenced equation and the levels equation, respectively. These are given by

$$Z_{di} = \begin{bmatrix} [y_{i1}] & \cdots & & 0 \\ & [y_{i1}, y_{i2}] & & \vdots \\ & & \ddots & \\ 0 & \cdots & & [y_{i1}, \cdots, y_{iT-2}] \end{bmatrix} \quad (10)$$

and

$$Z_{li} = \text{diag}[\Delta y_{i2}, \Delta y_{i3}, \dots, \Delta y_{iT-1}]. \quad (11)$$

Let Z_s be a $2N(T-2) \times m_s$ matrix consisting of $(Z_{s,1}, \dots, Z_{s,N})$ and Y be a stacked matrix $Y_i = (\Delta y_i, y_i)$. Then, the one-step system GMM estimator based on these moment conditions (7) is

$$\hat{\alpha}_s = [Y'_{-1}Z_sW_NZ'_sY_{-1}]^{-1} [Y'_{-1}Z_sW_NZ'_sY] \quad (12)$$

for some positive weight matrix W_N . The efficient two-step system GMM estimator is obtained in a similar way to the standard GMM procedure.

In panel data models, the estimated standard error can be substantially biased downward; therefore, we often rely on inference based on the less efficient one-step estimator. In this case, there is no one-step system GMM estimator that is asymptotically equivalent to the two-step estimator, unless $\sigma_\mu^2 = 0$. As a natural choice for W_N to yield the initial consistent estimator, Blundell and Bond (1998) used

$$W_N = \left[\frac{1}{N} \sum_{i=1}^N Z'_{si}H_sZ_{si} \right]^{-1}, \quad (13)$$

where

$$H_s = \begin{bmatrix} H_d & \mathbf{0} \\ \mathbf{0} & I_{T-2} \end{bmatrix}. \quad (14)$$

While the submatrix H_d —a $(T-2)$ square matrix which has twos in the main diagonal, minus ones in the first subdiagonals, and zeros otherwise—is used for the first-differenced equation, the identity matrix is used for the level estimation. This implies that the variance-covariance structure of residuals from the level estimation is not considered in the system estimation. Therefore, even if the weight matrix H_s works well for small values of σ_μ^2 , there exists the potential for efficiency loss when σ_μ^2 gets large.

3 A Suboptimal Weight Matrix

In large samples, the efficiency of the one-step system GMM estimator is not affected by the choice of the weight matrix, as long as the matrix is positive definite. Therefore, the efficiency gain for the two-step procedure may not be substantial asymptotically. However, there is no one-step system GMM estimator that is asymptotically equivalent to the two-step estimator, even in the special case of i.i.d. disturbances. Only in the case of $\sigma_\mu^2 = 0$ is an optimal weight matrix for the system GMM estimator given by

$$H_s = \begin{bmatrix} H_d & C \\ C & I_{T-2} \end{bmatrix} \quad (15)$$

where C is a $(T - 2)$ square matrix which has ones in the main diagonal, minus ones in the first lower subdiagonals and zeros otherwise.³ Unless $\sigma_\mu^2 = 0$, the identity matrix I_{T-2} in H_s should be replaced by the matrix J_{T-2} to achieve optimality, where

$$J_{T-2} = \begin{bmatrix} 1 + \rho & \rho & \rho & \cdots & \rho \\ \rho & 1 + \rho & \rho & \cdots & \rho \\ \rho & \rho & 1 + \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 + \rho \end{bmatrix}, \quad (16)$$

which yields the optimal weight matrix, H_{so} :

$$H_{so} = \begin{bmatrix} H_d & C \\ C & J_{T-2} \end{bmatrix}. \quad (17)$$

Although the system estimator in Blundell and Bond (1998) performs well as long as ρ is reasonably small, there are always cases where the variance of the individual effects, μ_i , is substantially larger than that of the classical error term, v_{it} . The use of the weight matrix H_{so} , therefore, can be described as inducing cross-sectional heterogeneity through ρ . Otherwise, using the matrix J_{T-2} can be explained as partially adopting a procedure of GLS (generalized least squares) to the level estimation, which is not done in Blundell and Bond (1998). However, since the variance ratio, ρ , is unknown in practice, we suggest the estimate of the optimal weight matrix, H_{so}

$$\widehat{H}_{so} = \begin{bmatrix} H_d & \mathbf{0} \\ \mathbf{0} & \hat{J}_{T-2} \end{bmatrix}, \quad (18)$$

where a natural estimator for the variance ratio $\hat{\rho}$ is readily available in the initial step of the system estimation. To obtain $\hat{\rho}$, we derive $\hat{\sigma}_v^2$ from the first-difference GMM estimation,

$$\hat{\sigma}_v^2 = \frac{\sum_{i=1}^N \Delta \hat{u}_i' \Delta \hat{u}_i}{2N(T-2)}, \quad (19)$$

while σ_μ^2 is given by

$$\frac{\sum_{i=1}^N [\tilde{u}_i' \tilde{u}_i - \Delta \tilde{u}_i' \Delta \tilde{u}_i / 2]}{N(T-2)}, \quad (20)$$

where $\Delta \hat{u}_i$ and \tilde{u}_i are residuals from the first difference and the level equation, respectively. Using this weight matrix, \widehat{H}_{so} , instead of the matrix H_s , may improve the efficiency of the second-step system estimation when ρ becomes large.⁴

³Also see Windmeijer (1998) for details.

⁴If ρ is small, the potential efficiency gain gets smaller

4 Efficiency Gains

To measure the efficiency gain, we use the KI upper bounds. Using the moment condition (7), the system GMM estimator $\hat{\alpha}_s$ for α_0 minimizes

$$\hat{\alpha}_s = \operatorname{argmin}_{\alpha_0} \left[\frac{1}{N} \sum_{i=1}^N f_i(\alpha) \right]' W_N \left[\frac{1}{N} \sum_{i=1}^N f_i(\alpha) \right] \quad (21)$$

where W_N is a positive definite weight matrix that satisfies $N \rightarrow \infty W_N = W$. Furthermore, if we assume that

$$\frac{1}{\sqrt{N}} f_i(\alpha_0) \rightarrow N(0, \Psi), \quad (22)$$

where the regularity conditions are in place and $F_\alpha = E(\partial f_i(\alpha)/\partial \alpha)$, $F_0 \equiv F_{\alpha_0}$, then $\sqrt{N}(\hat{\alpha}_s - \alpha_0)$ has a limiting normal distribution,

$$\sqrt{N}(\hat{\alpha}_s - \alpha_0) \rightarrow N(0, V_W), \quad (23)$$

where $V_W = (F_0' W F_0)^{-1} F_0' W \Psi W F_0 (F_0' W F_0)^{-1}$. An optimal choice for W is Ψ^{-1} , so the asymptotic variance matrix is given by $(F_0' W F_0)^{-1}$. Clearly, the following inequality holds for any positive matrix W :

$$(F_0' \Psi^{-1} F_0)^{-1} \leq (F_0' W F_0)^{-1} F_0' W \Psi W F_0 (F_0' W F_0)^{-1} \quad (24)$$

According to Liu and Neudecker (1997), the following inequality also holds:

$$(F_0' W F_0)^{-1} F_0' W \Psi W F_0 (F_0' W F_0)^{-1} \leq \frac{(\lambda_1 + \lambda_p)^2}{4\lambda_1\lambda_p} (F_0' \Psi^{-1} F_0)^{-1}, \quad (25)$$

and the KI upper bounds – $\text{KI}_{ub} = \frac{(\lambda_1 + \lambda_p)^2}{4\lambda_1\lambda_p}$ – are calculated, where $\lambda_i > 0$ ($i = 1, \dots, p$) are the eigenvalues of the $p \times p$ matrix ΨW .⁵ If we use an initial weight matrix equal to \hat{H}_{so} , Ψ is obtained by

$$\Psi = \frac{1}{N} \sum_{i=1}^N [Z_i' \hat{H}_{so} Z_i], \quad (26)$$

and the asymptotic variance matrix for using the suboptimal weighting matrix \hat{H}_{so} then is $(F_0' \Psi^{-1} F_0)^{-1}$. For $T = 4$, for example, with four overidentifying moment conditions, the matrices Ψ and W_1 are given by

$$\begin{aligned} \Psi &= \frac{1}{N} \sum_{i=1}^N [Z_{si}' H_{so} Z_{si}] \\ &= \begin{bmatrix} 2\sigma_y^2 & -\sigma_y^2 & -\delta & 0 & 0 \\ -\sigma_y^2 & 2\sigma_y^2 & 2\delta & 0 & 0 \\ -\delta & 2\delta & 2\sigma_y^2 & 0 & 0 \\ 0 & 0 & 0 & \frac{2(1+\rho)\sigma_v^2}{1+\alpha} & \frac{-\rho(1-\alpha)\sigma_v^2}{1+\alpha} \\ 0 & 0 & 0 & \frac{-\rho(1-\alpha)\sigma_v^2}{1+\alpha} & \frac{2(1+\rho)\sigma_v^2}{1+\alpha} \end{bmatrix} \end{aligned} \quad (27)$$

and

$$\begin{aligned} W_1 &= \frac{1}{N} \sum_{i=1}^N [Z_{si}' Z_{si}]^{-1} \\ &= \begin{bmatrix} \sigma_y^2 & 0 & 0 & 0 & 0 \\ 0 & \sigma_y^2 & \delta & 0 & 0 \\ 0 & \delta & \sigma_y^2 & 0 & 0 \\ 0 & 0 & 0 & \frac{2\sigma_v^2}{1+\alpha} & 0 \\ 0 & 0 & 0 & 0 & \frac{2\sigma_v^2}{1+\alpha} \end{bmatrix}^{-1}, \end{aligned} \quad (28)$$

⁵Also see Liu and Neudecker (1997) for details

where $\delta = \sigma_y^2 - \frac{\sigma_v^2}{1+\alpha}$. Hence,

$$\Psi W_1 = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 & 0 \\ -\frac{\delta}{\sigma_y^2} & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & (1+\rho) & \frac{-\rho(1-\alpha)}{2} \\ 0 & 0 & 0 & \frac{-\rho(1-\alpha)}{2} & (1+\rho) \end{bmatrix}. \quad (29)$$

We easily find that the eigenvalues of the upper block matrix in (29) are 1, 2 and 3, which are fixed for any values of δ and σ_y^2 . The eigenvalues of the lower block matrix are functions of ρ and α :

$$\begin{aligned} \lambda_{low} &= \left[(1+\rho) \mp \frac{-\rho(1-\alpha)}{2} \right] \\ &= \left[\frac{2+3\rho-\alpha\rho}{2}, \frac{2+\rho+\alpha\rho}{2} \right]. \end{aligned} \quad (30)$$

If $\rho \leq 1.3$, the eigenvalues of the lower block matrix in (29) are in [1,3] so that the minimum and maximum of the eigenvalues of the whole matrix ΨW_1 are 1 and 3, respectively. This implies that the efficiency loss of the one-step system GMM estimator with the identity matrix is around 30% compared to the suboptimal system GMM estimator.

5 Monte Carlo Experiments

This section illustrates the small-sample performance of the various system GMM estimators. Monte Carlo experiments were carried out based on a data generating process following Nerlove (1971) and Blundell and Bond (1998).

$$y_{it} = \alpha y_{it-1} + u_{it}, \quad (31)$$

for $i = 1, 2, \dots, N$ and $t = 2, 3, \dots, T$. For the random-effects specification, we generate $u_{it} = \mu_i + v_{it}$, where $\mu_i \sim iid(0, \sigma_\mu^2)$. All the innovations are independent over time and are homoscedastic; that is, $v_{it} \sim NID(0, 1)$. We generate the initial conditions y_{i1} as

$$y_{i1} = \frac{\mu_i}{1-\alpha} + w_{i1}, \quad (32)$$

where w_{i1} is an $NID(0, \sigma_{w1})$ random variable, independent of both μ_i and v_{it} with the variance σ_{w1} chosen to satisfy covariance stationarity.

The variance ratio, ρ , is characterized by $\frac{\sigma_\mu^2}{\sigma_v^2}$, so it depends only on σ_μ^2 . Throughout the experiments, eighteen parameter settings (i.e., $\alpha = 0.2, 0.5, 0.8$ and $\rho = 0, 0.5, 1, 2, 5, 10$) are simulated. To compare the small-sample performance, the five different system GMM estimation procedures are considered according to their weight matrix. Specifically, ISYS denotes the first-step estimator, which uses the identity matrix, while the one- and two-step system GMM estimation in Blundell and Bond (1998) are named SYS_1 and SYS_2 , respectively. Furthermore, SYS_3 uses the alternative suboptimal weighting matrix defined in (15). SYS_{sub} denotes the newly proposed suboptimal weight matrix (18), which uses the estimated ρ , while $TSYS_{sub}$ uses the true ρ .⁶

Tables 1 and 2 present the estimation results for $T = 5$ and 10, respectively. Clearly, the bias and the standard deviations of all the estimators are affected by the variance ratio ρ . While the biases of ISYS, SYS_1 , SYS_2 and SYS_3 are negligible when $\rho \leq 1$, they

⁶In one of the most widespread statistics programs, STATA, ISYS, SYS_1 and SYS_3 are derived by choosing $h(1)$, $h(2)$ and $h(3)$, which determine the first-step weight matrix from among three options.

Table 1: Small-sample properties of various GMM estimators (T=5)

	ρ	ISYS	SYS ₁	SYS ₂	SYS ₃	SYS _{sub}	TSYS _{sub}
$\alpha = 0.2$ Mean	0	0.1875	0.1960	0.1983	0.1970	0.1958	0.1960
	0.5	0.1889	0.1980	0.2030	0.2127	0.1945	0.1948
	1	0.1933	0.2076	0.2085	0.2287	0.1992	0.1991
	2	0.2122	0.2238	0.2166	0.2572	0.2048	0.2029
	5	0.2484	0.2688	0.2397	0.3254	0.2201	0.2112
	10	0.3031	0.3251	0.2840	0.4149	0.2428	0.2173
Std.	0	0.0875	0.0764	0.0757	0.0700	0.0765	0.0764
	0.5	0.0966	0.0851	0.0817	0.0839	0.0857	0.0850
	1	0.0986	0.0941	0.0876	0.0947	0.0948	0.0933
	2	0.1074	0.1059	0.0906	0.1152	0.1053	0.1027
	5	0.1286	0.1270	0.1091	0.1496	0.1223	0.1168
	10	0.1541	0.1556	0.1381	0.1877	0.1379	0.1249
$\alpha = 0.5$ Mean	0	0.4740	0.4895	0.4927	0.4913	0.4890	0.4895
	0.5	0.4824	0.5030	0.5053	0.5179	0.4959	0.4966
	1	0.4984	0.5157	0.5155	0.5371	0.5018	0.5011
	2	0.5227	0.5440	0.5372	0.5729	0.5180	0.5113
	5	0.5898	0.6074	0.5877	0.6604	0.5590	0.5289
	10	0.6664	0.6848	0.6604	0.7535	0.6217	0.5495
Std.	0	0.0981	0.0860	0.0816	0.0741	0.0861	0.0860
	0.5	0.1099	0.1016	0.0960	0.0948	0.1048	0.1024
	1	0.1182	0.1115	0.1043	0.1070	0.1164	0.1127
	2	0.1313	0.1261	0.1150	0.1235	0.1323	0.1264
	5	0.1458	0.1429	0.1359	0.1489	0.1558	0.1397
	10	0.1555	0.1595	0.1589	0.1669	0.1823	0.1539
$\alpha = 0.8$ Mean	0	0.7522	0.7786	0.7883	0.7867	0.7772	0.7786
	0.5	0.7851	0.8116	0.8031	0.8305	0.7990	0.8003
	1	0.8192	0.8429	0.8310	0.8649	0.8265	0.8210
	2	0.8693	0.8869	0.8730	0.9086	0.8685	0.8486
	5	0.9286	0.9408	0.9286	0.9531	0.9229	0.8837
	10	0.9559	0.9681	0.9609	0.9739	0.9550	0.9058
Std.	0	0.1070	0.0954	0.0887	0.0789	0.0963	0.0954
	0.5	0.1219	0.1123	0.1104	0.0939	0.1252	0.1147
	1	0.1246	0.1148	0.1177	0.0999	0.1326	0.1178
	2	0.1175	0.1094	0.1180	0.0932	0.1349	0.1158
	5	0.0941	0.0916	0.1167	0.0821	0.1218	0.1078
	10	0.0817	0.0697	0.0886	0.0625	0.0930	0.0956

Notes: (1) 5,000 replications with $N = 100$. (2) Std. refers to standard deviation.

Table 2: Small-sample properties of various GMM estimators (T=10)

	ρ	ISYS	SYS ₁	SYS ₂	SYS ₃	SYS _{sub}	TSYS _{sub}
$\alpha = 0.2$ Mean	0	0.1624	0.1935	0.1961	0.1991	0.1934	0.1935
	0.5	0.1642	0.1981	0.1996	0.2354	0.1951	0.1952
	1	0.1684	0.2038	0.2023	0.2658	0.1959	0.1958
	2	0.1794	0.2172	0.2121	0.3187	0.1987	0.1980
	5	0.2153	0.2561	0.2406	0.4391	0.2045	0.2010
	10	0.2708	0.3100	0.2855	0.5602	0.2130	0.2039
Std.	0	0.0538	0.0419	0.0437	0.0375	0.0419	0.0419
	0.5	0.0561	0.0470	0.0473	0.0475	0.0472	0.0470
	1	0.0587	0.0493	0.0477	0.0565	0.0496	0.0490
	2	0.0596	0.0518	0.0489	0.0758	0.0503	0.0493
	5	0.0654	0.0622	0.0568	0.1015	0.0547	0.0523
	10	0.0789	0.0761	0.0727	0.1142	0.0564	0.0519
$\alpha = 0.5$ Mean	0	0.4368	0.4871	0.4913	0.4958	0.4870	0.4871
	0.5	0.4452	0.4963	0.4983	0.5439	0.4907	0.4908
	1	0.4601	0.5108	0.5084	0.5843	0.4960	0.4951
	2	0.4910	0.5367	0.5278	0.6466	0.5045	0.5006
	5	0.5634	0.6005	0.5834	0.7523	0.5265	0.5074
	10	0.6465	0.6720	0.6539	0.8352	0.5600	0.5114
Std.	0	0.0589	0.0440	0.0446	0.0377	0.0441	0.0440
	0.5	0.0650	0.0532	0.0506	0.0513	0.0541	0.0530
	1	0.0669	0.0561	0.0530	0.0617	0.0574	0.0552
	2	0.0711	0.0586	0.0549	0.0683	0.0601	0.0561
	5	0.0799	0.0689	0.0680	0.0786	0.0690	0.0589
	10	0.0890	0.0760	0.0798	0.0711	0.0821	0.0589
$\alpha = 0.8$ Mean	0	0.7036	0.7770	0.7849	0.7939	0.7766	0.7770
	0.5	0.7530	0.8097	0.8091	0.8641	0.7963	0.7947
	1	0.7973	0.8387	0.8343	0.8976	0.8179	0.8081
	2	0.8502	0.8777	0.8724	0.9329	0.8538	0.8250
	5	0.9124	0.9310	0.9270	0.9673	0.9131	0.8474
	10	0.9501	0.9599	0.9580	0.9828	0.9498	0.8613
Std.	0	0.0656	0.0476	0.0454	0.0351	0.0478	0.0476
	0.5	0.0711	0.0577	0.0568	0.0412	0.0643	0.0576
	1	0.0691	0.0564	0.0570	0.0397	0.0684	0.0581
	2	0.0634	0.0533	0.0556	0.0361	0.0710	0.0578
	5	0.0493	0.0419	0.0451	0.0287	0.0625	0.0562
	10	0.0352	0.0318	0.0350	0.0215	0.0470	0.0534

Table 3: Estimation Results of ρ

α	$\rho = 0$	0.5	1	2	5	10
0.2	0.0099	0.5228	1.0167	1.9035	4.1814	6.9826
0.5	0.0223	0.5384	1.0126	1.7680	3.3160	4.6412
0.8	0.0514	0.6569	1.0037	1.2449	1.3974	1.3014

Notes: (1) 5,000 replications with $T = 5, N = 100$.

rapidly increase with ρ . The biases of SYS_{sub} and TSYS_{sub} show a much slower increase due to an increase in ρ . Even in the case of $\alpha = 0.8$, the two estimators show the smallest increase in mean. On the other hand, SYS_{sub} and TSYS_{sub} in most cases have smaller variance than any of the other estimators except SYS_2 .⁷ Consequently, we conclude that SYS_{sub} outperforms the conventional system estimators in terms of bias and efficiency. However, the advantage of the SYS_{sub} estimator decreases as α grows to unity because a high α leads to an unreliable estimate of ρ itself.

Table 3 presents the estimation results of ρ based on residuals from the first-step system GMM estimator. The mean of the estimated ρ has substantial bias when α is close to one, which yields no considerable improvement to using the suboptimal system procedure. Even though the suggested estimator, SYS_{sub} , depends on the results from the first-step estimation of ρ , in most cases, SYS_{sub} performs better than any of the other conventional system estimators widely in use.

6 Empirical Application: Estimation of Production Functions using Japanese Firm-level Panel Data

We apply the suboptimal system GMM estimation procedure (denoted SYS_{sub}) to the estimation of production functions using firm-level balanced panel data for 1,002 Japanese manufacturing firms. As highlighted by Griliches and Mairesse (1995), there are many econometric problems involved in the estimation of production functions, including unobserved heterogeneity between firms, simultaneity of the decisions about inputs and output, and measurement errors in inputs. We compare our result with the results from the different estimation approaches that have been proposed to deal with these problems, such as OLS, LSDV, GMM and system GMM.

We estimate

$$y_{it} = \beta_m M_{it} + \beta_l L_{it} + \beta_k K_{it} + \gamma_t + (\mu_i + v_{it} + m_{it}) \quad (33)$$

$$v_{it} = \alpha v_{i,t-1} + e_{it} \quad |\alpha| < 1, \quad (34)$$

where y_{it} is the log of firm i 's sales in year t , M_{it} is the log of intermediate inputs, L_{it} is the log of employment, K_{it} is the log of capital stock, and γ_t is a year-specific intercept reflecting, for example, a common technology shock. As for the error components, μ_i is an unobserved firm-specific effect, v_{it} is a possibly autoregressive productivity shock, and m_{it} is measurement error. We assume that m_{it} and e_{it} are serially uncorrelated. As all independent variables are potentially correlated with the individual-specific effects and with productivity shocks, no valid moment conditions for specification (34) exist as long as $\alpha \neq 0$. However, this model has a dynamic common factor representation:

$$y_{it} = \alpha y_{i,t-1} + \beta_m M_{it} - \alpha \beta_m M_{i,t-1} + \beta_l L_{it} - \alpha \beta_l L_{i,t-1} + \beta_k K_{it} - \alpha \beta_k K_{i,t-1} \quad (35)$$

⁷The two-step system estimators are known to have downward bias in the estimated standard errors.

$$+(\gamma_t - \alpha\gamma_{t-1}) + (\mu_i(1 - \alpha) + e_{it} + m_{it} - \alpha m_{i,t-1})$$

or

$$y_{it} = \pi_1 y_{i,t-1} + \pi_2 M_{it} + \pi_3 M_{i,t-1} + \pi_4 L_{it} + \pi_5 L_{i,t-1} + \pi_6 K_{it} + \pi_7 K_{i,t-1} \quad (36)$$

$$\dot{\gamma}_t + (\dot{\mu}_i + w_{it}),$$

subject to the three nonlinear common factor restrictions $\pi_3 = -\pi_1\pi_2$, $\pi_5 = -\pi_1\pi_4$ and $\pi_7 = -\pi_1\pi_6$. On the other hand, the error term $w_{it} = e_{it} + m_{it} - \alpha m_{i,t-1}$ is serially uncorrelated if there are no measurement errors or $w_{it} \sim \text{MA}(1)$ if there are measurement errors in some of the series. Although consistent estimates of the unrestricted parameters, $\pi = (\pi_1, \dots, \pi_7)$, are possible in either case, we assume there is no measurement error, i.e., $m_{it} = 0$, for convenience. Using the suboptimal system GMM methods outlined in the previous sections, we present the consistent estimates of π and $\text{var}(\pi)$.

Table 4 presents the various estimation results. The key element we are interested in is the sign of the coefficient estimates and their significance. Typically, the coefficients of labor, capital, and intermediate inputs will tend to be biased upward in pooled OLS, whereas the LSDV estimator controlled for unobserved heterogeneity provides very small estimates of capital (see the survey by Griliches and Mairesse (1995)).

In the estimation results reported in Table 4, the value of the LSDV estimate of capital is negative but significant. The low coefficient on capital may be caused by measurement errors in the calculation of capital stocks and the difficulty of rapid adjustments in response to exogenous shocks like demand shifts or productivity shocks. Our result obtained by the LSDV estimator is consistent with previous findings by Blundell and Bond (1998) and Black and Lynch (2001).

In order to control for these two problems, the first differences GMM (denoted GMM1) is used in the estimation of the production function. Unfortunately, the first differences GMM does not remedy the two distortions of the LSDV estimator, showing that the estimated coefficient of capital is smaller than that of the LSDV estimation. Overidentifying restrictions are also rejected. These findings result from the weak instruments problem in dynamic models where regressors in first differences are weakly autocorrelated and from the exacerbation of the measurement error problem caused by the elimination of the cross-sectional variation through first differencing.

These findings indicate that the measurement error downward bias in capital is clearly in excess of the upward bias caused by simultaneity. Blundell and Bond (2000) suggest an alternative estimator that corrects these problems in the first differenced GMM estimators, which they called the system GMM. Blundell and Bond (2000) and Alonso-Borrego and Sanchez-Mangas (2001), using UK and Spanish data, respectively, show that the system GMM estimation performs very well. In the system GMM estimation, the equation in differences is instrumented by lagged differences (while the equation in levels is additionally instrumented by suitably lagged differences. A reason that the system GMM works better than the first differenced GMM is that the second set of moment conditions reduces weak instruments problems and the large measurement error in capital.

Table 4 also reports the coefficient estimates from three different methods of the system GMM. When applying the system GMM using the one-step identity weight matrix (ISYS), we obtained the result that the coefficients on the lagged dependent variable and capital were larger than those of the first differenced GMM. The Sargan-statistic does not reject the validity of the instruments. However, the coefficient on capital is not still significant. Even when using SYS_1 , we still found a large coefficient on the lagged dependent variable and an insignificant coefficient on capital, while the coefficient on labor was considerably smaller than the result of ISYS, which is an unexpected result. By contrast, the specification of SYS_3 corrects for large measurement error in the differences of capital while the test of the validity of the instruments and second order autocorrelation is not

Table 4: Results of the Estimation of the Cobb-Douglas Production Function

	REG	OLS	LSDV	GMM1	ISYS	SYS ₁	SYS ₃	SYS _{sub}
coe.	LogQ ₁	0.8782	0.3384	0.1765	0.3033	0.4070	0.6938	0.3808
	LogM	0.7259	0.7495	0.7330	0.9078	0.9158	0.7954	0.8509
	LogL	0.1887	0.1101	0.1118	0.2154	0.0853	0.1407	0.1013
	LogK	0.0177	-0.038	-0.0273	0.0004	0.1060	0.1272	0.0588
	LogM ₁	-0.6313	-0.1329	-0.0264	-0.2245	-0.3445	-0.5042	-0.2300
	LogL ₁	-0.1693	-0.0574	-0.0627	-0.0718	-0.0576	-0.1037	-0.0272
	LogK ₁	-0.0057	0.0307	0.0906	-0.0458	-0.0842	-0.1138	-0.1108
ste.	LogQ ₁	0.0058	0.0084	0.0440	0.0359	0.0309	0.0152	0.0161
	LogM	0.0044	0.0045	0.0057	0.0233	0.0278	0.0160	0.0123
	LogL	0.0093	0.0088	0.0127	0.0413	0.0383	0.0223	0.0186
	LogK	0.0086	0.0081	0.0123	0.0369	0.0363	0.0219	0.0178
	LogM ₁	0.0062	0.0075	0.0324	0.0297	0.0293	0.0168	0.0159
	LogL ₁	0.0093	0.0087	0.0136	0.0441	0.0415	0.0232	0.0172
	LogK ₁	0.0087	0.0082	0.0123	0.0383	0.0386	0.0235	0.0167
t-val.	LogQ ₁	151.770	40.437	4.010	8.440	13.440	45.610	23.620
	LogM	165.393	165.922	128.450	39.010	31.690	49.610	71.824
	LogL	20.190	12.443	8.780	5.210	4.200	6.300	5.444
	LogK	2.057	-4.6703	-2.230	(0.010)	(1.250)	5.800	3.297
	LogM ₁	-102.015	-17.644	(-0.820)	-7.550	-9.420	-29.930	-14.465
	LogL ₁	-18.226	-6.616	-4.610	(-1.630)	(-1.280)	-4.470	(-1.583)
	LogK ₁	(-0.653)	3.753	7.380	(-1.200)	-2.810	-4.850	-6.631
p-val.			<i>m</i> ₁ -	0.000	0.000	0.000	0.000	0.000
			<i>m</i> ₂ -	0.228	0.041	0.167	0.000	0.153
			Sargan-	0.000	0.775	0.811	0.000	0.796

Notes: (1) The year dummy included in the estimation is not reported here. (2) 5% critical values are used in the specification tests. (3) ISYS, SYS₁ and SYS₃ are based on their one-step estimation. (4) The initially estimated value, $\hat{\rho} = 3.6237$, is used for SYS_{sub}. (5) *LogQ*₁ refers to the lagged levels at $t - 1$.

accepted. This indicates that the system GMM estimator using SYS_3 will be inconsistent. The limitation of the system GMM is that it cannot obtain consistent estimates because it does not consider the fixed effects in level equation.

The results show that the alternative estimation suggested in this paper helps to alleviate the econometric problems in the estimation of production functions such as unobserved heterogeneity, simultaneity, and measurement errors in intermediate inputs. For example, the estimate of the coefficient on capital is positive and significant, while the coefficient on inputs is quite realistic. The specification tests suggest that no second order correlation in the error terms is present and that the instruments are valid. In sum, our estimation results clearly show that the suboptimal system GMM estimator performs very well when compared with the first differences GMM or the three other system GMM estimators.

7 Conclusion

The weak instruments problem may cause substantial small-sample biases when using the first-difference GMM procedure to estimate autoregressive models for moderately persistent series from short panels. (Also see Blundell and Bond, 1998). However, these biases could be reduced by incorporating more informative moment conditions that are valid under quite general stationarity restrictions on the initial conditions. To this end, the system GMM estimation using lagged first differences as instruments for equations in addition to the usual lagged levels as instruments for the first differences equations is suggested as an alternative in Blundell and Bond (1999).

To go one step further, we considered a suboptimal system GMM estimation in the analysis of dynamic panel data sets with large cross-sectional variance. Since the small-sample properties of the first-difference GMM estimators depend on the initial weighting matrix, the performance of various system estimators with different weight matrices was investigated. Our Monte Carlo results indicate that the conventional system estimators are vulnerable to an increase in ρ . One of the most distinguishing features in these experiments was that biases and standard deviations increase with ρ in most cases. To overcome this deficiency, by inducing the variance of individual effects, μ_i , into the weight matrix, the SYS_{sub} estimation successfully weakens the increase of its biases and variances. Consequently, we expect that the SYS_{sub} estimation will provide useful parameter estimates for the practitioner.

In the estimation of the Cobb-Douglas production function for the 1,002 Japanese manufacturing firms, the suggested estimator provides the best parameter estimates in terms of precision.

Acknowledgement

The first author would like to thank Professor Katsuto Tanaka, Professor Taku Yamamoto and Doctor Kyongwon Kim for suggestions and valuable comments. We gratefully acknowledge the Grant-in-Aid from New Energy and Industrial Technology Development Organization (grant no. 0624006)

Appendix

Construction of variables

The source of our data on Japanese manufacturing firms for the empirical application is the DBJ database compiled by the Development Bank of Japan

Output: Firms' total sales are used as a proxy for gross output. Total sales are deflated by output deflators obtained from the SNA (System of National Accounts).

Intermediate inputs: Intermediate inputs are defined as (Cost of sales + Operating costs) - (Wages + Depreciation costs) and are provided in the SNA.

Labor input: As labor input, we used the average of man hours between year t and year $t-1$. Man hours are computed as each firms' total number of workers multiplied by the sectoral working hours obtained from the JIP.

The JIP 2006 Database was compiled as part of a RIETI research project. The detailed results of this project are reported in Fukao et al. (2006). The database contains annual information on 108 sectors, including 56 non-manufacturing sectors, from 1970 to 2002. These sectors cover the whole Japanese economy. The database includes detailed information on factor inputs, annual nominal and real input-output tables, as well as some additional statistics, such as capacity utilization rates, Japan's international trade by trade partner, inward and outward FDI, etc., at the detailed sectoral level. An Excel file version of the JIP2006 Database is available on RIETI's web site.

References

- [1] Ahn, S.C and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, *Journal of Econometrics* 68, 29-52.
- [2] Amemiya, T., 1976, The maximum likelihood, the minimum chi-square and the non-linear weighted least squares estimator in the general qualitative response model, *Journal of the American Statistical Association* 71, 347-351.
- [3] Alonso-Borrego, C. and R. Sanchez-Mangas, 2001, GMM Estimation of a Production Function with Panel Data: An Application to Spanish Manufacturing Firms, *Statistics and Econometrics Working Papers ws015527*, Universidad Carlos III.
- [4] Anderson, T.W. and C. Hsiao, 1981, Estimation of dynamic models with error components, *Journal of the American Statistical Association* 76, 598-606.
- [5] Anderson, T.W. and C. Hsiao, 1982, Formulation and estimation of dynamic models using panel data, *Journal of Econometrics* 18, 47-82.
- [6] Arellano, M. and S. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies* 58, 277-297.
- [7] Arellano, M. and O. Bover, 1995, Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics* 68, 29-52.
- [8] Balestra, P. and M. Nerlove, 1966, Pooling cross-section and time series data in the estimation of a dynamic model: the demand for natural gas, *Econometrica* 34, 585-612.
- [9] Baltagi, B.H. and Q. Li, 1995, Testing AR(1) against MA(1) disturbances in an error component model, *Journal of Econometrics* 48, 385-393.

- [10] Bhargava, A., L. Franzini and W. Narendranathan, 1982, Serial Correlation and Fixed effects model, *Review of Economic Studies* 49, 533-549.
- [11] Bhargava, A. and J.D. Sargan, 1983, Estimating dynamic random-effects models from panel data covering short time periods, *Econometrica* 51, 1635-1659.
- [12] Black, S. E. and L. M. Lynch, 2001, How to Compete: The Impact of Workplace Practices and Information Technology on Productivity, *Review of Economics and Statistics* 83, 435-445.
- [13] Blundell, R. and R.J. Smith, 1991, Conditions Initiales et Estimation Efficace dans les Modeles Dynamiques sur Donnees de Panel: une Application au Comportement d'Investissement des Entreprises, *Annales d'Economie et de Statistique* 20/21, 109-123.
- [14] Blundell, R. and S. Bond, 1998, Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* 87, 115-143.
- [15] Bowsher, G. 2002, On testing overidentifying restrictions in dynamic panel data models, *Economics Letters* 77, 211-220.
- [16] Fukao, K., S. Hamagata, T. Inui, K. Ito, H.U. Kwon, T. Makino, T. Miyagawa, Y. Nakanishi, and J. Tokui, 2006, Estimation Procedures and TFP Analysis of the JIP Database 2006 Provisional Version, paper presented at the EU KLEMS 3rd Consortium Meeting, May 7-9, 2006, Valencia (Spain).
- [17] Harris, M.N. and Laszlo Matyas, 2001, The Robustness of Estimators for Dynamic Panel Data Models to Misspecification, (Monash University, Dept. of Econometrics), Working Paper no. 9/96.
- [18] Griliches Z. and Mairesse, 1995, Production Functions: The Search for Identification, NBER Working Paper No.8075.
- [19] Hausman, J.A. and W. E. Taylor, 1981, Panel Data and Unobservable Individual Effects, *Econometrica* 49, 1377-1398.
- [20] Hayakawa, K, 2005, Small Sample Bias Properties of the System GMM Estimators in Dynamic Panel Data Models, Hi-Stat Discussion Paper Series No.82, Hitotsubashi University, Tokyo.
- [21] Hsiao, C., 1986, *Analysis of Panel Data* (Cambridge University Press, Cambridge).
- [22] Keane, M.P. and D.E. Runkle, 1992, On the estimation of panel data models with serial correlation when instruments are not strictly exogenous, *Journal of Business and Economic Statistics* 10, 1-9.
- [23] Kiviet, J.F, 1995, On bias, inconsistency and efficiency in various estimators of dynamic panel data models, *Journal of Econometrics* 68, 53-78.
- [24] Liu, S. and H. Neudecker, 1997, Experiments with Mixtures: Optimal Allocations for Becker's Models, *Metrika* 45, 53-66.
- [25] Nerlove, M., 1971a, Further evidence on the estimation of dynamic economic relations from a time series of cross-sections, *Econometrica* 39, 359-382.
- [26] Newey, W. K. and R.J. Smith, 2004, Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators, *Econometrica* 72, 219-255.

- [27] Nickell, S., 1981, Biases in dynamic models with fixed effects , *Econometrica* 49, 1399-1416.
- [28] Wansbeek, T.J. and P. Bekker, 1996, On IV, GMM and ML in a dynamic panel data model, *Economics Letters* 51, 145-152.
- [29] Windmeijer, F., 2004, A finite sample correction for the variance of linear efficient two-step GMM Estimator, *Journal of Econometrics*, Forthcoming.