

行政院國家科學委員會專題研究計畫 成果報告

變數誤差下隨機邊界模型之估計：一般動差法

計畫類別：個別型計畫

計畫編號：NSC91-2415-H-032-011-

執行期間：91年08月01日至92年07月31日

執行單位：淡江大學經濟系(所)

計畫主持人：陳怡宜

報告類型：精簡報告

處理方式：本計畫可公開查詢

中 華 民 國 92 年 12 月 21 日

A Method of Moments Estimator for a Stochastic Frontier Model with Errors in Variables

Yi-Yi Chen (91-2415-H-032-011)

1 Introduction

We consider the estimation of a stochastic frontier (SF) model in which one of the independent variables is measured with errors. With the measurement error, a typical SF model of the type of Aigner et al. (1977) can be specified as

$$y_i = \beta_0 + \tilde{\mathbf{x}}_i \tilde{\boldsymbol{\beta}} + \gamma z_i^* + v_i - u_i, \quad (1)$$

$$z_i = z_i^* + e_i, \quad (2)$$

where y_i is a scalar, $\tilde{\mathbf{x}}_i$ is a $1 \times (k - 1)$ vector of perfectly-measured variables, z_i^* is the latent variable, z_i is the observed value of z_i^* , e_i is measurement error, and $v_i - u_i$ is the composed error of the model. It is often assumed that v_i is a zero-mean symmetric random variable, and u_i is also a random variable that takes on either non-negative values ($u_i \geq 0$) for a production-frontier type model, or non-positive values ($u_i \leq 0$) for a cost-frontier type model. In this paper we focus on the one where $u_i \geq 0$; the results for $u_i \leq 0$ will be a straightforward extension.

If z_i has to be used in (1), then the estimated model statistics such as returns to scale, price elasticities, and input substitution elasticities would be erroneous. Moreover, the measurement error also causes the frontier function of $\beta_0 + \tilde{\mathbf{x}}_i \tilde{\boldsymbol{\beta}} + \gamma z_i^*$ to be mis-measured, which in turn causes bias to the frequently-used inefficiency index $E(u_i | v_i - u_i)$ and $E(\exp(-u_i) | v_i - u_i)$ because the conditional information on $v_i - u_i$ is wrong (Kumbhakar 1991).

In the context of production and cost frontier models, measurement errors may arise from the use of self-reporting data, and the use of proxy variables when true ones are not available. Data aggregation can also introduce unduly errors to accounting data. Furthermore, as the

literature is seeing a growing number of applications of the SF model in different fields of economics, the problem of errors in variables needs to be taken seriously, since in many of the applications the problem is well-known and the consequence is critical. For instance, in the search-based wage equations of Hofler and Murphy (1992) and Polachek and Robst (1998), imperfect instruments for innate ability are often used. In the financing-constrained investment model of Wang (2003), the variable of Tobin's Q is also known to be prone to measurement errors.

In this paper we propose a method of moments estimator to obtain consistent estimates of SF model parameters when one of the independent variables is measured with errors. The estimator uses the information on the moments of the joint distribution of the observed variables, which is a non-trivial extension of Erickson and Whited (2002). Attractive features of this estimator include that no additional data are required, that minimum assumptions on the error (e) distribution are imposed, and that the computation is inexpensive.

In the remaining part of the paper, we derive the estimator and provide Monte Carlo evidence on its performance. We then apply the estimator to an empirical example of an investment model with financing constraint, which is similar to Wang (2003).

2 The Estimator

We make the following assumptions regarding the models of (1) and (2) in deriving the estimator: (i) the random variable v_i has a symmetric zero-mean distribution; (ii) the random variable $u_i \geq 0$ has a single-parameter density function (such as half-normal or exponential) (Kopp and Mullahy 1990); (iii) the regression residual from z_i^* on $\tilde{\mathbf{x}}_i$ is not symmetrically distributed; and (iv) the random variables v_i , u_i , and e_i are independently distributed to each other and to $\tilde{\mathbf{x}}_i$ and z_i^* . Note that we do not need to assume a particular distribution function for e_i .

We first subtract the mean of u_i , $\bar{u} = E(u_i)$, from the composed error, so that the latter has a zero mean.

$$y_i = (\beta_0 - \bar{u}) + \tilde{\mathbf{x}}_i \tilde{\boldsymbol{\beta}} + \gamma z_i^* + [v_i - (u_i - \bar{u})] \quad (3)$$

$$= \mathbf{x}_i \boldsymbol{\beta} + \gamma z_i^* + (v_i - \tilde{u}_i), \quad (4)$$

where $\mathbf{x}_i = (1, \tilde{\mathbf{x}}_i)$ is $1 \times k$, $\boldsymbol{\beta} = (\beta_0 - \bar{u}, \tilde{\boldsymbol{\beta}})'$, and $\tilde{u}_i = u_i - \bar{u}$. The remaining estimation procedure proceeds in two steps. In the first step, the perfectly-measured variables are partial out from the model, and moment equations are derived based on which estimates of γ and the parameters of v_i and u_i are obtained. The $\boldsymbol{\beta}$ coefficients are recovered in the second step.

We multiply $M = I_N - \mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'$, where \mathbf{x} is a stacked matrix of \mathbf{x}_i , to both sides of the stacked equation (4) and obtain

$$y_i - \mathbf{x}_i\mu_y = \gamma(z_i^* - \mathbf{x}_i\mu_z) + (v_i - \tilde{u}_i), \quad (5)$$

where $\mu_y = [E(\mathbf{x}'_i\mathbf{x}_i)]^{-1}E[\mathbf{x}'_iy_i]$ and $\mu_z = [E(\mathbf{x}'_i\mathbf{x}_i)]^{-1}E[\mathbf{x}'_iz_i]$. Subtracting $\mathbf{x}_i\mu_z$ from (2), we have

$$(z_i - \mathbf{x}_i\mu_z) = (z_i^* - \mathbf{x}_i\mu_z) + e_i. \quad (6)$$

The following moment equations are derived from the second, the third, and the product moments of (5) and (6):

$$E[(y_i - \mathbf{x}_i\mu_y)^2] = \gamma^2 E[(z_i^* - \mathbf{x}_i\mu_z)^2] + E(v_i^2) + \phi_2(\lambda), \quad (7a)$$

$$E[(y_i - \mathbf{x}_i\mu_y)(z_i - \mathbf{x}_i\mu_z)] = \gamma E[(z_i^* - \mathbf{x}_i\mu_z)^2], \quad (7b)$$

$$E[(z_i - \mathbf{x}_i\mu_z)^2] = E[(z_i^* - \mathbf{x}_i\mu_z)^2] + E(e_i^2), \quad (7c)$$

$$E[(y_i - \mathbf{x}_i\mu_y)^2(z_i - \mathbf{x}_i\mu_z)] = \gamma^2 E[(z_i^* - \mathbf{x}_i\mu_z)^3], \quad (7d)$$

$$E[(y_i - \mathbf{x}_i\mu_y)(z_i - \mathbf{x}_i\mu_z)^2] = \gamma E[(z_i^* - \mathbf{x}_i\mu_z)^3], \quad (7e)$$

$$E[(y_i - \mathbf{x}_i\mu_y)^3] = \gamma^3 E[(z_i^* - \mathbf{x}_i\mu_z)^3] - \phi_3(\lambda), \quad (7f)$$

where $\phi_2(\lambda) \equiv E(\tilde{u}_i^2)$, $\phi_3(\lambda) \equiv E(\tilde{u}_i^3)$, and λ is the single-parameter of the underlying distribution of u_i . For example, for half-normal ($f(u_i) = (2/\sqrt{2\pi}) \cdot \exp(-0.5(u_i^2/\lambda))$ for $u_i \geq 0$) and exponential ($f(u_i) = \lambda \cdot \exp(-\lambda u_i)$ for $u_i \geq 0$) distributions, the functions are

$$\text{half-normal: } \quad \phi_2(\lambda) = \left(1 - \frac{2}{\pi}\right) \lambda, \quad \phi_3(\lambda) = -\sqrt{\frac{2}{\pi}} \left(1 - \frac{4}{\pi}\right) \lambda^{3/2}; \quad (8)$$

$$\text{exponential: } \quad \phi_2(\lambda) = \frac{1}{\lambda^2}, \quad \phi_3(\lambda) = \frac{2}{\lambda^3}. \quad (9)$$

In the case of a half-normal distribution, λ is familiarly denoted as σ_u^2 , the variance of the underlying normal distribution. In the estimation, the sample counterparts of the population moments are used. For instance, $(1/N) \sum_i (y_i - \mathbf{x}_i\hat{\mu}_y)^2$ is used for $E[(y_i - \mathbf{x}_i\mu_y)^2]$ and $(\sum_i \mathbf{x}'_i\mathbf{x}_i)^{-1} \sum_i \mathbf{x}'_iy_i$ is used for $\hat{\mu}_y$.

The above system of the six moment equations from (7a) to (7f) contains six unknown parameters: γ , $\sigma_v^2 \equiv E(v_i^2)$, λ , $E[(z_i^* - \mathbf{x}_i\mu_z)^2]$, $E[(z_i^* - \mathbf{x}_i\mu_z)^3]$, and $E(e_i^2)$. We assume that their values exist in the population and are non-zero. To solve for their values, we divide (7d) by (7e) to obtain $\hat{\gamma}$, which in turn solves $\hat{E}[(z_i^* - \mathbf{x}_i\mu_z)^2]$ from (7b). The value of $\hat{E}(e_i^2)$ is then solved from (7c). Substituting $\hat{\gamma}$ into (7e) solves $\hat{E}[(z_i^* - \mathbf{x}_i\mu_z)^3]$. The value of $\hat{\lambda}$ is then obtained from (7f). Finally, $\hat{\sigma}_v^2$ is solved from (7a).

The estimated values of the $k \times 1$ vector $\boldsymbol{\beta}$ can be recovered as follows.

$$\begin{aligned}\mu_y &= [E(\mathbf{x}'_i \mathbf{x}_i)]^{-1} E[\mathbf{x}'_i y_i] = [E(\mathbf{x}'_i \mathbf{x}_i)]^{-1} E[\mathbf{x}'_i (\gamma z_i^* + \mathbf{x}_i \boldsymbol{\beta} + (v_i - \tilde{u}_i))] \\ &= [E(\mathbf{x}'_i \mathbf{x}_i)]^{-1} E[\mathbf{x}'_i z_i \gamma + \mathbf{x}'_i \mathbf{x}_i \boldsymbol{\beta}],\end{aligned}$$

$$\Rightarrow \quad \mu_y = \mu_z \gamma + \boldsymbol{\beta}, \quad (10)$$

$$\text{therefore,} \quad \hat{\boldsymbol{\beta}} = \hat{\mu}_y - \hat{\mu}_z \hat{\gamma}. \quad (11)$$

We used in the derivation the assumption that e_i and $v_i - \tilde{u}_i$ are independent of \mathbf{x}_i . Note that the first element of $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}}[1]$, is an estimated value of $\beta_0 - \bar{u}$ (see (3)). Hence, the model's intercept β_0 is estimated as

$$\hat{\beta}_0 = \hat{\boldsymbol{\beta}}[1] + \hat{u}. \quad (12)$$

Here, $\hat{u} = \hat{E}(u_i) = \sqrt{2\hat{\lambda}/\pi}$ if u_i has a half normal distribution, and $\hat{u} = 1/\hat{\lambda}$ if the distribution is exponential.

The variance covariance matrix of the parameters is obtained as follows. For easier reference, we denote the system of moment equations from (7a) to (7f) of observation i as $E[m_i(\boldsymbol{\mu})] = c_i(\boldsymbol{\delta})$, where $\boldsymbol{\mu} \equiv (\mu'_y, \mu'_z)'$, and $\boldsymbol{\delta} \equiv (\gamma, \lambda, \sigma_v^2, E[(z_i^* - \mathbf{x}_i \mu_z)^2], E[(z_i^* - \mathbf{x}_i \mu_z)^3], E(e_i^2))$. Therefore (see also Erickson and Whited 2002)

$$\text{avar}(\hat{\boldsymbol{\delta}}) = \frac{1}{N} [\hat{C}' \hat{\Omega}^{-1} \hat{C}]^{-1}, \quad (13)$$

where

$$\hat{C} = \frac{1}{N} \sum_{i=1}^N \left. \frac{\partial c_i(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}'} \right|_{\hat{\boldsymbol{\delta}}}, \quad (14)$$

$$\hat{\Omega} = \frac{1}{N} \sum_{i=1}^N (m_i(\hat{\boldsymbol{\mu}}) - \bar{m}(\hat{\boldsymbol{\mu}}) + \bar{G}(\hat{\boldsymbol{\mu}}) \hat{\Psi}_{\mu i}) (m_i(\hat{\boldsymbol{\mu}}) - \bar{m}(\hat{\boldsymbol{\mu}}) + \bar{G}(\hat{\boldsymbol{\mu}}) \hat{\Psi}_{\mu i})', \quad (15)$$

$$\bar{G}(\hat{\boldsymbol{\mu}}) = \frac{1}{N} \sum_{i=1}^N \left. \frac{\partial m_i(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}'} \right|_{\hat{\boldsymbol{\mu}}}, \quad (16)$$

$$\hat{\Psi}_{\mu i} = \left(I_2 \otimes \frac{1}{N} \sum_{i=1}^N \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \times \text{vec}(\mathbf{x}'_i (y_i - \mathbf{x}_i \hat{\mu}_y), \mathbf{x}'_i (z_i - \mathbf{x}_i \hat{\mu}_z)). \quad (17)$$

For the variances of $\boldsymbol{\beta}$ (including the intercept), Erickson and Whited (2002) show that they can be estimated using the delta method based on (10). To proceed, let $\tilde{R} = \mu_y - \mu_z \gamma - \boldsymbol{\beta}$, which is a $k \times 1$ vector of (10), and we define $R = \tilde{R} + (\bar{u}, 0, \dots, 0)'$ and $\check{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\bar{u}, 0, \dots, 0)'$, where the second vector on the right-hand-side of the equal signs is $k \times 1$. This vector is added to \tilde{R} and $\boldsymbol{\beta}$ for the intercept adjustment (see (12)). We further define $\boldsymbol{\theta} = (\gamma, \lambda)'$ and

$T = (\boldsymbol{\mu}', \boldsymbol{\theta}')$. With these notations, the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\text{avar}(\hat{\boldsymbol{\beta}}) = \frac{1}{N} \left[\hat{C}'_0 \hat{\Omega}_0^{-1} \hat{C}_0 \right]^{-1}, \quad (18)$$

where

$$\hat{C}_0 = \frac{1}{N} \sum_{i=1}^N \left. \frac{\partial R}{\partial \tilde{\boldsymbol{\beta}}'} \right|_{\hat{\boldsymbol{\beta}}} = -1, \quad (19)$$

$$\hat{\Omega}_0 = \frac{1}{N} \sum_{i=1}^N (\bar{G}_0 \hat{\Psi}_{Ti}) (\bar{G}_0 \hat{\Psi}_{Ti})', \quad (20)$$

$$\bar{G}_0 = \frac{1}{N} \sum_{i=1}^N \left. \frac{\partial R}{\partial T'} \right|_{\hat{T}}, \quad (21)$$

$$\hat{\Psi}_{Ti} = (\hat{\Psi}'_{\mu i}, \hat{\Psi}_{\theta i}[1], \hat{\Psi}_{\theta i}[2])', \quad (22)$$

$$\hat{\Psi}_{\theta i} = -N \cdot \text{avar}(\hat{\boldsymbol{\delta}})^{-1} \cdot \hat{C} \cdot (m_i(\hat{\boldsymbol{\mu}}) - \bar{m}(\hat{\boldsymbol{\mu}}) + \bar{G}(\hat{\boldsymbol{\mu}}) \hat{\Psi}_{\mu i}). \quad (23)$$

The estimator is programmed using Stata 7.0 software.

A Monte Carlo experiment, designed similar to that of Erickson and Whited (2002), is conducted to show the estimator's performance. We consider the case in which there is one perfectly-measured variable x_i and one latent variable z_i^* measured by z_i . The base case has the following parameter values: $\{\beta_0 = 0.5, \beta_1 = 0.5, \gamma = 0.5\}$. To generate $\{x_i, z_i^*, e_i\}$, we first create three independent zero-mean normal variables with the variances equal to $\{\sigma_x^2 = 1, \sigma_{z^*}^2 = 1, \sigma_e^2 = 2\}$, respectively, and then the variables are exponentiated and standardized. The term u_i is created as a half-normal variable with the pre-truncation variance equal to $\sigma_u^2 = 2$. The term v_i is generated from a zero-mean normal distribution with the variance equal to $\sigma_v^2 = 1$.

We also examine three additional cases with a smaller value of σ_e^2 , a smaller value of σ_u^2 , and non-zero correlation between z_i^* and x_i . For each model parameter ϑ , we report the mean, the mean absolute error (MAE), and the probability (prob.) that $p(|\hat{\vartheta} - \vartheta| \leq 0.15 | \vartheta |)$. The latter measures the probability that the estimated value is within 15% above or below that of the true value. The results are based on 1,000 Monte Carlo replications with the sample size equal to 1,000 in each of the replications. The results are in Table 1.

The results indicate that the MoM estimates are indeed much better than the ML estimates. The ML estimates of the intercept are significantly biased upward while those of γ are significantly biased downward. The MoM estimates are much closer to the true values. We find that if z_i^* and x_i are not correlated, then the estimation of β_1 (coefficient of the perfectly-measured variable) is not affected by the model's measurement error. When the

correlation is non-zero, the MoM estimate of β_1 is much better. The results also show that the ML estimates of σ_u^2 are upward biased, implying a possible overestimation of the inefficiency effect.

3 Empirical Example

We apply the estimator to the estimation of an investment model with financing constraints, in which Tobin's Q is an important explanatory variable. Wang (2003) shows that when there is a financing constraint, a firm's actual investment falls below the neo-classical investment level (i.e., the *frontier*). The actual investment can thus be modeled as a one-sided deviation from the neo-classical investment frontier, with the deviation representing the financing constraint effect and captured by the u_i term. On the other hand, the measurement error problem of the Q variable has long been recognized in the literature, because the measured Q may not be a good approximation of the theoretical marginal Q ,

Annual data of 184 publicly traded manufacturing firms in Taiwan from 1990 to 1996 are used. The total number of observations is 1,036. This is the same data used by Wang (2003). The dependent variable $\ln(I/K)$ is the log of investment to capital ratio, and the independent variables include Q , $\ln(S/K)$, $\ln(S/K)$, and six year dummies, where Q is Tobin's Q and S is sales.

The estimation results from the method of moments (MoM), maximum likelihood estimator (MLE), and the ordinary least square (OLS) are presented in Table 2. For MLE and OLS, the measurement errors are not accounted for. To save space, estimates on the six time dummies are not reported. The result from MoM indicates that a marginal increase in Q entails a 30% increase in the investment rate. The estimate is more than twice as large as that obtained by MLE and OLS, which are 12% and 11%, respectively. Both of the latter two figures appear to be unreasonably small. On the other hand, judging from $E(\hat{\sigma}_u^2)$ of MoM and MLE, the financing constraint effect on investment does not appear to be sensitive to the measurement error of Tobin's Q in this particular application.

4 Conclusion

In this paper we propose a method of moments estimator for an SF model with errors in the variables. The errors in variable problem of an SF model is an important issue particularly when the model has been applied to many different fields in economics where the problem is known and critical. The Monte Carlo results show that the proposed estimator indeed

performs quite well. Our empirical example also indicates that the MoM estimator yields more reasonable parameter estimates.

References

- [1] Aigner, Dennis, Lovell, C. A. Knox, and Schmidt, Peter. (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics* **6**, pp. 21-37.
- [2] Erickson, Timothy , and Whited, Toni M. (2002). "Two-Step Gmm Estimation of the Errors-In-Variabes Model using High-Order Moments," *Econometric Theory* **18**, pp. 776-799.
- [3] Hofer, Richard A., and Murphy, Kevin J. (1992). "Underpaid and Overworked: Measuring the Effect of Imperfect Information on Wages," *Economic Inquiry* **30**, pp. 511-29.
- [4] Kopp, Raymond J., and Mullahy, John. (1990). "Moment-Based Estimation and Testing of Stochastic Frontier Models," *Journal of Econometrics* **46**, pp. 165-83.
- [5] Kumbhakar, Subal C. (1991). "Estimation of Technical Inefficiency in Panel Data Models with Firm- and Time-Specific Effects," *Economics Letters* **36**, pp. 43-48.
- [6] Polachek, Solomon W., and Robst, John. (1998). "Employee Labor Market Information: Comparing Direct World of Work Measures of Workers' Knowledge to Stochastic Frontier Estimates," *Labour Economics* **5**, pp. 231-42.
- [7] Wang, Hung-Jen. (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.

Table 1: Monte Carlo

	base						change: $\sigma_e^2 = 1$					
	MoM			MLE			MoM			MLE		
	mean	MAE	prob.	mean	MAE	prob.	mean	MAE	prob.	mean	MAE	prob.
β_0	0.494	0.145	0.341	0.784	0.286	0.036	0.493	0.139	0.368	0.751	0.254	0.074
β_1	0.500	0.017	1.000	0.499	0.016	1.000	0.500	0.017	1.000	0.500	0.016	1.000
γ	0.501	0.027	0.956	0.433	0.067	0.629	0.500	0.024	0.988	0.426	0.074	0.517
σ_u^2	2.039	0.492	0.378	2.693	0.715	0.157	2.025	0.456	0.404	2.509	0.573	0.254
σ_v^2	0.984	0.183	0.501	0.947	0.117	0.685	0.990	0.164	0.542	1.017	0.114	0.723
$E(u_i)$	1.124	0.143	0.679	1.305	0.184	0.428	1.122	0.132	0.713	1.259	0.149	0.600

	change: $\sigma_u^2 = 1$						change: $\text{corr}(z^*, x) = 0.37^\dagger$					
	MoM			MLE			MoM			MLE		
	mean	MAE	prob.	mean	MAE	prob.	mean	MAE	prob.	mean	MAE	prob.
β_0	0.527	0.169	0.250	0.912	0.412	0.000	0.498	0.132	0.390	0.523	0.115	0.402
β_1	0.500	0.015	1.000	0.500	0.015	1.000	0.493	0.071	0.672	0.593	0.094	0.365
γ	0.503	0.024	0.979	0.431	0.069	0.608	0.515	0.148	0.367	0.265	0.235	0.000
σ_u^2	1.152	0.434	0.209	1.912	0.912	0.005	2.039	0.450	0.435	2.117	0.396	0.462
σ_v^2	0.939	0.166	0.526	0.869	0.143	0.551	0.982	0.184	0.538	1.071	0.137	0.649
$E(u_i)$	0.831	0.170	0.398	1.100	0.302	0.026	1.127	0.129	0.729	1.152	0.111	0.790

Note: MAE: mean absolute error; prob.: $p(|\hat{\vartheta} - \vartheta| \leq 0.15 | \vartheta)$.

†: Two random variables are generated from a bivariate normal with correlation coefficient= 0.5, and then they are exponentiated and standardized to obtain z_i^* and x_i , for which the correlation coefficient is about 0.37.

Table 2: Investment Model with Financing Constraints

	MoM		MLE		OLS	
	coef.	std.err.	coef.	std.err.	coef.	std.err.
const.	-2.844***	0.562	-2.046***	0.142	-3.068***	0.141
Q	0.298***	0.114	0.118***	0.019	0.105***	0.020
$\ln(S/K)$	-0.759***	0.235	-0.649***	0.147	-0.515***	0.154
$\ln(S/K)_{-1}$	1.189***	0.190	1.121***	0.147	1.096***	0.154
σ_u^2	1.867***	0.270	1.785***	0.203	–	–
σ_v^2	0.366***	0.101	0.447***	0.057	–	–

Note: Dependent variable: $\ln(I/K)$. Significance: ***, 1% level.