Estimating fixed-effect panel stochastic frontier models by model transformation

Wang, Hung-Jen and Ho, Chia-Wen

December 2009

Estimating Fixed-Effect Panel Stochastic Frontier Models by Model Transformation*

Hung-Jen Wang† Chia-Wen Ho

December 2009

Abstract

Traditional panel stochastic frontier models do not distinguish between unobserved individual heterogeneity and inefficiency. They thus force all time-invariant individual heterogeneity into the estimated inefficiency. Greene (2005) proposes a true fixed-effect stochastic frontier model which, in theory, may be biased by the incidental parameters problem. The problem usually cannot be dealt with by model transformations owing to the nonlinearity of the stochastic frontier model. In this paper, we propose a class of panel stochastic frontier models which create an exception. We show that first-difference and within-transformation can be analytically performed on this model to remove the fixed individual effects, and thus the estimator is immune to the incidental parameters problem. Consistency of the estimator is obtained by either $N \to \infty$ or $T \to \infty$, which is an attractive property for empirical researchers.

Keywords: stochastic frontier models, fixed effects, panel data.

JEL Codes: C13, C16, C23.

*We thank William Greene, Ching-Fan Chung, and Chang-Ching Lin for helpful discussions. Comments and suggestions from Peter Schmidt, the associate editor, and two anonymous referees are particularly appreciated. Hung-Jen Wang is grateful for the financial support provided by National Science Council (NSC-93-2415-H-001-020).

†Corresponding Author. Hung-Jen Wang: Department of Economics, National Taiwan University, and Institute of Economics, Academia Sinica. 21 Hsu Chow Rd., Taipei 100, Taiwan. E-mail: wangh@ntu.edu.tw. Tel.: +886-2-23519641 ext. 667; fax: +886-2-23511826. Chia-Wen Ho: Department of Economics, The Ohio State University, 410 Arps Hall, 1945 N High Street, Columbus OH 43210. Email:ho.255@buckeyemail.osu.edu.

1 Introduction

An important advantage of using panel data in an empirical study is that effects of differences across individuals (individual effects) can be distinguished from effects changing over time within individuals. Although time-invariant and individual-specific effects are often unobservable, they frequently account for an important share of the heterogeneity in data. In the study of wage rates, for example, a worker's innate ability is an important determinant of his wage. Such ability is both time invariant and not directly observable to econometricians. For household consumption behaviors, time-constant personal/household tastes are important in explaining data variations. Regardless of the source of heterogeneity, failure to control for individual effects is likely to bias estimation results, especially when there is correlation between the effect and other explanatory variables in the model.

Unobservable individual effects also play an important role in the estimation of panel stochastic frontier models. In contrast to the conventional panel data literature, however, studies using stochastic frontier models often interpret individual effects as *inefficiency* (e.x., Schmidt and Sickle 1984), such as technical inefficiency in a stochastic production frontier model. In this approach, the model is estimated using traditional panel data methods which transform models to eliminate individual effects before estimation. After the model parameters are estimated, individual effects are recovered and then adjusted to conform to an inefficiency interpretation. This modeling and estimation strategy is easy to use, but at the cost of not allowing for individual effects (in the traditional sense) to exist alongside inefficiency effects. In other words, all the individual effects are attributed to inefficiency, and inefficiency accounts for all the time-invariant and individual-specific effects in the data. Another feature of this approach is that inefficiency is necessarily time-invariant which may be problematic when operating under the competitive market assumption.

This time-invariant inefficiency assumption has been relaxed in a number of subsequent studies including Kumbhakar (1990) and Battese and Coelli (1992). These studies specify inefficiency (u_{it}) as a product of two components. One of the components is a function of time and the other is an individual specific effect so that $u_{it} = G(t) \cdot u_i$. In these models, however, the time-varying pattern of inefficiency is the same for all individuals, so the problem of inseparable inefficiency and individual heterogeneity remains.

In all these models, the inability to separate inefficiency and individual heterogeneity is likely to limit their applicability in empirical studies. This point is lucidly made in Greene (2005) which conducts a cross-country comparison of health care service efficiency and argues that the (in)efficiency effect and the time-invariant country-specific effect are different and should be accounted for separately in the estimation. If, for example, the country-specific heterogeneity is not adequately controlled for, then the estimated inefficiency may be picking up country-specific heterogeneity in addition to or even instead of inefficiency. In this way, the inability of a model to estimate individual effects in addition to the inefficiency effect poses a problem for empirical research. Greene then proposes the "true fixed-effect" model which is essentially a standard fixed-effect panel data model augmented by the inefficiency effect (u_{it}). The latter effect is allowed to change over time and across individuals in the model.

However, including both the inefficiency effect and fixed individual effects in the model significantly complicates its estimation. For a fixed-effect model, the number of fixed-effect parameters (also called incidental parameters since their values are usually not of direct interest) increases with the number of individuals (N). In this situation, the conventional asymptotic result, which relies on $N \to \infty$, cannot be applied and estimates of the incidental parameters are necessarily inconsistent for a fixed T. For many estimators, inconsistency may also contaminate the estimates of the model's other parameters; the issue is referred to as the incidental parameters problem (Neyman and Scott 1948). For instance, for linear models with normal errors, the maximum likelihood estimator (MLE) of the slope coefficients is still consistent, but that of the variance-covariance matrix is inconsistent (Kendall and Stuart 1973, Mak 1982). For nonlinear models, such as the binomial logit, the MLE of all of the model parameters is inconsistent in general. Incidental parameters would not be an issue and MLE would be consistent if $T \to \infty$, but this condition is seldom met in empirical applications.

Aside from the statistical issue, there is also a related computational problem. It arises because the number of parameters to be estimated is at least N, and so maximizing the model's log-likelihood function may be difficult when N is large.[1]

The literature proposes some solutions to the incidental parameters problem for some of the

[1]Recent developments in computer algorithms have relaxed this constraint to some extent. For instance, the algorithm adopted by Greene (2005) is able to handle large problems and is available in the LimDep package.

models. The key to these solutions usually lies in removing the incidental parameters before estimation. One popular approach, which is widely used in linear models, is to transform the model by first-differencing or by within-transformation and then obtaining the marginal MLE (MMLE). Alternatively, a conditional likelihood may be formed if a sufficient statistic exists for the fixed effects, yielding a conditional MLE (CMLE). The likelihood functions of MMLE and CMLE do not contain incidental parameters, and the estimators are thus consistent (e.g., Cornwell and Schmidt 1992).

These methods, however, are not readily applicable to stochastic frontier models. For the MMLE, the transformation is usually intractable because of the nonlinearity of the model. For the CMLE, the sufficient statistic is yet to be found. On the other hand, Greene (2005) suggested that the model may be estimated by MLE where individual dummies are included for the fixed effects. The numerical issue of estimating a large number of parameters is then handled by an advanced numerical maximization algorithm. Using a Monte Carlo experiment on a cross-country health-care data set, Greene (2005) found that the incidental parameters problem does not affect the slope coefficients of a stochastic frontier model, while there is also evidence suggesting that the variance parameters are more likely to be affected when T is not large.

In this paper, we propose a different panel stochastic frontier model that has the true fixed-effect model specification and yet allows model transformations to be done while keeping the likelihood function tractable. After transforming the model by either first-difference or within-transformation, the fixed effects are removed before estimation based on which we obtain consistent MMLE for the panel stochastic frontier model. Our model differs from Greene's in three aspects. (1) Removing the fixed-effect parameters avoids the incidental parameters problem entirely, and consistency can be obtained by $N \to \infty$. (2) The model we consider is flexible in the sense that it allows the pre-truncation mean of the inefficiency variable to be non-zero (e.g., truncated-normal) and it accommodates exogenous determinants of inefficiency in the model. (3) No special maximization routine is required.

The proposed model shares important characteristics of the scaling-property model proposed by Wang and Schmidt (2002). The authors discussed in the paper the theoretical appeals of the scaling property in the context of cross-sectional data. Alvarez et al. (2006) discuss the use of scaling-property model in the panel data context. Here, we show that the property can be manipulated such that model transformations of either first-difference or within-transformation can be performed analytically. We conduct a Monte Carlo experiment to evaluate the performance of the estimator, paying particular attention to the effects from different values of N and T. We also compare the results to those of the dummy-variable based approach. Finally, we illustrate the use of the estimator in a capital investment model with a financing constraint using data from Taiwan.

The rest of the paper is organized as the follows. Section 2 presents the model and shows how the first-difference and within-transformation of the model can be performed. This section also provides marginal likelihood functions and the formula for estimating the inefficiency index for both of the transformed models. Section 3 provides Monte Carlo results on the models, which is followed by an empirical example in section 4. Conclusions of the paper are given in section 5.

2 The Model

Consider a stochastic frontier model with the following specifications:

$$y_{it} = \alpha_i + \boldsymbol{x}_{it}\boldsymbol{\beta} + \varepsilon_{it}, \tag{1}$$

$$\varepsilon_{it} = v_{it} - u_{it}, \tag{2}$$

$$v_{it} \sim N(0, \sigma_v^2), \tag{3}$$

$$u_{it} = h_{it} \cdot u_i^*, \tag{4}$$

$$h_{it} = f(\boldsymbol{z}_{it}\boldsymbol{\delta}), \tag{5}$$

$$u_i^* \sim N^+(\mu, \sigma_u^2), \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{6}$$

In this setup, α_i is individual i's fixed unobservable effect, $\boldsymbol{x}_{it}$ is a $1 \times K$ vector of explanatory variables, v_{it} is a zero-mean random error, u_{it} is a stochastic variable measuring inefficiency, and h_{it} is a positive function of a $1 \times L$ vector of non-stochastic inefficiency determinants ($\boldsymbol{z}_{it}$). Neither of the vectors of $\boldsymbol{x}_{it}$ and $\boldsymbol{z}_{it}$ contain constants (intercepts) because they are not identified. The notation "+" indicates that the underlying distribution is truncated from below at zero so that realized values of the random variable u_i^* are positive. If we set μ equal to 0, then u_i^* follows a half-normal distribution. The random variable u_i^* is independent of all T observations on v_{it}, and both u_i^* and v_{it} are independent of all T observations on $\{\boldsymbol{x}_{it}, \boldsymbol{z}_{it}\}$. For example, in a study of technical inefficiency of production, y_{it} is the log of output, $\boldsymbol{x}_{it}$ is a vector of log inputs and other factors affecting production, u_{it} is the technical inefficiency which measures the percentage (when multiplied by 100) of output loss due to inefficiency, and $\boldsymbol{z}_{it}$ is a vector of variables explaining the inefficiency.

The above model can be seen as a panel extension of the cross-sectional model of Wang and Schmidt (2002) (which attributed the idea to Simar et al. 1994). The extension shows up in the inclusion of the individual effects (α_i) and in the specification of the time invariant "basic" distribution u_i^*. As will be shown later, the time invariant assumption of u_i^* holds the key to a tractable model transformation.[2]

The above model exhibits the "scaling property" that, conditional on $\boldsymbol{z}_{it}$, the one-sided error term equals a scaling function h_{it} multiplied by a one-sided error distributed independently of $\boldsymbol{z}_{it}$. With this property, the shape of the underlying distribution of inefficiency is the same for all individuals, but the scale of the distribution is stretched or shrunk by observation-specific factors $\boldsymbol{z}_{it}$. The time invariant specification of u_i^* allows the inefficiency u_{it} to be correlated over time for a given individual. Compared to the independence assumption of u_{it} used in some other panel models, the correlated inefficiency is another appealing property of the current model. Wang and Schmidt (2002) and Alvarez et al. (2006) discussed other advantages of the scaling property.

Whether the scaling property holds in the data is ultimately an empirical question. Nevertheless, note that the specification nests some of the models in the literature as special cases. By setting $\mu = 0$, the model is the same as that in Reifschneider and Stevenson (1991), Caudill and Ford (1993) and Caudill, Ford, and Gropper (1995). Using a time trend variable in the place of $\boldsymbol{z}_{it}$, i.e., $f(\boldsymbol{z}_{it}\boldsymbol{\delta}) = f(z_t\delta)$, the model essentially mimics the one proposed by Kumbhakar (1990) and Battese and Coelli (1992).

In the next two sections, we show that the fixed individual effect α_i can be removed from the model by either first-differencing or within-transforming the model.

[2] Alvarez et al. (2006) assumed that the basic distribution is u_{it}^*, although they briefly mentioned u_i^* as another possible modeling strategy for panel data (p.205).

2.1 First-Difference

We first define the following notation: $\Delta w_{it} = w_{it} - w_{it-1}$, and the stacked vector of Δw_{it} for a given i and $t = 2, \ldots, T$ is denoted as $\Delta \tilde{w}_i = (\Delta w_{i2}, \Delta w_{i3}, \ldots, \Delta w_{iT})'$. Provided that the scaling function h_{it} is not constant,[3] the model after the first-difference is

$$\Delta \tilde{y}_i = \Delta \tilde{\boldsymbol{x}}_i \boldsymbol{\beta} + \Delta \tilde{\varepsilon}_i, \tag{7}$$

$$\Delta \tilde{\varepsilon}_i = \Delta \tilde{v}_i - \Delta \tilde{u}_i, \tag{8}$$

$$\Delta \tilde{v}_i \sim \mathrm{MN}(0, \Sigma), \tag{9}$$

$$\Delta \tilde{u}_i = \Delta \tilde{h}_i u_i^*, \tag{10}$$

$$u_i^* \sim N^+(\mu, \sigma_u^2), \qquad i = 1, \ldots, N. \tag{11}$$

The first-difference introduces correlations of Δv_{it} within the ith panel, and the $(T-1) \times (T-1)$ variance-covariance matrix of the multivariate normal distribution of $\Delta \tilde{v}_i = (\Delta v_{i2}, \Delta v_{i3}, \ldots, \Delta v_{iT})'$ is

$$\Sigma = \begin{bmatrix} 2\sigma_v^2 & -\sigma_v^2 & 0 & \ldots & 0 \\ -\sigma_v^2 & 2\sigma_v^2 & -\sigma_v^2 & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -\sigma_v^2 \\ 0 & 0 & \ldots & -\sigma_v^2 & 2\sigma_v^2 \end{bmatrix}. \tag{12}$$

The matrix has $2\sigma_v^2$ on the diagonal and $-\sigma_v^2$ on the off-diagonals.

It is noteworthy that the model in (7) to (11) *looks* similar to the cross-sectional model of Wang and Schmidt (2002) except for the multivariate normal distribution and the obvious transformation of variables. More importantly, the truncated normal distribution of u_i^* is not affected by the transformation. This key aspect of the model leads to a tractable likelihood function. Alternatively, consider a typical and simple model specification in which (4) to (6) are replaced by $u_{it} \sim N^+(0, \breve{\sigma}_u^2)$. Even in its simplest form, first-differencing u_{it} will not result in a known distribution, and the joint distribution involving the Δv_{it} terms would be intractable.

[3]The condition requires that z_{it} contains at least one variable which changes values over time.

After tedious but straightforward derivation, the marginal log-likelihood function of panel i in the model is

$$\ln L_i^D = -\frac{1}{2}(T-1)\ln(2\pi) - \frac{1}{2}\ln(T) - \frac{1}{2}(T-1)\ln(\sigma_v^2) - \frac{1}{2}\Delta\tilde{\varepsilon}_i'\Sigma^{-1}\Delta\tilde{\varepsilon}_i + \frac{1}{2}\left(\frac{\mu_*^2}{\sigma_*^2} - \frac{\mu^2}{\sigma_u^2}\right) + \ln\left(\sigma_*\Phi\left(\frac{\mu_*}{\sigma_*}\right)\right) - \ln\left(\sigma_u\Phi\left(\frac{\mu}{\sigma_u}\right)\right), \tag{13}$$

where

$$\mu_* = \frac{\mu/\sigma_u^2 - \Delta\tilde{\varepsilon}_i'\Sigma^{-1}\Delta\tilde{h}_i}{\Delta\tilde{h}_i'\Sigma^{-1}\Delta\tilde{h}_i + 1/\sigma_u^2}, \tag{14}$$

$$\sigma_*^2 = \frac{1}{\Delta\tilde{h}_i'\Sigma^{-1}\Delta\tilde{h}_i + 1/\sigma_u^2}, \tag{15}$$

$$\Delta\tilde{\varepsilon}_i = \Delta\tilde{y}_i - \Delta\tilde{\boldsymbol{x}}_i\boldsymbol{\beta}. \tag{16}$$

In the expressions, Φ is the cumulative density function of a standard normal distribution. The marginal log-likelihood function of the model is obtained by summing the above function over $i = 1, \ldots, N$. The model parameters are estimated by numerically maximizing the marginal log-likelihood function of the model.

The Inefficiency Index

For many empirical applications of stochastic frontier models, it is of great importance to compute observation-specific technical inefficiency. The conditional expectation estimator suggested in Jondrow et al. (1982), $\mathrm{E}(u_i|\varepsilon_i)$ evaluated at $\varepsilon_i = \hat{\varepsilon}_i$, is often adopted for this purpose (for simplicity we use notation implying a cross-sectional model here). For the model presented above, the similar $\mathrm{E}(u_{it}|\varepsilon_{it})$ evaluated at $\varepsilon_{it} = \hat{\varepsilon}_{it}$ can be used, noting that $\hat{\varepsilon}_{it} = y_{it} - \hat{\alpha}_i - \boldsymbol{x}_{it}\hat{\boldsymbol{\beta}}$ where the value of $\hat{\alpha}_i$ is discussed later in (31).

Instead of conditioning on the level of ε_{it}, an alternative (*modified*) way to estimate the inefficiency index is to perform the conditional expectation of u_{it} on the vector of differenced ε_{it}, i.e., $\Delta\tilde{\boldsymbol{\varepsilon}}_i = \Delta\tilde{y}_i - \Delta\tilde{\boldsymbol{x}}_i\boldsymbol{\beta}$. Note that $\Delta\tilde{\boldsymbol{\varepsilon}}_i$ does not contain α_i. The advantages of using the modified estimator is that (1) the vector $\Delta\tilde{\boldsymbol{\varepsilon}}_i$ contains all the information of individual i in the sample, and that (2) the estimator depends on $\hat{\boldsymbol{\beta}}$ (for which the variance is of order $1/((N-1)T)$) but not $\hat{\alpha}_i$

(for which the variance order is $1/T$). The second property is particular appealing when T of the sample is not large. The derivation of the equation is again tedious but straightforward:

$$\mathrm{E}(u_{it}|\Delta\tilde{\boldsymbol{\varepsilon}}_i) = h_{it}\left[\mu_* + \frac{\phi(\frac{\mu_*}{\sigma_*})\sigma_*}{\Phi(\frac{\mu_*}{\sigma_*})}\right], \tag{17}$$

which is evaluated at $\Delta\tilde{\boldsymbol{\varepsilon}}_i = \Delta\hat{\tilde{\boldsymbol{\varepsilon}}}_i$.

2.2 Within-Transformation

By within-transformation, the sample mean of each panel is subtracted from every observation in the panel. The transformation thus removes the time-invariant individual effect from the model. The following notation is helpful in discussing the model: $w_{i\boldsymbol{.}} = (1/T)\sum_{t=1}^{T} w_{it}$, $w_{it\boldsymbol{.}} = w_{it} - w_{i\boldsymbol{.}}$, and that the stacked vector of $w_{it\boldsymbol{.}}$ for a given i is $\tilde{w}_{i\boldsymbol{.}} = (w_{i1\boldsymbol{.}}, w_{i2\boldsymbol{.}}, \ldots, w_{iT\boldsymbol{.}})'$. The model after the transformation is

$$\tilde{y}_{i\boldsymbol{.}} = \tilde{\boldsymbol{x}}_{i\boldsymbol{.}}\boldsymbol{\beta} + \tilde{\varepsilon}_{i\boldsymbol{.}}, \tag{18}$$

$$\tilde{\varepsilon}_{i\boldsymbol{.}} = \tilde{v}_{i\boldsymbol{.}} - \tilde{u}_{i\boldsymbol{.}}, \tag{19}$$

$$\tilde{v}_{i\boldsymbol{.}} \sim \mathrm{MN}(0, \Pi), \tag{20}$$

$$\tilde{u}_{i\boldsymbol{.}} = \tilde{h}_{i\boldsymbol{.}} u_i^*, \tag{21}$$

$$u_i^* \sim N^+(\mu, \sigma_u^2), \qquad i = 1, \ldots, N. \tag{22}$$

The variance-covariance matrix of $\tilde{v}_{i\boldsymbol{.}}$ is

$$\begin{aligned}
\Pi &= \begin{bmatrix} \sigma_v^2(1-1/T) & \sigma_v^2(-1/T) & \ldots & \sigma_v^2(-1/T) \\ \sigma_v^2(-1/T) & \sigma_v^2(1-1/T) & \ldots & \sigma_v^2(-1/T) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_v^2(-1/T) & \sigma_v^2(-1/T) & \ldots & \sigma_v^2(1-1/T) \end{bmatrix} \\
&= \sigma_v^2\left[I_T - \frac{\iota'\iota}{T}\right] \\
&= \sigma_v^2 M,
\end{aligned} \tag{23}$$

where ι is a $T \times 1$ vector of 1's. For (21), note that

$$u_{it\bullet} = u_{it} - u_{i\bullet} = h_{it}u_i^* - u_i^* \left(\frac{1}{T} \sum_{t=1}^{T} h_{it} \right) = (h_{it} - h_{i\bullet})u_i^* = h_{it\bullet}u_i^*. \tag{24}$$

Equation (21) is the stacked vector of $u_{it\bullet}$.

The above model is complicated by the fact that M is a singular idempotent matrix and is not invertible. Here we use the singular multivariate normal distribution of Khatri (1968) to solve the problem. The density function of the vector $\tilde{v}_{i\bullet}$ which is defined on a $(T-1)$ dimensional subspace is

$$g(\tilde{v}_{i\bullet}) = \frac{1}{(\sqrt{2\pi})^{(T-1)} \sqrt{\sigma_v^{2(T-1)}}} \exp\left\{ -\frac{1}{2} \tilde{v}_{i\bullet}' \Pi^- \tilde{v}_{i\bullet} \right\}, \tag{25}$$

where Π^- indicates the generalized inverse of Π, and $(T-1)\sigma_v^2$ is the product of nonzero eigenvalues of Π.[4] The model's marginal likelihood function is then derived based on the joint distribution of $\tilde{v}_{i\bullet}$ and $\tilde{u}_{i\bullet}$. The marginal log-likelihood function of the ith panel is

$$\begin{aligned} \ln L_i^W = &- \frac{1}{2}(T-1)\ln(2\pi) - \frac{1}{2}(T-1)\ln(\sigma_v^2) - \frac{1}{2}\tilde{\varepsilon}_{i\bullet}' \Pi^- \tilde{\varepsilon}_{i\bullet} \\ &+ \frac{1}{2}\left(\frac{\mu_{**}^2}{\sigma_{**}^2} - \frac{\mu^2}{\sigma_u^2} \right) + \ln\left(\sigma_{**}\Phi\left(\frac{\mu_{**}}{\sigma_{**}} \right) \right) - \ln\left(\sigma_u \Phi\left(\frac{\mu}{\sigma_u} \right) \right), \end{aligned} \tag{26}$$

where

$$\mu_{**} = \frac{\mu/\sigma_u^2 - \tilde{\varepsilon}_{i\bullet}' \Pi^- \tilde{h}_{i\bullet}}{\tilde{h}_{i\bullet}' \Pi^- \tilde{h}_{i\bullet} + 1/\sigma_u^2}, \tag{27}$$

$$\sigma_{**}^2 = \frac{1}{\tilde{h}_{i\bullet}' \Pi^- \tilde{h}_{i\bullet} + 1/\sigma_u^2}, \tag{28}$$

$$\tilde{\varepsilon}_{i\bullet} = \tilde{y}_{i\bullet} - \tilde{\boldsymbol{x}}_{i\bullet}\boldsymbol{\beta}. \tag{29}$$

The marginal log-likelihood function of the model is obtained by summing the above function over $i = 1, \ldots, N$.

[4]Eigenvalues of an idempotent matrix are either 0 or 1, with the number of eigenvalues that are 1 equal to the rank of the matrix. The rank of the matrix M is $T-1$, so there is a total of $T-1$ eigenvalues equal to σ_v^2 for the matrix $\Pi = \sigma_v^2 M$.

The Inefficiency Index

As we discussed in the case of the first-differenced model, the formula of Jondrow et al. (1982) can be applied here after $\hat{\alpha}_i$ is recovered to obtain an observation-specific inefficiency index. The estimator may not work very well for small samples because of the large sample assumption used in recovering $\hat{\alpha}_i$. Again, we propose a modified estimator which does not require $\hat{\alpha}_i$ and thus does not suffer from the approximation problem. The estimator is based on the conditional expectation of u_{it} on $\tilde{\varepsilon}_{i\cdot} = \tilde{y}_{i\cdot} - \tilde{\boldsymbol{x}}_{i\cdot}\boldsymbol{\beta}$:

$$\mathrm{E}(u_{it}|\tilde{\varepsilon}_{i\cdot}) = h_{it}\left[\mu_{**} + \frac{\phi(\frac{\mu_{**}}{\sigma_{**}})\sigma_{**}}{\Phi(\frac{\mu_{**}}{\sigma_{**}})}\right], \tag{30}$$

which is evaluated at $\tilde{\varepsilon}_{i\cdot} = \hat{\tilde{\varepsilon}}_{i\cdot}$

Recovering Values of Individual Fixed Effects

Although the individual effects α_i's are not estimated in the model, their values can be recovered after the model's other parameters are estimated by either of the transformed models proposed above. A T-consistent estimator of α_i may be obtained by solving the first order condition for α_i from the un-transformed log-likelihood function of the model assuming all other parameters are known. Doing so we have:

$$\hat{\alpha}_i = y_{i\cdot} - \boldsymbol{x}_{i\cdot}\hat{\boldsymbol{\beta}} + \hat{\mu}_{***}\hat{h}_{i\cdot} + \hat{\sigma}_{***}\hat{h}_{i\cdot}\frac{\phi\left(\frac{\hat{\mu}_{***}}{\hat{\sigma}_{***}}\right)}{\Phi\left(\frac{\hat{\mu}_{***}}{\hat{\sigma}_{***}}\right)}, \tag{31}$$

where

$$\hat{\mu}_{***} = \frac{\hat{\mu}\hat{\sigma}_u^{-2} - \hat{\sigma}_v^{-2T}\sum_t \hat{\varepsilon}_{it}\hat{h}_{it}}{\hat{\sigma}_v^{-2T}\sum_t \hat{h}_{it}^2 + \hat{\sigma}_u^{-2}}, \tag{32}$$

$$\hat{\sigma}_{***}^2 = \frac{\hat{\sigma}_v^{2T}}{\sum_t \hat{h}_{it}^2 + \hat{\sigma}_v^{2T}\hat{\sigma}_u^{-2}}. \tag{33}$$

The hat symbol indicates the values estimated from either the first-difference model or the within-transformation model.

2.3 Equivalence of the Two Models

Although the two models proposed above may seem different, the likelihood functions are actually the same ($L_i^D \propto L_i^W$). To prove the equivalence of the estimates, we first observe that the models' likelihood functions, as stated in (13) to (16) and (26) to (29), differ only in terms involving the inverse of the variance-covariance matrices. In particular, if the following equations can be established, then the equivalence of the likelihood functions is obtained (for generality, Π^{-1} is used in lieu of Π^{-} in this section):

$$\tilde{\boldsymbol{\varepsilon}}_i' \Pi^{-1} \tilde{\boldsymbol{\varepsilon}}_i = \Delta \tilde{\boldsymbol{\varepsilon}}_i' \Sigma^{-1} \Delta \tilde{\boldsymbol{\varepsilon}}_i, \tag{34}$$

$$\tilde{\boldsymbol{h}}_i' \Pi^{-1} \tilde{\boldsymbol{h}}_i = \Delta \tilde{\boldsymbol{h}}_i' \Sigma^{-1} \Delta \tilde{\boldsymbol{h}}_i, \tag{35}$$

$$\tilde{\boldsymbol{\varepsilon}}_i' \Pi^{-1} \tilde{\boldsymbol{h}}_i = \Delta \tilde{\boldsymbol{\varepsilon}}_i' \Sigma^{-1} \Delta \tilde{\boldsymbol{h}}_i. \tag{36}$$

A proof is sketched as follows.

Let D be a $T - 1 \times T$ matrix of the first-difference projection matrix,

$$D = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & -1 & 1 \end{bmatrix}. \tag{37}$$

The first-difference model is obtained by projecting the original model onto D. Specifically,

$$\Delta \tilde{\boldsymbol{\varepsilon}}_i = D \boldsymbol{\varepsilon}_i, \tag{38}$$

$$\text{and} \quad \Sigma = \sigma_v^2 D D'. \tag{39}$$

The within-transformation projection matrix can also be constructed from D (see also (23)):

$$D'(DD')^{-1} D = I_T - \frac{\iota' \iota}{T}. \tag{40}$$

Using the projection matrix, we have

$$\tilde{\varepsilon}_i = D'(DD')^{-1}D\varepsilon_i, \tag{41}$$

$$\text{and} \quad \Pi = \sigma_v^2 D'(DD')^{-1}D. \tag{42}$$

Using the results, it is easy to show that (34) is true.

$$\begin{aligned}
\tilde{\varepsilon}_i'\Pi^{-1}\tilde{\varepsilon}_i &= (D'(DD')^{-1}D\varepsilon_i)'(D'(DD')^{-1}D)(D'(DD')^{-1}D\varepsilon_i)\sigma_v^{-2} \\
&= (D\varepsilon_i)'((DD')^{-1})'(D\varepsilon_i)\sigma_v^{-2} \\
&= (D\varepsilon_i)'(DD')^{-1}(D\varepsilon_i)\sigma_v^{-2} \\
&= \Delta\tilde{\varepsilon}_i'\Sigma^{-1}\Delta\tilde{\varepsilon}_i.
\end{aligned} \tag{43}$$

Similarly, it is easy to show that (35) and (36) are also true. Therefore, the log-likelihood functions are the same for the first-difference and the within-transformation models.

Remarks

We end this section with two remarks. First, although the models are derived assuming balanced panels, the results can be easily modified for unbalanced panel models. The only required modification is changing T to T_i (≥ 2) where T_i is individual i's number of observations in the data. Secondly, the lost degrees of freedom from the removal of individual effects in the model transformations are accounted for in the MLE. No additional adjustment needs to be taken. For the first-differenced model, the (automatic) adjustment is obvious since only $T-1$ observations are used in the estimation from each panel. For the within-transformed model, (25) is derived based on a $(T-1)$ dimensional subspace. Therefore, the marginal log-likelihood function loses one degree of freedom for each panel.

For the Monte Carlo analysis in the following section, both the first-difference and the within-transformation models were programmed using Stata 10 software. The programs are available from the authors upon request.

3 A Monte Carlo Study

In this section, we conduct Monte Carlo experiments on fixed-effect panel stochastic frontier models. We first conduct a small-scale experiment on a simple and untransformed model for which the fixed effects are estimated by dummy variables. The results are complements to Greene's (2005) study, and show bias from the incidental parameters problem. We then carry out a more extensive Monte Carlo study on models for which the individual effects are removed before estimation by first-difference and within model transformation.

3.1 First-Difference and Within-Transformation

We consider a panel stochastic frontier model with the following specification:

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it}, \tag{44}$$

$$\varepsilon_{it} = v_{it} - u_{it}, \tag{45}$$

$$v_{it} \sim N(0,\ \sigma_v^2), \tag{46}$$

$$u_{it} = \exp(\delta z_{it}) \cdot u_i^*, \tag{47}$$

$$u_i^* \sim N^+(\mu,\ \sigma_u^2), \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T. \tag{48}$$

To generate data for simulation, we first draw fixed-effect parameters (α_i's) from a uniform distribution in $[0, 1]$. The x_{it} is then drawn from a $N(\alpha_i, 1)$ normal distribution for the ith panel, $i = 1, \ldots, N$. This data generation method induces correlations between α_i and x_{it}, and the correlation coefficient is around 0.27 in our samples.[5] The z_{it} is generated from $N(0, 1)$. For the base case, the selected parameter values are $\{\beta = 0.5,\ \delta = 0.5,\ \sigma_v^2 = 0.1,\ \mu = 0.5,\ \sigma_u^2 = 0.2,\ N = 100,\ T = 5\}$. The simulation is conducted with 1,000 replications.

The relatively small values chosen for N and T are not uncommon in empirical applications. Since we are interested in observing the effects of changes in N and T on parameter estimation, we also considered alternative values for N and T. Specifically, we also included $N = 200$, 300,

[5] The correlation is created to simulate scenarios in which the fixed effect specification is often called upon. We also experimented with cases where there is no correlation between α_i and x_{it}. The estimator still applies and the results (not shown) are similar to what we reported in the paper. Results are available from the authors upon request.

and $T = 10$, 15. The variance ratio σ_u^2/σ_v^2 may also affect model estimation. Therefore, we hold σ_v^2 fixed at 0.1 and consider an alternative value of σ_u^2 equal to 0.15. All possible combinations of the alternative values of N, T, and σ_u^2 are included in the study. Simulation results of the various cases are sorted by values of T and are presented in Table 1 ($T = 5$), Table 2 ($T = 10$), and Table 3 ($T = 15$).

Table 1 presents cases with small T ($T = 5$). Notice that estimation results are similar for the first-difference and the within-transformation models, which confirms the equivalence of the models. When both T and N are small ($N = 100$), results show that β, δ, and σ_v^2 are estimated well, although the MSEs of $\hat{\mu}$ and $\hat{\sigma}_u^2$ are somewhat larger. Large variances contribute to the large MSEs in both cases, although noticeable bias is also observed for $\hat{\sigma}_u^2$. The correlation coefficients between the estimated inefficiency index and the true index are around 0.86 to 0.875, which is quite high given the small sample size.

Increases in N quickly reduce the MSEs of $\hat{\mu}$ and $\hat{\sigma}_u^2$. For $\hat{\sigma}_u^2$, the MSE of the first-difference model with $\sigma_u^2 = 0.2$ falls from 0.027 to 0.011 when N increases from 100 to 200, and the figure continues to fall to 0.006 when N is 300. The improvement is significant, and stems from smaller variance and the smaller bias of the estimate. The reduction in bias from larger N is an important properly of the estimators. As we will show in the next subsection, the bias reduction does not take place in the dummy-variable model. The correlation coefficient between the estimated and the true inefficiency index also improves when N increases as expected.

It is well known that stochastic frontier models are difficult to estimate when σ_u^2 is small. The panel on the right of Table 1 shows the results with a small σ_u^2 ($\sigma_u^2/\sigma_v^2 = 1.5$). Compared to the models with $\sigma_u^2/\sigma_v^2 = 2$, the correlation coefficient between the estimated and the true inefficiency index is smaller in all cases. Otherwise, the parameter estimates are qualitatively similar.

Table 2 presents results with a larger T ($T = 10$). Because the first-difference and the within-transformed models are shown to be identical, we only report results from the first-difference model. As expected, when the parameter configuration is held unchanged, estimation results improve with a larger T. For instance, the MSE of $\hat{\sigma}_u^2$ from the first-difference model with $N = 100$ falls from 0.027 to 0.009 when T increases from 5 to 10. As with the results of an increase in N, the reduction in MSE stems from both a smaller bias and smaller variances in the estimate. The rest of the table shows cases with larger values of N, and the results are as expected: smaller MSEs for all the parameters and larger correlation coefficients between the estimated and the true values of inefficiency index. Table 3 presents results of models with a $T = 15$. Regardless of the size of N, all parameters are estimated very well.

Finally, we add Table 4 which reports results from models with larger values of σ_u^2. It is obvious from the table that larger σ_u^2 makes μ and σ_u^2, both of which parameters of u_i^*, to be estimated less precisely. On the other hand, the expected inefficiency index ($\mathrm{E}(u_{it})$) is computed conditional on the composed error of $\varepsilon_{it} = v_{it} - u_{it}$, and so the conditional information is more useful if u_{it} accounts for a larger share of ε_{it}'s variance. The result is a higher correlation between the true and the estimated inefficiency index when σ_u^2 increases. The above observations are also found in the preceding tables.

3.2 Dummy Variable Models: A Comparison

As shown in the previous subsection, parameters are estimated very well in transformed models, and the estimation consistency improves with increases in either N or T. To further understand the performance of the estimators, in this subsection we provide simulation results of the model in which the fixed individual effects are estimated by dummy variables.[6] The model suffers from the incidental parameters problem, and simulation results show the consequences of not removing incidental parameters prior to estimation.

Since we wish to observe how values of N and T affect estimation, we simulate models with different configurations of $N = 100, 200, 300$ and $T = 5, 10, 15$. We choose $\sigma_u^2 = 0.2$ for all the models and keep other parameters the same as those used in the previous section. Results are presented in Table 5 (for selected models) and in Figure 1. For comparison, we also reproduce results of the corresponding first-difference models in the table and the figure.

For Model 1 ($N = 100$ and $T = 5$), the β is estimated very well and the estimate of σ_v^2 is also reasonably sound. The rest of the parameters, however, are very poorly estimated: $\hat{\delta} = 0.810$ ($\delta = 0.5$), $\hat{\mu} = 0.232$ ($\mu = 0.5$), and $\hat{\sigma}_u^2 = 0.056$ ($\sigma_u^2 = 0.20$). Note that these are all parameters

[6] This is similar to the "true fixed effect" model of Greene (2005) in that the fixed effects are estimated by dummy variables. The only difference is in the specification of u_{it} which has a truncated-normal distribution with exogenous determinants in the current model. Greene assumes an i.i.d. half-normal distribution. Further investigation is needed to determine to what extent the results observed here pertain to Greene's model.

of u_{it}. The biases are large and significant. The correlation coefficient between the estimated inefficiency ($\mathrm{E}(u_{it}|\varepsilon_{it})$ evaluated at $\varepsilon_{it} = \hat{\varepsilon}_{it}$) and the true inefficiency is 0.711, clearly smaller than the value of 0.871 obtained from the first-difference model.

The above finding is consistent with Greene (2005) in which the author showed that the incidental parameters problem does not cause bias to the slope coefficients. The estimation problem arises mainly in the error variances estimation. However, since estimated inefficiency of a stochastic frontier model is based on the error variance, the empirical consequence of the incidental parameters problem cannot be ignored.

Model 2 keeps N the same and increases T to 15. As discussed earlier, larger T helps the dummy-variable model gain consistency. Table shows that the estimation indeed improves with T equal to 15, but the overall result is still unsatisfactory. For example, while $\hat{\delta}$ falls from 0.810 in Model 1 to 0.627 in Model 2, it still overestimates the true value by 25%. Given that effects of the inefficiency determinants often play an important role in empirical stochastic frontier analysis, this result should be alarming to empirical researchers. $\hat{\sigma}_u^2$ also suffers from a large bias of about -47%. The correlation coefficient between the true and the estimated index is 0.726, which is only a slight increase from the value of 0.711 obtained from Model 1. The first-difference model, on the other hand, reaps substantial gains from a larger T as shown in the table.

Model 3 increases N to 300 and keeps T at 5. As expected, the dummy variable model does not benefit from an increase in N. There is no appreciable change in the parameter estimation compared with Model 1. The first-difference model, on the contrary, improves substantially from an increase in N. Model 4 increases both N and T. Results of the dummy-variable model show improvements when compared to Model 1, which is likely due to the effect of the large T.

It is worthwhile to note that, for all the cases presented in Table 5, the dummy variable model tends to overestimate the importance of exogenous determinant of inefficiency (δ too large) while underestimate σ_u^2. Because the inefficiency index is fundamentally affected by the variance parameters, large biases in δ, μ, and σ_u^2 have negative consequences on the estimated inefficiency. Using the figures in Table 5, it can be shown that the sample mean of the inefficiency index from the dummy variable model is about 18% to 40% smaller than that of the first-difference model. The correlation coefficient between the true and the estimated inefficiency index is also much smaller with the dummy variable model.

Figure 1 plots the point estimates of $\hat{\delta}$, $\hat{\mu}$, and $\hat{\sigma}_u^2$ from all possible combinations of N and T in the simulation. Graphs in the left column have N fixed at 100 while T changes from 5 to 10 to 15. Graphs in the right column have T fixed at 5 as N changes from 100 to 200 to 300. The figure clearly shows that the estimates of the dummy-variable model do not benefit from increases in N (because of the incidental parameters problem). While the point estimates improve with larger T, with $T = 15$ the performance is still inferior to that of the first-difference model.

4 Empirical Example

We apply the estimator to a capital investment model of Taiwan for the period of 1999-2005. Wang (2003) solved a profit maximization problem for the firm's investment decision and showed that the solution has a frontier-type interpretation. The frontier in the model represents the frictionless level of investment and the one-sided deviation captures the financing constraint effects. Wang (2003) estimates the model using Taiwanese data from the period between 1989-1996, a period in which the economy underwent a series of reforms aimed at financial liberations. Results show that effects of financing constraints on investment reduced gradually along with the progress of financial reforms, and that both cash flow and asset size are helpful in explaining individual firms' financing constraints.

Here we estimate the model using data from 1999-2005, a period in which Taiwan's economic growth rate slowed down substantially. The slow down began with the Asian financial crisis in late 1997 and continued with a recession in 2002 which was one of the worst in Taiwan's recent history. Chen and Wang (2007) shows that Taiwan's credit market shrunk dramatically following the Asian financial crisis, and identifies an inward shift of supply (as opposed to demand) as the main cause of the credit slow down. Against this background, it is interesting to determine whether the effects of financing constraints on firms' investment changed during this period.

The model specification is similar to Wang (2003). Referring to (1) – (6), the dependent variable is $\ln(I/K)_{it}$ which is the log of the investment to capital ratio. Capital K_{it} is measured at the beginning of each period and is used to normalize most of the variables in the model in order to control for heteroscedasticity. The explanatory variables ($\boldsymbol{x}_{it}$) include the log of Tobin's Q ($\ln Q_{it}$) and the current and lag sales to capital ratios ($\ln(S/K)_{it}$, $\ln(S/K)_{it-1}$). The investment literature shows that Tobin's Q is a sufficient statistic of investment in the absence of market imperfection (e.g., Hayashi 1985, Osterberg 1989, Chirinko 1993, and Gilchrist and Himmelberg 1995), and sales variables are added to further explain investment behavior. The Tobin's Q variable is computed based on the method of Lewellen and Badrinath (1997). Implementation of this method is described in the appendix of Wang (2003). The firm fixed-effects (α_i) are also included in the model specification.

The $h(\cdot)$ function in (5) is specified as $\exp(\boldsymbol{z}_{it}\boldsymbol{\delta})$, where the $\boldsymbol{z}_{it}$ vector includes the cash flow ratio variable ($(CF/K)_{it}$) and the asset size variable ($\ln(asset)_{it}$). Wang (2003), following the literature, hypothesized that the extent of a firm's financing constraint on investment is inversely related to both of the variables. The random variable v_{it} is assumed to follow a zero-mean normal distribution and u_i^* is assumed to follow a half-normal distribution. Both of the variances are parameterized as follows in the estimation:

$$\sigma_v^2 = \exp(c_v), \qquad \sigma_u^2 = \exp(c_u),$$

where c_v and c_u are unconstrained constant parameters.

The empirical data is from the Taiwan Economic Journal Data Bank. The sample consists of data of 206 Taiwanese manufacturing firms publicly traded on the Taiwan Stock Exchange. Because a lag variable is used in the model specification, the actual estimation period is from the year 2000 to 2005 (T=6). There are a total of 1220 observations. Summary statistics are reported in Table 6.

Table 7 shows the estimation results from the dummy-variable model (Model 1) and the within-transformation model (Model 2). As indicated by Model 2 (our preferred model; log-likelihood value = -1363.732), the estimated coefficients on the Tobin's Q and sales ratio variables are all positive and significant at least at the 10% level. Regarding the effect of financing constraints on investment, the asset variable's coefficient is negative and significant, implying that the degree of financing constraint is smaller for larger firms. This result is consistent with findings in the literature (Gertler and Gilchrist 1994, Carpenter et al. 1994, and Gilchrist and Himmelberg 1995).

One reason for this is that larger firms are likely better equipped in providing collateral to mitigate the information problem in the capital market. Larger firms also tend to be older and more mature, so that the market has better access to and assessment of the firm's information.

On the other hand, the coefficient of the cash flow variable is estimated rather imprecisely. A possible explanation for our data is that Taiwanese firms in the sample period were cash-strapped in general due to recessions, and therefore there was not enough variation in cash flow across firms to show significant covariance with the extent of financing constraints. In any case, the validity of using cash flow to gauge financing constraints is controversial in the literature (e.x., Fazzari et al. 2000, Kaplan and Zingales 2000) and empirical results are mixed.

As for the dummy-variable model (Model 1; log-likelihood value = -1512.152), the coefficients of the Q and sales variables are quite close to those of Model 2. On the other hand, the coefficient of the log of asset is much larger in size compared to Model 2 while the estimate of σ_u^2 is much smaller. The mean of the conditional expectation of $\exp(-u_{it})$ also shows that the dummy-variable model implies a higher investment efficiency (lower financing constraints).

These observations are consistent with the simulation results (see, in particular, Model 3 in Table 5 for similar N and T), which show that the β coefficients are always similar in both models while the dummy-variable model tends to overestimate the impact of the inefficiency determinants (e.g., $\ln Assets_{it}$) and underestimate the size of σ_u^2.

We end this section by discussing the model's time-varying characteristic of the inefficiency index. As mentioned earlier, the model's inefficiency is a product of a time-varying function $f(\boldsymbol{z}_{it}\boldsymbol{\delta})$ and a time-invariant random variable u_i^*. This combination yields a specification that is in between the time-constant assumption of inefficiency ($u_{it} = u_i$, i.e., Schmidt and Sickles 1984) and the observation-independent assumption (e.g., Greene 2005). In a related context, Greene (2002, 2005) found that the difference between the predictions of model with time varying vs. time invariant inefficiency is vast and unsettling.

To see if the time-varying or the time-constant properties of the model dominates in the estimated efficiency, we compute, as an approximation, the mean and the standard deviation of the efficiency index within each firm. In particular, we assess the cross-time variation of the efficiency index by calculating the standard deviation of the index for each firm (separately) and then average the figures across firms. This yields a mean standard deviation of 0.020 (which would be 0 if the model has a time constant specification of inefficiency). On the other hand, the mean of the efficiency index across firms is 0.571. A one standard-deviation above and below the mean puts the efficiency index between 0.551 to 0.591 for this sample. Although not a precise measure, these numbers suggest that the time variation of inefficiency is on the lower side. However, the numbers are not totally unreasonable given that they are from firms in a six-year span. Whether the low time variation is due to data or is a property intrinsic to the proposed model specification remains an issue for further investigation.

5 Conclusion

Recent literature has emphasized the importance of separating inefficiency and fixed individual effects in a panel stochastic frontier model. In this paper, we propose a class of panel stochastic frontier models that take account of both time-varying inefficiency and time-invariant individual effects. An important feature of these models is that simple transformations can be performed to remove the fixed individual effects prior to estimation. The first-difference and within-transformation methods, which cannot normally be used on stochastic frontier models due to their complicated error structure, eliminate the problem of incidental parameters brought about by the inclusion of fixed individual effects in the model.

The transformed models proposed in this paper in general performed quite well in our Monte Carlo study. Most importantly, consistency of the parameter estimates can be improved by increasing either T or N (or both). In addition, because the fixed individual effects are removed by model transformations, the number of parameters to be estimated is no more than that of a cross-sectional model. Our models' desirable statistical properties and their ease of estimation should appeal to empirical researchers.

Similar to Greene's (2005) finding, our Monte Carlo results indicate that while the incidental parameters problem does not affect the estimation of slope coefficients, it does introduce bias to the estimated model residuals. The situation can not be remedied with a larger N, and can only be improved by increasing T. Since the inefficiency estimation is based on model residuals and the estimation is often at the core of a stochastic frontier analysis study, the incidental parameters problem should concern empirical researchers particularly when T is not large.

References

[1] Alvarez , A., Amsler , C., Orea , L., and Schmidt , P. (2006). "Interpreting and Testing the Scaling Property in Models where Inefficiency Depends on Firm Characteristics," *Journal of Productivity Analysis* **25**, pp. 201-212.

[2] Battese, G.E., and Coelli, T.J. (1992). "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India," *Journal of Productivity Analysis* **3**, pp. 153-69.

[3] Carpenter, R.E., Fazzari, S.M., and Petersen, B.C. (1994). "Inventory Investment, Internal-finance Fluctuations, and the Business Cycle," *Brookings Papers on Economic Activity* **2**, pp. 75-137.

[4] Caudill, S.B., and Ford, J.M. (1993). "Biases in Frontier Estimation Due to Heteroscedasticity," *Economics Letters* **41**, pp. 17-20.

[5] Caudill, S.B., Ford, J.M., and Gropper, D.M. (1995). "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity," *Journal of Business and Economic Statistics* **13**, pp. 105-11.

[6] Chen, N.-K., and Wang, H.-J. (2008). "Identifying the Demand and Supply Effects of Financial Crises on Bank Credit–Evidence from Taiwan," *Southern Economic Journal* **75**, pp. 26-49.

[7] Chirinko, R.S. (1993). "Business Fixed Investment Spending: Modeling Strategies, Empirical Results, and Policy Implications," *Journal of Economic Literature* **31**, pp. 1875-1911.

[8] Cornwell, C., and Schmidt, P. (1992). "Models for Which the MLE and the Conditional MLE Coincide," *Empirical Economics* **17**, pp. 67-75.

[9] Fazzari, S.M., Hubbard, R.G., and Petersen, B.C. (2000). "Investment-Cash Flow Sensitivities Are Useful a Comment on Kaplan and Zingales," *Quarterly Journal of Economics* **115**, pp. 695-705.

[10] Gertler, M., and Gilchrist, S. (1994). "Monetary Policy, Business Cycles, and the Behavior of Small Manufacturing Firms," *Quarterly Journal of Economics* **109**, pp. 309-340.

[11] Gilchrist, S., and Himmelberg, C.P. (1995). "Evidence on the role of cash flow for investment," *Journal Of Monetary Economics* **36**, pp. 541-572.

[12] Greene, W. (2002). "Fixed and Random Effects in Stochastic Frontier Models," Stern School of Business, New York University.

[13] Greene, W. (2005). "Reconsidering Heterogeneity In Panel Data Estimators Of The Stochastic Frontier Model," *Journal of Econometrics* **126**, pp. 269-303.

[14] Hayashi, F. (1985). "Corporate Finance Side of the Q Theory of Investment," *Journal of Public Economics* **27**, pp. 261-80.

[15] Jondrow, J., Lovell, C.A.K., Materov, I.S., and Schmidt, P. (1982). "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model," *Journal of Econometrics* **19**, pp. 233-38.

[16] Kaplan, S.N., and Zingales, L. (2000). "Investment-Cash Flow Sensitivities Are Not Valid Measures of Financing Constraints," *Quarterly Journal of Economics* **115**, pp. 707-712.

[17] Kendall, M.G., and Stuart, A. (1973). The Advanced Theory of Statistics: London: Griffin.

[18] Khatri, C.G. (1968). "Some Results for the Singular Normal Multivariate Regression Models," *Sankhya* **30**, pp. 267-280.

[19] Kumbhakar, S.C. (1990). "Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency," *Journal of Econometrics* **46**, pp. 201-211.

[20] Lewellen, W.G., and Badrinath, S.G. (1997). "On the Measurement of Tobin's q," *Journal of Financial Economics* **44**, pp. 77-122.

[21] Mak, T.K. (1982). "Estimation in the Presence of Incidental Parameters," *The Canadian Journal of Statistics* **10**, pp. 121-132.

[22] Neyman, J., and Scott, E.L. (1948). "Consistent Estimation from Partially Consistent Observations," *Econometrica* **16**, pp. 1-32.

[23] Osterberg, W.P. (1989). "Tobin's q, Investment, and the Endogenous Adjustment of Financial Structure," *Journal of Public Economics* **40**, pp. 293-318.

[24] Reifschneider, D., and Stevenson, R. (1991). "Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency," *International Economic Review* **32**, pp. 715-23.

[25] Schmidt, P., and Sickles, R.C. (1984). "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics* **2**, pp. 367-74.

[26] Simar, L., Lovell, C.A.K., and Eeckaut, P.V. (1994). "Stochastic Frontiers Incorporating Exogenous Influences on Efficiency," *Discusssion Papers No.9403, Institut de Statistique, University Catholique de Louvain.*

[27] Wang, H.-J. (2003). "A Stochastic Frontier Analysis of Financing Constraints on Investment: The Case of Financial Liberalization in Taiwan," *Journal of Business and Economic Statistics* **21**, pp. 406-19.

[28] Wang, H.-J., and Schmidt, P. (2002). "One-Step and Two-Step Estimation of The Effects of Exogenous Variables on Technical Efficiency Levels," *Journal of Productivity Analysis* **18**, pp. 129-44.

Table 1: T=5

$\beta = 0.5$, $\delta = 0.5$, $\mu = 0.5$, $\sigma_v^2 = 0.1$

	$\sigma_u^2/\sigma_v^2 = 2$				$\sigma_u^2/\sigma_v^2 = 1.5$			
	N=100				N=100			
	first-difference		within-transf.		first-difference		within-transf.	
	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\hat{\beta}$	0.500 (0.017)	3.0×10^{-4}	0.500 (0.017)	3.0×10^{-4}	0.500 (0.017)	2.9×10^{-4}	0.500 (0.017)	2.9×10^{-4}
$\hat{\delta}$	0.502 (0.084)	0.007	0.502 (0.084)	0.007	0.502 (0.088)	0.008	0.502 (0.087)	0.008
$\hat{\mu}$	0.491 (0.180)	0.032	0.489 (0.186)	0.035	0.497 (0.164)	0.027	0.496 (0.164)	0.027
$\hat{\sigma}_u^2$	0.232 (0.163)	0.027	0.233 (0.167)	0.029	0.177 (0.134)	0.019	0.176 (0.133)	0.018
$\hat{\sigma}_v^2$	0.099 (0.008)	6.4×10^{-5}	0.099 (0.008)	6.5×10^{-5}	0.099 (0.008)	6.4×10^{-5}	0.099 (0.008)	6.4×10^{-5}
$E(u_{it}\|\Theta)^*$	0.709 (0.137)	0.104	0.708 (0.137)	0.104	0.670 (0.137)	0.096	0.669 (0.136)	0.095
corr*	0.871		0.871		0.860		0.860	
	N=200				N=200			
	first-difference		within-transf.		first-difference		within-transf.	
	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\hat{\beta}$	0.499 (0.013)	1.6×10^{-4}	0.499 (0.013)	1.6×10^{-4}	0.499 (0.013)	1.6×10^{-4}	0.499 (0.013)	1.6×10^{-4}
$\hat{\delta}$	0.497 (0.057)	0.003	0.497 (0.057)	0.003	0.497 (0.060)	0.004	0.498 (0.060)	0.004
$\hat{\mu}$	0.500 (0.117)	0.014	0.497 (0.115)	0.013	0.502 (0.105)	0.011	0.501 (0.105)	0.011
$\hat{\sigma}_u^2$	0.218 (0.103)	0.011	0.219 (0.103)	0.011	0.165 (0.081)	0.007	0.165 (0.082)	0.007
$\hat{\sigma}_v^2$	0.100 (0.006)	3.0×10^{-5}	0.100 (0.005)	3.0×10^{-5}	0.100 (0.005)	3.0×10^{-5}	0.100 (0.005)	2.9×10^{-5}
$E(u_{it}\|\Theta)^*$	0.705 (0.093)	0.091	0.703 (0.091)	0.091	0.665 (0.091)	0.083	0.663 (0.091)	0.083
corr*	0.874		0.875		0.864		0.864	
	N=300				N=300			
	first-difference		within-transf.		first-difference		within-transf.	
	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\hat{\beta}$	0.500 (0.010)	9.9×10^{-5}	0.500 (0.010)	9.9×10^{-5}	0.500 (0.010)	9.9×10^{-5}	0.500 (0.010)	9.7×10^{-5}
$\hat{\delta}$	0.499 (0.047)	0.002	0.500 (0.048)	0.002	0.499 (0.049)	0.002	0.501 (0.050)	0.002
$\hat{\mu}$	0.498 (0.097)	0.009	0.496 (0.096)	0.009	0.499 (0.085)	0.007	0.495 (0.085)	0.007
$\hat{\sigma}_u^2$	0.211 (0.078)	0.006	0.210 (0.078)	0.006	0.159 (0.061)	0.004	0.159 (0.062)	0.004
$\hat{\sigma}_v^2$	0.100 (0.005)	2.2×10^{-5}	0.100 (0.005)	2.2×10^{-5}	0.100 (0.005)	2.2×10^{-5}	0.100 (0.005)	2.2×10^{-5}
$E(u_{it}\|\Theta)^*$	0.699 (0.073)	0.088	0.697 (0.074)	0.088	0.659 (0.071)	0.080	0.656 (0.072)	0.080
corr*	0.875		0.875		0.865		0.865	

* $\mathrm{E}(u_{it}|\Theta) = \mathrm{E}(u_{it}|\Delta\tilde{\varepsilon}_i)$ evaluated at $\Delta\tilde{\varepsilon}_i = \Delta\hat{\tilde{\varepsilon}}_i$ for the first-difference model, $\mathrm{E}(u_{it}|\Theta) = \mathrm{E}(u_{it}|\tilde{\varepsilon}_{i\cdot})$ evaluated at $\tilde{\varepsilon}_{i\cdot} = \hat{\tilde{\varepsilon}}_{i\cdot}$ for the within-transformation model. corr=corr$(E(u_{it}|\Theta), u_{it})$. Standard deviations are in the parenthesis.

Table 2: T=10

$\beta = 0.5$, $\delta = 0.5$, $\mu = 0.5$, $\sigma_v^2 = 0.1$

	$\sigma_u^2/\sigma_v^2 = 2$				$\sigma_u^2/\sigma_v^2 = 1.5$			
	N=100		N=200		N=100		N=200	
	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\hat{\beta}$	0.499 (0.011)	1.2×10^{-4}	0.500 (0.008)	6.5×10^{-5}	0.499 (0.011)	1.2×10^{-4}	0.500 (0.008)	6.4×10^{-5}
$\hat{\delta}$	0.502 (0.049)	0.002	0.501 (0.034)	0.001	0.502 (0.052)	0.003	0.502 (0.036)	0.001
$\hat{\mu}$	0.490 (0.136)	0.019	0.496 (0.081)	0.007	0.494 (0.112)	0.013	0.497 (0.071)	0.005
$\hat{\sigma}_u^2$	0.213 (0.096)	0.009	0.204 (0.059)	0.003	0.160 (0.072)	0.005	0.154 (0.046)	0.002
$\hat{\sigma}_v^2$	0.099 (0.005)	2.5×10^{-5}	0.100 (0.003)	1.2×10^{-5}	0.099 (0.005)	2.5×10^{-5}	0.100 (0.003)	1.2×10^{-5}
$E(u_{it}\|\Theta)^*$	0.697 (0.083)	0.053	0.693 (0.058)	0.049	0.657 (0.081)	0.050	0.653 (0.056)	0.046
corr*	0.931		0.933		0.922		0.924	
	N=300				N=300			
	Mean	MSE			Mean	MSE		
$\hat{\beta}$	0.500 (0.006)	4.0×10^{-5}			0.500 (0.006)	4.0×10^{-5}		
$\hat{\delta}$	0.502 (0.029)	0.001			0.502 (0.031)	0.001		
$\hat{\mu}$	0.498 (0.068)	0.005			0.499 (0.059)	0.003		
$\hat{\sigma}_u^2$	0.202 (0.049)	0.002			0.151 (0.037)	0.001		
$\hat{\sigma}_v^2$	0.100 (0.003)	8.6×10^{-6}			0.100 (0.003)	8.6×10^{-6}		
$E(u_{it}\|\Theta)^*$	0.692 (0.049)	0.048			0.652 (0.048)	0.046		
corr*	0.933				0.924			

* E($u_{it}|\Theta$) = E($u_{it}|\Delta\tilde{\varepsilon}_i$) evaluated at $\Delta\tilde{\varepsilon}_i = \Delta\hat{\tilde{\varepsilon}}_i$ for the first-difference model. corr=corr($E(u_{it}|\Theta), u_{it}$). Standard deviations are in the parenthesis.

Table 3: T=15

$\beta = 0.5,\ \delta = 0.5,\ \mu = 0.5,\ \sigma_v^2 = 0.1$

	$\sigma_u^2/\sigma_v^2 = 2$				$\sigma_u^2/\sigma_v^2 = 1.5$			
	N=100		N=200		N=100		N=200	
	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\hat{\beta}$	0.500 (0.009)	7.6×10^{-5}	0.500 (0.006)	3.5×10^{-5}	0.500 (0.009)	7.6×10^{-5}	0.500 (0.006)	3.5×10^{-5}
$\hat{\delta}$	0.499 (0.037)	0.001	0.498 (0.025)	0.001	0.499 (0.040)	0.002	0.498 (0.026)	0.001
$\hat{\mu}$	0.500 (0.106)	0.011	0.497 (0.073)	0.005	0.502 (0.089)	0.008	0.498 (0.061)	0.004
$\hat{\sigma}_u^2$	0.207 (0.073)	0.005	0.207 (0.051)	0.003	0.156 (0.055)	0.003	0.156 (0.038)	0.002
$\hat{\sigma}_v^2$	0.100 (0.004)	1.4×10^{-5}	0.100 (0.003)	7.9×10^{-6}	0.100 (0.004)	1.4×10^{-5}	0.100 (0.003)	7.8×10^{-6}
$E(u_{it}\|\Theta)^*$	0.700 (0.070)	0.035	0.695 (0.048)	0.033	0.660 (0.068)	0.034	0.654 (0.046)	0.032
corr*	0.954		0.955		0.947		0.948	

	N=300		N=300	
	Mean	MSE	Mean	MSE
$\hat{\beta}$	0.500 (0.005)	2.4×10^{-5}	0.500 (0.005)	2.4×10^{-5}
$\hat{\delta}$	0.500 (0.022)	4.6×10^{-4}	0.500 (0.023)	0.001
$\hat{\mu}$	0.497 (0.060)	0.004	0.497 (0.051)	0.003
$\hat{\sigma}_u^2$	0.203 (0.040)	0.002	0.152 (0.029)	0.001
$\hat{\sigma}_v^2$	0.100 (0.002)	5.0×10^{-6}	0.100 (0.002)	5.1×10^{-6}
$E(u_{it}\|\Theta)^*$	0.692 (0.041)	0.032	0.651 (0.040)	0.031
corr*	0.955		0.948	

* $\mathrm{E}(u_{it}|\Theta) = \mathrm{E}(u_{it}|\Delta\tilde{\varepsilon}_i)$ evaluated at $\Delta\tilde{\varepsilon}_i = \Delta\hat{\tilde{\varepsilon}}_i$ for the first-difference model. corr=corr$(E(u_{it}|\Theta), u_{it})$. Standard deviations are in the parenthesis.

Table 4: Larger σ_u^2

$\beta = 0.5$, $\delta = 0.5$, $\mu = 0.5$, $\sigma_v^2 = 0.1$

	$\sigma_u^2/\sigma_v^2 = 3$				$\sigma_u^2/\sigma_v^2 = 3$			
	N=100, T=5		N=100, T=15		N=300, T=5		N=300, T=15	
	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\hat{\beta}$	0.500 (0.017)	3.0×10^{-4}	0.500 (0.009)	7.7×10^{-5}	0.500 (0.010)	1.0×10^{-4}	0.500 (0.005)	2.4×10^{-5}
$\hat{\delta}$	0.501 (0.077)	0.006	0.499 (0.033)	0.001	0.499 (0.043)	0.002	0.500 (0.019)	3.7×10^{-4}
$\hat{\mu}$	0.480 (0.221)	0.049	0.499 (0.141)	0.020	0.496 (0.117)	0.014	0.494 (0.081)	0.007
$\hat{\sigma}_u^2$	0.343 (0.221)	0.051	0.308 (0.109)	0.012	0.312 (0.107)	0.012	0.304 (0.061)	0.004
$\hat{\sigma}_v^2$	0.099 (0.008)	6.6×10^{-5}	0.100 (0.004)	1.4×10^{-5}	0.100 (0.005)	2.2×10^{-5}	0.100 (0.002)	5.0×10^{-6}
$E(u_{it}\|\Theta)^*$	0.783 (0.139)	0.116	0.775 (0.073)	0.037	0.773 (0.075)	0.100	0.766 (0.045)	0.034
corr*	0.888		0.964		0.892		0.964	
	$\sigma_u^2/\sigma_v^2 = 4$				$\sigma_u^2/\sigma_v^2 = 4$			
	N=100, T=5		N=100, T=15		N=300, T=5		N=300, T=15	
	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
$\hat{\beta}$	0.500 (0.017)	3.0×10^{-4}	0.500 (0.009)	7.7×10^{-5}	0.500 (0.010)	1.0×10^{-4}	0.500 (0.005)	2.5×10^{-5}
$\hat{\delta}$	0.501 (0.072)	0.005	0.499 (0.030)	0.001	0.500 (0.040)	0.002	0.500 (0.018)	3.1×10^{-4}
$\hat{\mu}$	0.469 (0.272)	0.075	0.495 (0.172)	0.030	0.494 (0.139)	0.019	0.493 (0.098)	0.010
$\hat{\sigma}_u^2$	0.453 (0.282)	0.082	0.411 (0.143)	0.020	0.413 (0.134)	0.018	0.406 (0.080)	0.007
$\hat{\sigma}_v^2$	0.099 (0.008)	6.6×10^{-5}	0.100 (0.004)	1.5×10^{-5}	0.100 (0.005)	2.2×10^{-5}	0.100 (0.002)	5.0×10^{-6}
$E(u_{it}\|\Theta)^*$	0.849 (0.141)	0.125	0.842 (0.076)	0.038	0.839 (0.077)	0.109	0.834 (0.046)	0.035
corr*	0.901		0.970		0.904		0.970	

* $\mathrm{E}(u_{it}|\Theta) = \mathrm{E}(u_{it}|\Delta\tilde{\varepsilon}_i)$ evaluated at $\Delta\tilde{\varepsilon}_i = \Delta\hat{\tilde{\varepsilon}}_i$ for the first-difference model. corr=corr$(E(u_{it}|\Theta), u_{it})$. Standard deviations are in the parenthesis.

Table 5: Dummy Variable Model: A Comparison

	Model 1: N=100,T=5				Model 2: N=100,T=15			
	Dummy		first-difference		Dummy		first-difference	
	Mean (std.)	MSE	Mean (std.)	MSE	Mean (std.)	MSE	Mean (std.)	MSE
$\hat{\beta}$	0.498 (0.027)	0.001	0.500 (0.017)	3.0×10^{-4}	0.500 (0.010)	9.2×10^{-5}	0.500 (0.009)	7.6×10^{-5}
$\hat{\delta}$	0.810 (0.140)	0.115	0.502 (0.084)	0.007	0.627 (0.043)	0.018	0.499 (0.037)	0.001
$\hat{\mu}$	0.232 (0.383)	0.218	0.491 (0.180)	0.032	0.371 (0.120)	0.031	0.500 (0.106)	0.011
$\hat{\sigma}_u^2$	0.056 (0.092)	0.029	0.232 (0.163)	0.027	0.106 (0.066)	0.013	0.207 (0.073)	0.005
$\hat{\sigma}_v^2$	0.090 (0.011)	2.3×10^{-4}	0.099 (0.008)	6.4×10^{-5}	0.095 (0.005)	5.2×10^{-5}	0.100 (0.004)	1.4×10^{-5}
$E(u\|\Theta)^*$	0.449 (0.502)	1.331	0.709 (0.137)	0.104	0.545 (0.062)	0.192	0.700 (0.070)	0.035
corr*	0.711		0.871		0.726		0.954	
	Model 3: N=300,T=5				Model 4: N=300,T=15			
	Dummy		first-difference		Dummy		first-difference	
	Mean (std.)	MSE	Mean (std.)	MSE	Mean (std.)	MSE	Mean (std.)	MSE
$\hat{\beta}$	0.498 (0.012)	1.6×10^{-4}	0.500 (0.010)	9.9×10^{-5}	0.499 (0.006)	3.2×10^{-5}	0.500 (0.005)	2.4×10^{-5}
$\hat{\delta}$	0.823 (0.091)	0.113	0.499 (0.047)	0.002	0.626 (0.033)	0.017	0.500 (0.022)	4.6×10^{-4}
$\hat{\mu}$	0.255 (0.157)	0.085	0.498 (0.097)	0.009	0.371 (0.106)	0.028	0.497 (0.060)	0.004
$\hat{\sigma}_u^2$	0.041 (0.068)	0.029	0.211 (0.078)	0.006	0.106 (0.062)	0.013	0.203 (0.040)	0.002
$\hat{\sigma}_v^2$	0.090 (0.008)	1.7×10^{-4}	0.100 (0.005)	2.2×10^{-5}	0.095 (0.003)	4.0×10^{-5}	0.100 (0.002)	5.0×10^{-6}
$E(u\|\Theta)^*$	0.422 (0.078)	0.251	0.699 (0.073)	0.088	0.542 (0.052)	0.194	0.692 (0.041)	0.032
corr*	0.718		0.875		0.724		0.955	

* $E(u_{it}|\Theta) = E(u_{it}|\Delta\tilde{\varepsilon}_i)$ evaluated at $\Delta\tilde{\varepsilon}_i = \Delta\hat{\tilde{\varepsilon}}_i$ for the first-difference model, $E(u_{it}|\Theta) = E(u_{it}|\varepsilon_{it})$ evaluated at $\varepsilon_{it} = \hat{\varepsilon}_{it}$ for the dummy model. corr=corr$(E(u_{it}|\Theta), u_{it})$. Standard deviations are in the parenthesis.

Table 6: Summary Statistics

	mean	median	std dev
$\ln(I/K)_{it}$	-3.260	-3.057	1.548
$\ln Q_{it}$	0.797	0.642	0.799
$\ln(S/K)_{it}$	0.561	0.471	1.105
$(CF/K)_{it}$	0.294	0.136	1.578
$\ln Assets_{it}$	0.620	6.658	1.154

Note: Source: Taiwan Economic Journal Data Bank. The sample consists of data of 206 Taiwanese manufacturing firms from the year 2000 to the year 2005.

Table 7: Empirical Results

		Model 1 (dummy)		Model 2 (within)	
		coeff.	std. err.	coeff.	std. err.
frontier					
	$\ln Q_{it}$	0.684***	0.093	0.655***	0.108
	$\ln(S/K)_{it}$	0.142*	0.079	0.150*	0.090
	$\ln(S/K)_{it-1}$	0.346***	0.078	0.340***	0.089
constraints					
	$\ln(CF/K)_{it}$	-0.042	0.297	0.018	0.030
	$\ln Assets_{it}$	-1.845***	0.441	-0.652***	0.180
	c_v	-0.674***	0.094	-0.158***	0.045
	c_u	5.231***	1.196	8.523***	1.686
mean efficiency					
	$E(\exp(-u_{it})\|\Theta)$	0.639		0.573	

Note 1: Model 1 includes firm dummies in the estimation (i.e., the dummy variable model), and Model 2 is the within-transformation model. Estimates of dummy-variable coefficients of Model 1 are not reported to save space. $\Theta = \varepsilon_{it}$ for Model 1, and $\Theta = \tilde{\varepsilon}_{i\cdot}$ for Model 2. $c_v = \ln(\sigma_v^2)$, $c_u = \ln(\sigma_u^2)$.
Note 2: Significance: ***: 1% level, **: 5% level; *: 10% level.

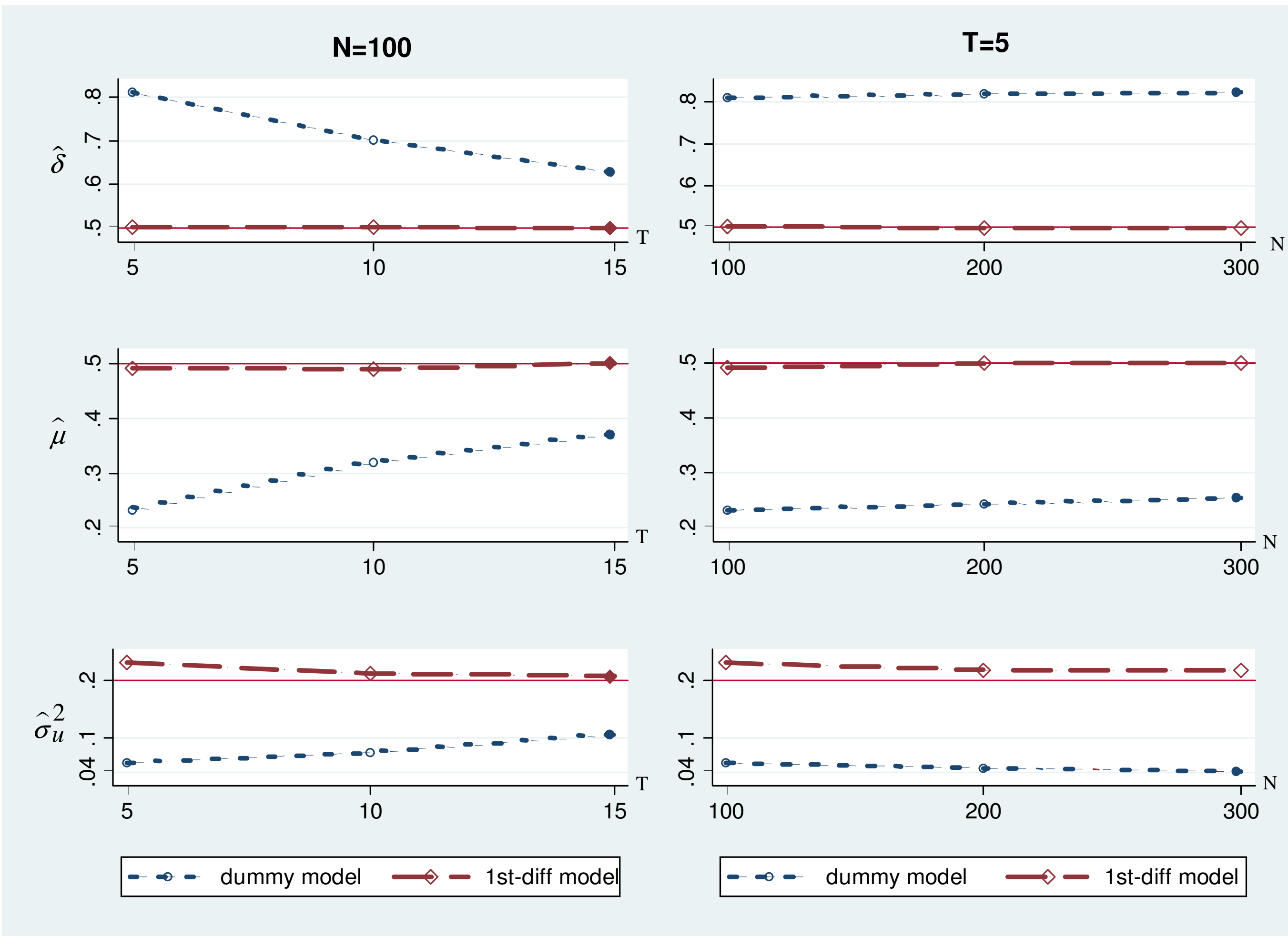

Figure 1: Model Comparisons