*Journal of Economic Theory and Econometrics*, Vol. 26, No. 2, Jun. 2015, 57–78

# An Alternative System GMM Estimation in Dynamic Panel Models*

Hosung Jung† Hyeog Ug Kwon‡ Gyehyung Jeon§

**Abstract** The system GMM estimator in dynamic panel data models which combines two sets of 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 asymptotic efficiency. Our Monte Carlo experiments show that the small sample properties of the suboptimal system estimator are much more reliable than any other conventional system GMM estimator in terms of bias.

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

**JEL Classification** C23

*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). We also thank SERI(Samsung Economic Research Institute) for financial support.

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

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

§Department of Economics, Seoul National University, Korea, Doctoral student, e-mail: peterjeon82@gmail.com

*Received Jan 05, 2015, Revised June 22, 2015, Accepted June 29, 2015*

## 1. INTRODUCTION

It is generally known that increasing the set of linearly independent instruments will always improve the asymptotic efficiency of various IV and GMM estimators (Arellano and Bover, 1995; Anderson and Hsiao, 1981, 1982 ; 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, 2005). 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 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 ($\text{SYS}_{sub}$ hereafter) estimation.[2]

To investigate the magnitude of the efficiency gain, Kantorovich inequality (KI hereafter) upper bounds based on the KI (Windmeijer, 2000, pp.178-180) 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 reduce bias from using the $\text{SYS}_{sub}$ 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 $\text{SYS}_{sub}$ estimator is relatively

[1] See Blundell and Bond (1998) for details.

[2] As 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 $\text{SYS}_{sub}$ does not suffer from a downward bias of its estimated standard error.

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 3 and 4, we propose the $\mathrm{SYS}_{sub}$ estimation and consider the efficiency gain against using the identity matrix as an initial weight matrix. Section 5 reports the simulation results, while Section 6 present an empirical application to a production function. Section 7 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, $u_{it}$:

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

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

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

and

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

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

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

We define the variance ratio, $\rho = \frac{\sigma_\mu^2}{\sigma_v^2}$, 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 $i$,

$$\begin{aligned} E\left[y_{i,t-j}\Delta u_{it}\right] &= 0 \quad \text{for} \quad (j=2,\ldots,t-1; t=3,\ldots,T) && (5)\\ E\left[\Delta y_{i,t-1}u_{it}\right] &= 0 \quad \text{for} \quad (t=3,\ldots,T). && (6)\end{aligned}$$

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

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

where

$$q_i = \begin{bmatrix} \Delta u_i \\ u_i \end{bmatrix}. \tag{8}$$

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

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

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

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

and

$$Z_{li} = \operatorname{diag}\left[\Delta y_{i2}, \Delta y_{i3}, \ldots, \Delta y_{i,T-1}\right]. \tag{11}$$

Let $Z_s = (Z'_{s,1},\ldots,Z'_{s,N})$ be a $m_s \times 2N(T-2)$ matrix and $Y$ be a stacked matrix $Y_i = (\Delta y_i, y_i)'$, where $\Delta y_i = (\Delta y_{i3}, \Delta y_{i4}, ..., \Delta y_{iT})$ and $y_i = (y_{i3}, y_{i4}, ..., y_{iT})$. Also let $Y_{-1}$ be a stacked matrix $Y_{i,-1} = (\Delta y_{i,-1}, y_{i,-1})'$, where $\Delta y_{i,-1} = (\Delta y_{i2}, \Delta y_{i3}, ...,$ $\Delta y_{i,T-1})$ and $y_{i,-1} = (y_{i2}, y_{i3}, ..., y_{i,T-1})$. Then, the one-step system GMM estimator based on these moment conditions (7) is

$$\hat{\alpha}_s = \left[Y'_{-1}Z_sW_NZ'_sY_{-1}\right]^{-1}\left[Y'_{-1}Z_sW_NZ'_sY\right] \tag{12}$$

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

Windmeijer (2005, p.25) reported that in panel data models, the estimated asymptotic standard errors of the efficient two-step GMM estimator can be substantially biased downward in small samples; 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, ignoring the variance-covariance structure of residuals, one can use

$$W_N = \left[\frac{1}{N}\sum\nolimits_{i=1}^{N} Z'_{si} H_I Z_{si}\right]^{-1}, \tag{13}$$

where

$$H_I = \begin{bmatrix} H_d & \mathbf{0} \\ \mathbf{0} & I_{T-2} \end{bmatrix}. \tag{14}$$

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 $H_I$ works well for small values of $\sigma_\mu^2$, there exists the potential for efficiency loss when $\sigma_\mu^2$ gets large.

## 3. A SUBOPTIMAL WEIGHT MATRIX

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

$$H_s = \begin{bmatrix} H_d & C \\ C' & I_{T-2} \end{bmatrix} \tag{15}$$

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

[3]Also see Windmeijer (2000, pp.180-182) for details.

identity matrix $I_{T-2}$ in $H_s$ should be replaced by the matrix $J_{T-2}$ to achieve optimality, where

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

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

$$H_{so} = \begin{bmatrix} H_d & C \\ C' & J_{T-2} \end{bmatrix}. \tag{17}$$

Although the system estimator in Blundell and Bond (1998) performs well as long as $\rho$ is reasonably small, there are always cases where the variance of the individual effects, $\mu_i$, 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 $\rho$. 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, $\rho$, is unknown in practice, we suggest using the estimates of the optimal weight matrix $H_{so}$, saying

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

where a natural estimator for the variance ratio $\hat{\rho}$ is readily available in the initial step of the system estimation.[4] To obtain $\hat{\rho}$, we derive $\sigma_v^2$ from the first-difference GMM estimation,

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

while $\sigma_\mu^2$ is given by

$$\hat{\sigma}_\mu^2 = \frac{\sum_{i=1}^{N} \left[\tilde{u}_i' \tilde{u}_i - \Delta\hat{u}_i' \Delta\hat{u}_i / 2\right]}{N(T-2)}, \tag{20}$$

[4]We drop C in (18), because we assume that the efficiency change caused by using 0 matrix instead of C would not be so big. In order to confirm this, we additionally compare the small-sample properties of possible weight matrices in Appendix B. We find that the estimated suboptimal weight matrix with C performs better when both $\alpha$ and $\rho$ are small, but the weight mtrix without C performs better when either $\alpha$ or $\rho$ is large.

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 $\rho$ becomes large.[5]

## 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 $\alpha_0$ minimizes

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

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

$$\frac{1}{\sqrt{N}} f_i(\alpha_0) \to \mathrm{N}(0,\Psi), \tag{22}$$

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,

$$\sqrt{N}(\hat{\alpha}_s - \alpha_0) \to \mathrm{N}(0, V_W), \tag{23}$$

where $V_W = (F_0'WF_0)^{-1}F_0'W\Psi WF_0(F_0'WF_0)^{-1}$. An optimal choice for $W$ is $\Psi^{-1}$, so the asymptotic variance matrix is given by $(F_0'WF_0)^{-1}$. Clearly, the following inequality holds for any positive matrix $W$:

$$(F_0'\Psi^{-1}F_0)^{-1} \le (F_0'WF_0)^{-1}F_0'W\Psi WF_0(F_0'WF_0)^{-1} \tag{24}$$

According to Liu and Neudecker (1996), the following inequality also holds:

$$(F_0'WF_0)^{-1}F_0'W\Psi WF_0(F_0'WF_0)^{-1} \le \frac{(\lambda_1+\lambda_p)^2}{4\lambda_1\lambda_p}(F_0'\Psi^{-1}F_0)^{-1}, \tag{25}$$

and the KI upper bounds – $\mathrm{KI}_{ub} = \frac{(\lambda_1+\lambda_p)^2}{4\lambda_1\lambda_p}$ – are calculated, where $\lambda_i > 0\,(i = 1,\ldots,p)$ are the eigenvalues of the $p\times p$ matrix $\Psi W$.[6] If we use an initial weight matrix equal to $\hat{H}_{so}$, $\Psi$ is obtained by

$$\Psi = plim\frac{1}{N}\sum_{i=1}^{N}\left[Z_{si}'\hat{H}_{so}Z_{si}\right], \tag{26}$$

[5] If $\rho$ is small, the potential efficiency gain gets smaller

[6] Also see Liu and Neudecker (1997) for details

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 $\Psi$ and $W_1$ are given by

$$
\begin{aligned}
\Psi &= plim\frac{1}{N}\sum_{i=1}^{N}\left[Z_{si}^{'}\hat{H}_{so}Z_{si}\right] \\
&\rightarrow \begin{bmatrix}
2\sigma_y^2 & -\sigma_y^2 & -\delta & 0 & 0 \\
-\sigma_y^2 & 2\sigma_y^2 & 2\delta & 0 & 0 \\
-\delta & 2\delta & 2\sigma_y^2 & 0 & 0 \\
0 & 0 & 0 & \frac{2(1+\rho)\sigma_v^2}{1+\alpha} & \frac{-\rho(1-\alpha)\sigma_v^2}{1+\alpha} \\
0 & 0 & 0 & \frac{-\rho(1-\alpha)\sigma_v^2}{1+\alpha} & \frac{2(1+\rho)\sigma_v^2}{1+\alpha}
\end{bmatrix}
\end{aligned}
\tag{27}
$$

and

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

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

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

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 $\delta$ and $\sigma_y^2$. The eigenvalues of the lower block matrix are functions of $\rho$ and $\alpha$:

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

If $\rho \leq 1.\dot{3}$, 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 $\Psi W_1$ 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).

$$y_{it} = \alpha y_{i,t-1} + u_{it}, \tag{31}$$

for $i = 1,2,\ldots,N$ and $t = 2,3,\ldots,T$. 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, $v_{it} \sim \mathrm{NID}(0,1)$. We generate the initial conditions $y_{i1}$ as

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

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

The variance ratio, $\rho$, is characterized by $\frac{\sigma_\mu^2}{\sigma_v^2}$, so it depends only on $\sigma_\mu^2$. 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 $\mathrm{SYS}_1$ and $\mathrm{SYS}_2$, respectively. Furthermore, $\mathrm{SYS}_3$ uses the alternative suboptimal weighting matrix defined in (15). $\mathrm{SYS}_{sub}$ denotes the newly proposed suboptimal weight matrix (18), which uses the estimated $\rho$, while $\mathrm{TSYS}_{sub}$ uses the true $\rho$.[7]

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 $\rho$. While the biases of ISYS, $\mathrm{SYS}_1$, $\mathrm{SYS}_2$ and $\mathrm{SYS}_3$ are

[7] In one of the most widespread statistics programs, STATA, ISYS, $\mathrm{SYS}_1$ and $\mathrm{SYS}_3$ 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)

| | | $\rho$ | ISYS | $SYS_1$ | $SYS_2$ | $SYS_3$ | $SYS_{sub}$ | $TSYS_{sub}$ |
|---|---|---|---|---|---|---|---|---|
| $\alpha = 0.2$ | Mean | 0 | 0.1865 | 0.1939 | 0.1962 | 0.1956 | 0.1937 | 0.1939 |
| | | 0.5 | 0.1879 | 0.1990 | 0.2026 | 0.2132 | 0.1952 | 0.1955 |
| | | 1 | 0.1979 | 0.2113 | 0.2113 | 0.2308 | 0.2028 | 0.2027 |
| | | 2 | 0.2075 | 0.2232 | 0.2154 | 0.2544 | 0.2044 | 0.2028 |
| | | 5 | 0.2436 | 0.2638 | 0.2382 | 0.3210 | 0.2176 | 0.2083 |
| | | 10 | 0.3050 | 0.3298 | 0.2870 | 0.4197 | 0.2458 | 0.2196 |
| | Std. | 0 | 0.0858 | 0.0753 | 0.0752 | 0.0687 | 0.0754 | 0.0753 |
| | | 0.5 | 0.0951 | 0.0858 | 0.0817 | 0.0825 | 0.0865 | 0.0857 |
| | | 1 | 0.1004 | 0.0925 | 0.0850 | 0.0941 | 0.0930 | 0.0917 |
| | | 2 | 0.1109 | 0.1058 | 0.0936 | 0.1156 | 0.1054 | 0.1030 |
| | | 5 | 0.1260 | 0.1283 | 0.1076 | 0.1534 | 0.1203 | 0.1150 |
| | | 10 | 0.1489 | 0.1553 | 0.1397 | 0.1859 | 0.1389 | 0.1246 |
| $\alpha = 0.5$ | Mean | 0 | 0.4733 | 0.4872 | 0.4914 | 0.4908 | 0.4867 | 0.4872 |
| | | 0.5 | 0.4845 | 0.5031 | 0.5050 | 0.5165 | 0.4958 | 0.4966 |
| | | 1 | 0.4959 | 0.5161 | 0.5147 | 0.5365 | 0.5014 | 0.5006 |
| | | 2 | 0.5205 | 0.5419 | 0.5343 | 0.5725 | 0.5167 | 0.5097 |
| | | 5 | 0.5872 | 0.6092 | 0.5878 | 0.6615 | 0.5597 | 0.5296 |
| | | 10 | 0.6683 | 0.6884 | 0.6634 | 0.7520 | 0.6266 | 0.5530 |
| | Std. | 0 | 0.0975 | 0.0855 | 0.0821 | 0.0747 | 0.0857 | 0.0855 |
| | | 0.5 | 0.1117 | 0.1024 | 0.0962 | 0.0945 | 0.1052 | 0.1030 |
| | | 1 | 0.1179 | 0.1110 | 0.1021 | 0.1066 | 0.1161 | 0.1119 |
| | | 2 | 0.1302 | 0.1251 | 0.1133 | 0.1240 | 0.1306 | 0.1243 |
| | | 5 | 0.1446 | 0.1418 | 0.1362 | 0.1479 | 0.1549 | 0.1409 |
| | | 10 | 0.1533 | 0.1535 | 0.1554 | 0.1610 | 0.1763 | 0.1494 |
| $\alpha = 0.8$ | Mean | 0 | 0.7511 | 0.7769 | 0.7859 | 0.7857 | 0.7754 | 0.7769 |
| | | 0.5 | 0.7885 | 0.8139 | 0.8063 | 0.8324 | 0.8016 | 0.8026 |
| | | 1 | 0.8215 | 0.8437 | 0.8308 | 0.8670 | 0.8268 | 0.8211 |
| | | 2 | 0.8688 | 0.8863 | 0.8709 | 0.9082 | 0.8674 | 0.8483 |
| | | 5 | 0.9291 | 0.9404 | 0.9281 | 0.9528 | 0.9230 | 0.8833 |
| | | 10 | 0.9600 | 0.9670 | 0.9589 | 0.9765 | 0.9527 | 0.9032 |
| | Std. | 0 | 0.1063 | 0.0955 | 0.0889 | 0.0775 | 0.0966 | 0.0955 |
| | | 0.5 | 0.1224 | 0.1133 | 0.1112 | 0.0943 | 0.1260 | 0.1155 |
| | | 1 | 0.1184 | 0.1113 | 0.1149 | 0.0959 | 0.1296 | 0.1150 |
| | | 2 | 0.1155 | 0.1086 | 0.1173 | 0.0939 | 0.1347 | 0.1161 |
| | | 5 | 0.0930 | 0.0893 | 0.1098 | 0.0781 | 0.1151 | 0.1057 |
| | | 10 | 0.0772 | 0.0756 | 0.0949 | 0.0650 | 0.0948 | 0.0975 |

*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)

| | | $\rho$ | ISYS | $SYS_1$ | $SYS_2$ | $SYS_3$ | $SYS_{sub}$ | $TSYS_{sub}$ |
|---|---|---|---|---|---|---|---|---|
| $\alpha = 0.2$ | Mean | 0 | 0.1604 | 0.1912 | 0.1943 | 0.1974 | 0.1912 | 0.1912 |
| | | 0.5 | 0.1619 | 0.1974 | 0.1991 | 0.2348 | 0.1945 | 0.1945 |
| | | 1 | 0.1682 | 0.2039 | 0.2032 | 0.2659 | 0.1962 | 0.196 |
| | | 2 | 0.1806 | 0.2181 | 0.2126 | 0.3196 | 0.1994 | 0.1987 |
| | | 5 | 0.2163 | 0.2544 | 0.2385 | 0.4376 | 0.2031 | 0.1999 |
| | | 10 | 0.2714 | 0.3095 | 0.2853 | 0.5634 | 0.2097 | 0.2008 |
| | Std. | 0 | 0.0535 | 0.0429 | 0.0446 | 0.0383 | 0.0429 | 0.0429 |
| | | 0.5 | 0.0562 | 0.0470 | 0.0471 | 0.0472 | 0.0470 | 0.0467 |
| | | 1 | 0.0573 | 0.0482 | 0.0465 | 0.0575 | 0.0482 | 0.0477 |
| | | 2 | 0.0598 | 0.0517 | 0.0488 | 0.0750 | 0.0505 | 0.0495 |
| | | 5 | 0.0668 | 0.0607 | 0.0557 | 0.1020 | 0.0538 | 0.0516 |
| | | 10 | 0.0800 | 0.0758 | 0.0727 | 0.1141 | 0.0575 | 0.0530 |
| $\alpha = 0.5$ | Mean | 0 | 0.4403 | 0.4884 | 0.4929 | 0.4974 | 0.4884 | 0.4884 |
| | | 0.5 | 0.4451 | 0.4972 | 0.4995 | 0.5476 | 0.4913 | 0.4914 |
| | | 1 | 0.4591 | 0.5104 | 0.5089 | 0.5855 | 0.4959 | 0.4951 |
| | | 2 | 0.4918 | 0.5373 | 0.5283 | 0.6466 | 0.5047 | 0.5005 |
| | | 5 | 0.5630 | 0.5981 | 0.5806 | 0.7500 | 0.5245 | 0.5058 |
| | | 10 | 0.6475 | 0.6707 | 0.6522 | 0.8325 | 0.5602 | 0.5123 |
| | Std. | 0 | 0.0591 | 0.0437 | 0.0442 | 0.0371 | 0.0437 | 0.0437 |
| | | 0.5 | 0.0648 | 0.0519 | 0.0499 | 0.0503 | 0.0528 | 0.0517 |
| | | 1 | 0.0674 | 0.0554 | 0.0521 | 0.0600 | 0.0567 | 0.0546 |
| | | 2 | 0.0698 | 0.0595 | 0.0559 | 0.0707 | 0.0606 | 0.0565 |
| | | 5 | 0.0820 | 0.0703 | 0.0697 | 0.0771 | 0.0699 | 0.0596 |
| | | 10 | 0.0889 | 0.0791 | 0.0833 | 0.0743 | 0.0858 | 0.0624 |
| $\alpha = 0.8$ | Mean | 0 | 0.7010 | 0.7759 | 0.7840 | 0.7941 | 0.7756 | 0.7759 |
| | | 0.5 | 0.7529 | 0.8094 | 0.8090 | 0.8636 | 0.7959 | 0.7945 |
| | | 1 | 0.7966 | 0.8408 | 0.8359 | 0.8985 | 0.8204 | 0.8099 |
| | | 2 | 0.8488 | 0.8790 | 0.8735 | 0.9324 | 0.8552 | 0.8258 |
| | | 5 | 0.9145 | 0.9302 | 0.9264 | 0.9666 | 0.9121 | 0.8471 |
| | | 10 | 0.9496 | 0.9581 | 0.9558 | 0.9815 | 0.9471 | 0.8591 |
| | Std. | 0 | 0.0661 | 0.0475 | 0.0453 | 0.0348 | 0.0477 | 0.0475 |
| | | 0.5 | 0.0704 | 0.0566 | 0.0552 | 0.0411 | 0.0634 | 0.0569 |
| | | 1 | 0.0707 | 0.0579 | 0.0583 | 0.0400 | 0.0696 | 0.0592 |
| | | 2 | 0.0633 | 0.0529 | 0.0552 | 0.0366 | 0.0706 | 0.0578 |
| | | 5 | 0.0483 | 0.0426 | 0.0457 | 0.0279 | 0.0632 | 0.0565 |
| | | 10 | 0.0355 | 0.0322 | 0.0351 | 0.0213 | 0.0498 | 0.0544 |

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

Table 3: Estimation Results of $\rho$

| $\alpha$ | $\rho = 0$ | 0.5 | 1 | 2 | 5 | 10 |
|---|---|---|---|---|---|---|
| 0.2 | 0.0132 | 0.5217 | 0.9961 | 1.9010 | 4.2034 | 6.7902 |
| 0.5 | 0.0235 | 0.5393 | 0.9990 | 1.7962 | 3.2832 | 4.4200 |
| 0.8 | 0.0529 | 0.6559 | 0.9667 | 1.2758 | 1.3355 | 1.5223 |

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

negligible when $\rho \leq 1$, they rapidly increase with $\rho$. The biases of $\text{SYS}_{sub}$ and $\text{TSYS}_{sub}$ show a much slower increase due to an increase in $\rho$. Even in the case of $\alpha = 0.8$, the two estimators show the smallest increase in mean. Consequently, we conclude that $\text{SYS}_{sub}$ outperforms the conventional system estimators in terms of bias. However, the advantage of the $\text{SYS}_{sub}$ estimator decreases as $\alpha$ grows to unity because a high $\alpha$ leads to an unreliable estimate of $\rho$ itself. Especially, efficiency loss gets larger as $\alpha$ grows to unity.

Table 3 presents the estimation results of $\rho$ based on residuals from the first-step system GMM estimator. The mean of the estimated $\rho$ has substantial bias when $\alpha$ is close to one, which yields no considerable improvement to using the suboptimal system procedure. Even though the suggested estimator, $\text{SYS}_{sub}$ depends on the results from the first-step estimation of $\rho$, $\text{SYS}_{sub}$ performs better in terms of bias than any of the other conventional system estimators.

## 6. EMPIRICAL APPLICATION: ESTIMATION OF PRODUCTION FUNCTIONS USING JAPANESE FIRM-LEVEL PANEL DATA

We apply the suboptimal system GMM estimation procedure(denoted $\text{SYS}_{sub}$ ) 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

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

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

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 $\gamma_t$ is a year-specific intercept reflecting, for example, a common technology shock. As for the error components, $\mu_i$ 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:

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

or

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

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 $w_{it} \sim MA(1)$ if there are measurement errors in some of the series. Although consistent estimates of the unrestricted parameters, $\pi = (\pi_1, \ldots, \pi_7)$, 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 $\pi$ and $var(\pi)$.

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 (Griliches and Mairesse, 1995; Nickell, 1981).

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

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. | | | $m_1$- | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | | | $m_2$- | 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, $SYS_1$ and $SYS_3$ 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 $t-1$.

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 levels (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 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 $\rho$. One of the most distinguishing features in these experiments was that biases and standard deviations increase with $\rho$ in most cases. To overcome this deficiency, by inducing the variance of individual effects, $\mu_i$, into the weight matrix, the $\text{SYS}_{sub}$ estimation successfully weakens the increase of its biases. But the estimator depends on first-step estimation of $\rho$ that if $\alpha$ grows to unity, efficiency loss gets larger. Consequently, we expect that the $\text{SYS}_{sub}$ estimation will provide useful parameter estimates for the practitioner as a sensitivity check to conventional estimators.

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.

There is another approach to improve the performance of the dynamic panel data model. This is limiting the lag depth, or transforming the instrument matrix. For example, Mehrhoff (2009) suggests a solution to the problem of too many instruments in dynamic panel data GMM. Our estimation technique of limiting the lag depth or transforming the instrument matrix could perform better and we leave it to future research to examine it.

## REFERENCES

Ahn, S.C and P. Schmidt, 1995, Efficient estimation of models for dynamic panel data, *Journal of Econometrics* 68, 29-52.

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.

Anderson, T.W. and C. Hsiao, 1981, Estimation of dynamic models with error components, *Journal of the American Statistical Association* 76, 598-606.

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

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.

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

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.

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

Griliches Z. and Mairesse, 1995, Production Functions: The Search for Identification, NBER Working Paper No.8075.

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).

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.

Kiviet, J. F., 2007, Judging contending estimators by simulation: tournaments in dynamic panel data models, *In G. D. A. Philips and E. Tzavalis (Eds.) The Refinement of Econometric Estimation and Test Procedures* 282-318, Cambridge: Cambridge University Press.

Liu, S. and H. Neudecker, 1996, Several Matrix Knatorovich-Type Inequalities, *Journal of Mathematical Analysis and Applications* 197, 23-26.

Mehrhoff, J., 2009, A solution to the problem of too many instruments in dynamic panel data GMM, *Discussion Paper Series* No. 1, Economic Studies, Deutsche Bundesbank.

Nerlove, M., 1971, Further evidence on the estimation of dynamic economic relations from a time series of cross-sections, *Econometrica* 39, 359-382.

Newey, W. K. and R.J. Smith, 2004, Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators, *Econometrica* 72, 219-255.

Nickell, S., 1981, Biases in dynamic models with fixed effects , *Econometrica* 49, 1399-1416.

Windmeijer, F., 2000, Efficiency Comparisons for a System GMM Estimator in Dynamic Panel Data Models, *Innovations in Multivariate Statistical Analysis*, Advanced Studies in Theoretical and Applied Econometrics Vol.36, 175-184.

Windmeijer, F., 2005, A finite sample correction for the variance of linear efficient two-step GMM Estimator, *Journal of Econometrics*, 126, 25-51.

# APPENDIX

## A. 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.

## B. SMALL-SAMPLE PROPERTIES OF ALTERNATIVE WEIGHTING MATRICES

When choosing the weight matrix in the substitution for the matrix in (17) there are various options. In this section, with a different point of view and insight, we illustrate the small-sample performances using the variety of suboptimal weighting matrices.

The suboptimal matrices are represented in (A1) to (A6). $H_1$ uses the estimate for the variance ratio $\hat{\rho}$, instead of the unknown parameter $\rho$ in the matrix (17). The estimator using $H_1$ is named as $SYS_{full}$. $SYS_3$, defined in Section 5, uses $H_2$. $H_3$ is defined in (18), which we suggest in this paper, and the estimator using it is named as $SYS_{sub}$ in Section 5. $H_4$ is defined in (14), and the one-step estimation in Blundell and Bond (1998) uses this method. The one-step estimator is named as $SYS_1$ in Section 5. ISYS, defined in Section 5, uses $H_5$. Kiviet (2007) suggestes $H_6$ as an alternative suboptimal weighting matrix, which uses a positive value instead of the unknown parameter $\rho$ in the matrix (17). We use 1, 5, 10 as positive values in the matrix $H_6$. The estimators are named as $SYS_K^1$, $SYS_K^5$, $SYS_K^{10}$, respectively.

$$H_1 = \begin{bmatrix} H_d & C \\ C' & \hat{J}_{T-2} \end{bmatrix} \tag{A,1}$$

$$H_2 = \begin{bmatrix} H_d & C \\ C' & I_{T-2} \end{bmatrix} \tag{A,2}$$

$$H_3 = \begin{bmatrix} H_d & \mathbf{0} \\ \mathbf{0} & \hat{J}_{T-2} \end{bmatrix} \tag{A,3}$$

$$H_4 = \begin{bmatrix} H_d & \mathbf{0} \\ \mathbf{0} & \hat{J}_{T-2} \end{bmatrix} \tag{A,4}$$

$$H_5 = \begin{bmatrix} I_{T-2} & \mathbf{0} \\ \mathbf{0} & I_{T-2} \end{bmatrix} \tag{A,5}$$

$$H_6 = \begin{bmatrix} H_d & C \\ C' & \tilde{J}_{T-2} \end{bmatrix} \tag{A,6}$$

Tables A.1 and A.2 present the estimation results for T=5 and 10, respectively. When the variance ratio $\rho$ is low, the sizes of the bias of the estimators are negligible except that of $\text{SYS}_K^1$. But the sizes differ as $\rho$ increases. The sizes of the bias of $\text{SYS}_{sub}$, $\text{SYS}_K^5$, and $\text{SYS}_K^{10}$ show a much slower increase following an increase in $\rho$. In most cases, the standard deviations of $\text{SYS}_{full}$, $\text{SYS}_3$, and $\text{SYS}_1$ are smaller than that of $\text{SYS}_{sub}$ but the differences are negligible. In terms of bias, $\text{SYS}_{sub}$ performs better than $\text{SYS}_{full}$, $\text{SYS}_3$, and $\text{SYS}_1$. The sizes of bias and variances of $\text{SYS}_K^5$ and $\text{SYS}_K^{10}$, especially those of $\text{SYS}_K^{10}$, are smaller than $\text{SYS}_{sub}$ when the variance ratio $\rho$ is high, and larger when $\rho$ is low. Kiviet's suggestion with large positive value can be a good estimation when $\rho$ is very high. However, with the absence of the knowledge of the exact value of $\rho$, $\text{SYS}_{sub}$ performs more stable than $\text{SYS}_K^5$ and $\text{SYS}_K^{10}$.

Table A.1: Small-Sample Properties of Alternative Weighting Matrices (T=5)

| | | $\rho$ | $\text{SYS}_{full}$ | $\text{SYS}_3$ | $\text{SYS}_{sub}$ | $\text{SYS}_1$ | ISYS | $\text{SYS}_K^1$ | $\text{SYS}_K^5$ | $\text{SYS}_K^{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha = 0.2$ | Mean | 0 | 0.193 | 0.196 | 0.194 | 0.194 | 0.187 | -0.041 | 0.189 | 0.188 |
| | | 0.5 | 0.201 | 0.213 | 0.195 | 0.199 | 0.188 | 0.127 | 0.188 | 0.185 |
| | | 1 | 0.210 | 0.231 | 0.203 | 0.211 | 0.198 | 0.228 | 0.196 | 0.190 |
| | | 2 | 0.215 | 0.254 | 0.204 | 0.223 | 0.208 | 0.327 | 0.199 | 0.190 |
| | | 5 | 0.236 | 0.321 | 0.218 | 0.264 | 0.244 | 0.468 | 0.222 | 0.203 |
| | | 10 | 0.273 | 0.420 | 0.246 | 0.330 | 0.305 | 0.572 | 0.264 | 0.232 |
| | Std. | 0 | 0.069 | 0.069 | 0.075 | 0.075 | 0.086 | 0.666 | 0.078 | 0.079 |
| | | 0.5 | 0.080 | 0.083 | 0.087 | 0.086 | 0.095 | 0.662 | 0.089 | 0.093 |
| | | 1 | 0.087 | 0.094 | 0.093 | 0.093 | 0.100 | 0.651 | 0.093 | 0.099 |
| | | 2 | 0.102 | 0.116 | 0.105 | 0.106 | 0.111 | 0.666 | 0.102 | 0.108 |
| | | 5 | 0.123 | 0.153 | 0.120 | 0.128 | 0.126 | 0.598 | 0.111 | 0.116 |
| | | 10 | 0.148 | 0.186 | 0.139 | 0.155 | 0.149 | 0.549 | 0.123 | 0.122 |
| $\alpha = 0.5$ | Mean | 0 | 0.487 | 0.491 | 0.487 | 0.487 | 0.473 | 0.174 | 0.479 | 0.478 |
| | | 0.5 | 0.502 | 0.517 | 0.496 | 0.503 | 0.485 | 0.378 | 0.479 | 0.471 |
| | | 1 | 0.512 | 0.537 | 0.501 | 0.516 | 0.496 | 0.464 | 0.484 | 0.470 |
| | | 2 | 0.532 | 0.573 | 0.517 | 0.542 | 0.521 | 0.549 | 0.502 | 0.482 |
| | | 5 | 0.586 | 0.662 | 0.560 | 0.609 | 0.587 | 0.675 | 0.550 | 0.517 |
| | | 10 | 0.661 | 0.752 | 0.627 | 0.688 | 0.668 | 0.736 | 0.619 | 0.572 |
| | Std. | 0 | 0.076 | 0.075 | 0.086 | 0.086 | 0.098 | 0.689 | 0.092 | 0.095 |
| | | 0.5 | 0.095 | 0.095 | 0.105 | 0.102 | 0.112 | 0.641 | 0.111 | 0.122 |
| | | 1 | 0.106 | 0.107 | 0.116 | 0.111 | 0.118 | 0.600 | 0.117 | 0.130 |
| | | 2 | 0.124 | 0.124 | 0.131 | 0.125 | 0.130 | 0.583 | 0.122 | 0.136 |
| | | 5 | 0.154 | 0.148 | 0.155 | 0.142 | 0.145 | 0.507 | 0.131 | 0.142 |
| | | 10 | 0.182 | 0.161 | 0.176 | 0.154 | 0.153 | 0.449 | 0.137 | 0.144 |
| $\alpha = 0.8$ | Mean | 0 | 0.781 | 0.786 | 0.775 | 0.777 | 0.751 | 0.421 | 0.760 | 0.756 |
| | | 0.5 | 0.812 | 0.832 | 0.802 | 0.814 | 0.789 | 0.684 | 0.762 | 0.733 |
| | | 1 | 0.840 | 0.867 | 0.827 | 0.844 | 0.822 | 0.761 | 0.785 | 0.747 |
| | | 2 | 0.880 | 0.908 | 0.867 | 0.886 | 0.869 | 0.810 | 0.827 | 0.785 |
| | | 5 | 0.932 | 0.953 | 0.923 | 0.940 | 0.929 | 0.870 | 0.894 | 0.857 |
| | | 10 | 0.960 | 0.977 | 0.953 | 0.967 | 0.960 | 0.903 | 0.938 | 0.910 |
| | Std. | 0 | 0.081 | 0.078 | 0.097 | 0.096 | 0.106 | 0.637 | 0.114 | 0.125 |
| | | 0.5 | 0.111 | 0.094 | 0.126 | 0.113 | 0.122 | 0.492 | 0.125 | 0.148 |
| | | 1 | 0.118 | 0.096 | 0.130 | 0.111 | 0.118 | 0.442 | 0.120 | 0.141 |
| | | 2 | 0.126 | 0.094 | 0.135 | 0.109 | 0.116 | 0.402 | 0.115 | 0.136 |
| | | 5 | 0.109 | 0.078 | 0.115 | 0.089 | 0.093 | 0.326 | 0.094 | 0.114 |
| | | 10 | 0.090 | 0.065 | 0.095 | 0.076 | 0.077 | 0.312 | 0.074 | 0.089 |

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

Table A.2: Small-Sample Properties of Alternative Weighting Matrices (T=10)

| | | $\rho$ | $\text{SYS}_{full}$ | $\text{SYS}_3$ | $\text{SYS}_{sub}$ | $\text{SYS}_1$ | ISYS | $\text{SYS}_K^1$ | $\text{SYS}_K^5$ | $\text{SYS}_K^{10}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha = 0.2$ | Mean | 0 | 0.196 | 0.197 | 0.191 | 0.191 | 0.160 | 0.013 | 0.187 | 0.187 |
| | | 0.5 | 0.204 | 0.235 | 0.195 | 0.197 | 0.162 | 0.120 | 0.187 | 0.186 |
| | | 1 | 0.205 | 0.266 | 0.196 | 0.204 | 0.168 | 0.198 | 0.189 | 0.185 |
| | | 2 | 0.209 | 0.320 | 0.199 | 0.218 | 0.181 | 0.270 | 0.194 | 0.188 |
| | | 5 | 0.215 | 0.438 | 0.203 | 0.254 | 0.216 | 0.418 | 0.208 | 0.195 |
| | | 10 | 0.225 | 0.563 | 0.210 | 0.310 | 0.271 | 0.521 | 0.231 | 0.207 |
| | Std. | 0 | 0.039 | 0.038 | 0.043 | 0.043 | 0.054 | 0.357 | 0.042 | 0.042 |
| | | 0.5 | 0.043 | 0.047 | 0.047 | 0.047 | 0.056 | 0.354 | 0.045 | 0.047 |
| | | 1 | 0.045 | 0.058 | 0.048 | 0.048 | 0.057 | 0.367 | 0.046 | 0.047 |
| | | 2 | 0.049 | 0.075 | 0.051 | 0.052 | 0.060 | 0.387 | 0.047 | 0.049 |
| | | 5 | 0.055 | 0.102 | 0.054 | 0.061 | 0.067 | 0.385 | 0.050 | 0.050 |
| | | 10 | 0.061 | 0.114 | 0.058 | 0.076 | 0.080 | 0.383 | 0.053 | 0.051 |
| $\alpha = 0.5$ | Mean | 0 | 0.496 | 0.497 | 0.488 | 0.488 | 0.440 | 0.243 | 0.482 | 0.482 |
| | | 0.5 | 0.507 | 0.548 | 0.491 | 0.497 | 0.445 | 0.389 | 0.477 | 0.473 |
| | | 1 | 0.514 | 0.586 | 0.496 | 0.510 | 0.459 | 0.461 | 0.481 | 0.473 |
| | | 2 | 0.526 | 0.647 | 0.505 | 0.537 | 0.492 | 0.529 | 0.490 | 0.477 |
| | | 5 | 0.555 | 0.750 | 0.525 | 0.598 | 0.563 | 0.653 | 0.520 | 0.494 |
| | | 10 | 0.603 | 0.833 | 0.560 | 0.671 | 0.648 | 0.733 | 0.565 | 0.523 |
| | Std. | 0 | 0.038 | 0.037 | 0.044 | 0.044 | 0.059 | 0.345 | 0.043 | 0.043 |
| | | 0.5 | 0.047 | 0.050 | 0.053 | 0.052 | 0.065 | 0.342 | 0.051 | 0.054 |
| | | 1 | 0.053 | 0.060 | 0.057 | 0.055 | 0.067 | 0.346 | 0.053 | 0.056 |
| | | 2 | 0.062 | 0.071 | 0.061 | 0.060 | 0.070 | 0.337 | 0.054 | 0.057 |
| | | 5 | 0.078 | 0.077 | 0.070 | 0.070 | 0.082 | 0.324 | 0.057 | 0.058 |
| | | 10 | 0.101 | 0.074 | 0.086 | 0.079 | 0.089 | 0.296 | 0.063 | 0.061 |
| $\alpha = 0.8$ | Mean | 0 | 0.790 | 0.794 | 0.776 | 0.776 | 0.701 | 0.499 | 0.765 | 0.763 |
| | | 0.5 | 0.827 | 0.864 | 0.796 | 0.809 | 0.753 | 0.694 | 0.753 | 0.732 |
| | | 1 | 0.853 | 0.899 | 0.820 | 0.841 | 0.797 | 0.748 | 0.771 | 0.740 |
| | | 2 | 0.890 | 0.932 | 0.855 | 0.879 | 0.849 | 0.804 | 0.804 | 0.764 |
| | | 5 | 0.941 | 0.967 | 0.912 | 0.930 | 0.915 | 0.862 | 0.866 | 0.821 |
| | | 10 | 0.967 | 0.982 | 0.947 | 0.958 | 0.950 | 0.905 | 0.911 | 0.871 |
| | Std. | 0 | 0.036 | 0.035 | 0.048 | 0.048 | 0.066 | 0.316 | 0.049 | 0.051 |
| | | 0.5 | 0.054 | 0.041 | 0.063 | 0.057 | 0.070 | 0.288 | 0.059 | 0.068 |
| | | 1 | 0.061 | 0.040 | 0.070 | 0.0580 | 0.071 | 0.281 | 0.059 | 0.068 |
| | | 2 | 0.064 | 0.037 | 0.071 | 0.053 | 0.063 | 0.250 | 0.056 | 0.065 |
| | | 5 | 0.054 | 0.028 | 0.063 | 0.043 | 0.048 | 0.225 | 0.050 | 0.058 |
| | | 10 | 0.042 | 0.021 | 0.050 | 0.032 | 0.036 | 0.189 | 0.040 | 0.051 |

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