

# A Transformed Random Effects Model with Applications<sup>1</sup>

Zhenlin Yang<sup>a</sup> and Jianhua Huang<sup>b</sup>

<sup>a</sup>*School of Economics, Singapore Management University*

email: `zlyang@smu.edu.sg`

<sup>b</sup>*Department of Statistics, Texas A & M University*

email: `jianhua@stat.tamu.edu`

December, 2004

## Abstract

This paper proposes a transformed random effects model for analyzing non-normal panel data where both the response and (some of) the covariates are subject to transformations for inducing flexible functional form, normality, homoscedasticity and simple model structure. We develop maximum likelihood procedure for model estimation and inference, along with a computational devise which makes the estimation procedure feasible in cases of large panels. We give model specification tests which take into account the fact that parameter values for error components cannot be negative. We illustrate the model and methods with two applications: state production and wage distribution. The empirical results strongly favor the new model to the standard ones where either linear or log-linear functional form is employed. Monte Carlo simulation shows that maximum likelihood inference is quite robust against mild departure from normality.

**Key Words:** Computational devise; Flexible functional form; Maximum likelihood estimation; One-sided LM tests; Robustness.

**JEL Classification:** C23, C51

---

<sup>1</sup>Zhenlin Yang gratefully acknowledges the support from a research grant (Grant number: C208/MS63E046) from Singapore Management University. Jianhua Huang is grateful to the Office of Research and the School of Economics, Singapore Management University, for their hospitality during his visit.

## 1. Introduction.

With the increasing availability of richer panel data sets, panel data regression models have become increasingly popular among the applied researchers due to their capabilities and flexibilities in dealing with complex issues in economic modeling. However, little has been done on the choices of functional form in panel data set-up as evidenced by the recent monographs (see, e.g., Baltagi, 2001; Arellano, 2003; Hsiao, 2003; Frees, 2004). This is in contrast to the cross-sectional or time-series studies where standard econometrics text books treat the choice of functional form as a standard topic (see, e.g., Davidson and MacKinnon, 1993; Greene, 2000).

It is well known that the purposes of transforming the economic data in a cross-sectional or time-series regression is to induce normality, flexible functional form, homoscedastic errors, and simple model structure. The same ideas can be applied to panel data. However, panel data has its own unique features such as unobservable individual and time effects that may not disappear even if transformations are applied to the data. We thus consider the model,

$$h(Y_{it}, \lambda) = \sum_{j=1}^{k_1} \beta_j Z_{itj} + \sum_{j=k_1+1}^k \beta_j h(X_{itj}, \lambda) + u_{it}, \quad (1)$$

where  $i = 1, 2, \dots, N$ ,  $t = 1, 2, \dots, T$ ,  $h(\cdot, \lambda)$  is a monotonic transformation (e.g., Box and Cox, 1964), known except the indexing parameter  $\lambda$ , called the *transformation parameter*. The  $Z$  variables contain the column of ones, dummy variables, etc., and the  $X$  variables are some of the continuous type of variables that need to be transformed. We assume the transformed observations follow a two-way random effects model, i.e.,

$$u_{it} = \mu_i + \eta_t + v_{it} \quad (2)$$

where  $\mu_i \stackrel{iid}{\sim} N(0, \sigma_\mu^2)$  (independent and identically distributed normal random variates with means zero and variances  $\sigma_\mu^2$ ),  $\eta_t \stackrel{iid}{\sim} N(0, \sigma_\eta^2)$ , and  $v_{it} \stackrel{iid}{\sim} N(0, \sigma_v^2)$ . The  $\mu_i$ ,  $\eta_t$  and  $v_{it}$  characterize, respectively, the unobservable individual effects, unobservable time effects and the model errors, which are assumed to be independent of each other.

The model specified by (1) and (2) clearly gives a useful extension of the standard random effects model by allowing the distribution of  $Y_{it}$  to be in a broad family (transformed normal family) not just normal or lognormal. It also allows easy testing of the traditional economic theories of lognormality for production function, firm-size distribution, income distribution, etc., as governed by Cobb-Douglas production function and Gibrat's Law. Baltagi (1997) considered a submodel with only individual random effects and presented a Lagrange multiplier test for testing whether the functional form is linear or log-linear. Breusch and Pagan (1980), Breusch (1987), Baltagi and Li (1992), among others, have treated the random effects model (one or two-way) without transformation.<sup>2</sup>

Despite of various works in the static panel models with or without parametric transformations on the model variables, there is a need of simple and easily implementable procedures for estimating the full transformation model, especially in the cases of large panels. The results presented in this paper seem to have fulfilled such a need, making the applied research work much easier. Also, it is desirable to have tests for functional form (not only linear or log-linear) or error components allowing for the presence of the other in such a general framework. Furthermore, parameter values for error components cannot be negative, and hence one-sided LM tests for model specification are desirable. On computing side, estimation of the random effects model involve the calculation of many  $NT \times NT$  matrices which can easily exhaust the computer memory when the panel ( $NT$ ) is large. A method is given to overcome this computational difficulty.

This paper is organized as follows. Section 2 develops the maximum likelihood estimation procedure for the model. Also in this section, a computational devise is presented which avoids the calculations of all the  $NT \times NT$  matrices, allowing for the estimation procedure to be implementable for large panel data sets. Section 3

---

<sup>2</sup>Some related works on the choice of functional form are Abrevaya (1999) who proposed a leapfrog estimation of a fixed-effects model with an unknown response transformation and Giannakas *et al.* (2003) who considered the choice of functional form in stochastic frontier model using panel data.

presents model specification tests which take into account the one-sided nature for the error components. Section 4 presents two empirical applications using a state production data and a wage data. Section 5 presents Monte Carlo results for the finite sample performance of parameter estimates and inferences. Section 5 concludes the paper. The empirical results strongly favor the new model to the standard ones where either linear or log-linear functional form is employed. Monte Carlo simulation shows that maximum likelihood inference is quite robust against mild departure from normality.

## 2. Model Estimation

For the regular error components model, the standard method of estimation for the regression coefficients may be the feasible GLS, which takes the advantage of the block diagonality between the estimates of regression coefficients and the estimates of variance components. This simple method, however, cannot be applied to the transformed panel regression model due to the fact that the estimates of the transformation parameter and the regression coefficients are highly correlated. Another commonly used method, the instrumental variable (IV) method (or GMM in general), is also not applicable to our model as the response involves an unknown parametric transformation (Davidson and MacKinnon, 1993, p. 243). We thus, turn to the maximum likelihood estimation (MLE) method.

### 2.1 Maximum likelihood estimation

Stacking the data into columns with  $i$  and  $t$  being, respectively, the slower and faster running indices, the above model can be compactly written in matrix form,

$$h(Y, \lambda) = X(\lambda)\beta + u \quad (3)$$

$$u = Z_\mu\mu + Z_\eta\eta + v \quad (4)$$

where  $Z_\mu = I_N \otimes 1_T$  and  $Z_\eta = 1_N \otimes I_T$  with  $I_n$  being an  $n \times n$  identity matrix,  $1_n$  an  $n$ -vector of ones, and  $\otimes$  the Kroneker product. Define  $J_n = 1_n 1_n'$ . The log likelihood

function after dropping the constant term takes the form

$$\ell(\beta, \sigma_\mu^2, \sigma_\eta^2, \sigma_v^2, \lambda) = -\frac{1}{2} \log |\Sigma| - \frac{1}{2} [h(y, \lambda) - X(\lambda)\beta]' \Sigma^{-1} [h(y, \lambda) - X(\lambda)\beta] + J(\lambda), \quad (5)$$

where  $J(\lambda) = \sum_{i=1}^N \sum_{t=1}^T \log h_{y_{it}}(\lambda)$  is the log Jacobian of the transformation, and  $\Sigma$  is the variance-covariance matrix of  $u$  which takes the form

$$\Sigma = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\eta^2(J_N \otimes I_T) + \sigma_v^2(I_N \otimes I_T).$$

Model estimation corresponds to the maximization of the log likelihood function (5). Clearly, the addition of transformation in the model makes parameter estimation a very challenging problem. Direct maximization of (5) may be impractical and method of simplification should be sought after. Following Baltagi and Li (1992), among others, we consider a reparameterization of the model and a spectral decomposition of  $\Sigma$  to simplify the model estimation process. Define  $Q = I_{NT} - \frac{1}{T}I_N \otimes J_T - \frac{1}{N}J_N \otimes I_T + \frac{1}{NT}J_{NT}$ ,  $P_1 = \frac{1}{T}I_N \otimes J_T - \frac{1}{NT}J_{NT}$ ,  $P_2 = \frac{1}{N}J_N \otimes I_T - \frac{1}{NT}J_{NT}$ , and  $P_3 = \frac{1}{NT}J_{NT}$ . Let  $\theta_1 = \sigma_v^2/(T\sigma_\mu^2 + \sigma_v^2)$ ,  $\theta_2 = \sigma_v^2/(N\sigma_\eta^2 + \sigma_v^2)$ , and  $\theta_3 = \sigma_v^2/(T\sigma_\mu^2 + N\sigma_\eta^2 + \sigma_v^2)$ , where  $\theta_3 = \theta_1\theta_2/(\theta_1 + \theta_2 - \theta_1\theta_2)$ . We have

$$\begin{aligned} \Sigma &= \sigma_v^2 \Omega, \text{ with } \Omega = Q + \frac{1}{\theta_1}P_1 + \frac{1}{\theta_2}P_2 + \frac{1}{\theta_3}P_3, \\ \Sigma^{-1} &= \sigma_v^{-2} \Omega^{-1}, \text{ with } \Omega^{-1} = Q + \theta_1 P_1 + \theta_2 P_2 + \theta_3 P_3, \\ |\Sigma|^{-1} &= (\sigma_v^2)^{-NT} \theta_1^{N-1} \theta_2^{T-1} \theta_3. \end{aligned}$$

It should be emphasized that the availability of the analytical inverse and determinant for the  $NT \times NT$  matrix  $\Omega$  greatly simplifies the computational process as direct calculations of  $|\Omega|$  and  $\Omega^{-1}$  can be extremely time consuming for large  $NT$  which is often the case for economic panel data. With these analytical expressions, the log likelihood (5) can be rewritten as,

$$\begin{aligned} \ell(\beta, \sigma_v^2, \theta_1, \theta_2, \lambda) &= -\frac{NT}{2} \log \sigma_v^2 + \frac{N}{2} \log \theta_1 + \frac{T}{2} \log \theta_2 - \frac{1}{2} \log(\theta_1 + \theta_2 - \theta_1\theta_2) \\ &\quad - \frac{1}{2\sigma_v^2} [h(y, \lambda) - X(\lambda)\beta]' \Omega^{-1} [h(y, \lambda) - X(\lambda)\beta] + J(\lambda). \end{aligned} \quad (6)$$

Note that  $\sigma_\mu^2 > 0$ ,  $0 \leq \sigma_\mu^2 < \infty$  and  $0 \leq \sigma_\eta^2 < \infty$ . Thus,  $0 < \theta_1 \leq 1$  and  $0 < \theta_2 \leq 1$ .

Further simplification can be done by concentrating out the parameters  $\beta$  and  $\sigma_v^2$  in the log likelihood function, thus considerably reducing the dimension of maximization. To simplify the notation, we define  $e = [h(y, \lambda) - X(\lambda)\beta]$  and  $e_\lambda = \partial e / \partial \lambda$ . The score function  $S(\beta, \sigma_v^2, \theta_1, \theta_2, \lambda)$  has the elements,

$$\begin{aligned} S_\beta = \frac{\partial \ell}{\partial \beta} &= \frac{1}{\sigma_v^2} X'(\lambda) \Omega^{-1} e \\ S_{\sigma_v^2} = \frac{\partial \ell}{\partial \sigma_v^2} &= \frac{1}{2\sigma_v^4} e' \Omega^{-1} e - \frac{NT}{2\sigma_v^2} \\ S_{\theta_1} = \frac{\partial \ell}{\partial \theta_1} &= \frac{N}{2\theta_1} - \frac{1 - \theta_2}{2(\theta_1 + \theta_2 - \theta_1 \theta_2)} - \frac{1}{2\sigma_v^2} e' [P_1 + (\partial \theta_3 / \partial \theta_1) P_3] e \\ S_{\theta_2} = \frac{\partial \ell}{\partial \theta_2} &= \frac{T}{2\theta_2} - \frac{1 - \theta_1}{2(\theta_1 + \theta_2 - \theta_1 \theta_2)} - \frac{1}{2\sigma_v^2} e' [P_2 + (\partial \theta_3 / \partial \theta_2) P_3] e \\ S_\lambda = \frac{\partial \ell}{\partial \lambda} &= J_\lambda(\lambda) - \frac{1}{\sigma_v^2} e'_\lambda \Omega^{-1} e \end{aligned}$$

Given  $\theta_1$ ,  $\theta_2$  and  $\lambda$ ,  $\ell$  is maximized at

$$\begin{aligned} \hat{\beta}(\theta_1, \theta_2, \lambda) &= [X'(\lambda) \Omega^{-1} X(\lambda)]^{-1} X'(\lambda) \Omega^{-1} h(Y, \lambda), \\ \hat{\sigma}_v^2(\theta_1, \theta_2, \lambda) &= \frac{1}{NT} \hat{e}' \Omega^{-1} \hat{e}, \end{aligned}$$

where  $\hat{e}$  is  $e$  with  $\beta$  replaced by  $\hat{\beta}(\theta_1, \theta_2, \lambda)$ . Similarly, we define  $\hat{e}_\lambda$ . Substituting  $\hat{\beta}(\theta_1, \theta_2, \lambda)$  and  $\hat{\sigma}_v^2(\theta_1, \theta_2, \lambda)$  into the log likelihood function (6) for  $\beta$  and  $\sigma_v^2$ , respectively, gives the concentrated log likelihood

$$\begin{aligned} \ell_{\max}(\theta_1, \theta_2, \lambda) &= \frac{N}{2} \log \theta_1 + \frac{T}{2} \log \theta_2 - \frac{1}{2} \log(\theta_1 + \theta_2 - \theta_1 \theta_2) \\ &\quad - \frac{NT}{2} \log \hat{e}'(Q + \theta_1 P_1 + \theta_2 P_2) \hat{e} + J(\lambda). \end{aligned} \quad (7)$$

Maximizing  $\ell_{\max}(\theta_1, \theta_2, \lambda)$ , subject to  $0 < \theta_1 \leq 1$  and  $1 < \theta_2 \leq 1$ , gives the MLEs  $\hat{\theta}_1$ ,  $\hat{\theta}_2$  and  $\hat{\lambda}$ , which upon substitution gives the MLEs  $\hat{\beta} = \hat{\beta}(\hat{\theta}_1, \hat{\theta}_2, \hat{\lambda})$  and  $\hat{\sigma}_v^2 = \hat{\sigma}_v^2(\hat{\theta}_1, \hat{\theta}_2, \hat{\lambda})$  for  $\beta$  and  $\sigma_v^2$ , respectively. Further, the MLEs of  $\sigma_\mu^2$  and  $\sigma_\eta^2$  can be obtained through the relations:  $\sigma_\mu^2 = \frac{1}{T} \sigma_v^2 (\frac{1}{\theta_1} - 1)$  and  $\sigma_\eta^2 = \frac{1}{N} \sigma_v^2 (\frac{1}{\theta_2} - 1)$ .

Maximization of  $\ell_{\max}$  can be further facilitated by providing the analytical gradients. Substituting  $\hat{\beta}(\theta_1, \theta_2, \lambda)$  and  $\hat{\sigma}_v^2(\theta_1, \theta_2, \lambda)$  into the last three elements of the score function and simplifying give the concentrated scores for  $\theta_1$ ,  $\theta_2$  and  $\lambda$ :

$$S_{\theta_1}(\theta_1, \theta_2, \lambda) = \frac{1}{2} \left( \frac{N}{\theta_1} - \frac{1 - \theta_2}{\theta_1 + \theta_2 - \theta_1 \theta_2} - \frac{NT \hat{e}' P_1 \hat{e}}{\hat{e}'(Q + \theta_1 P_1 + \theta_2 P_2) \hat{e}} \right) \quad (8)$$

$$S_{\theta_2}(\theta_1, \theta_2, \lambda) = \frac{1}{2} \left( \frac{T}{\theta_2} - \frac{1 - \theta_1}{\theta_1 + \theta_2 - \theta_1 \theta_2} - \frac{NT\hat{e}'P_2\hat{e}}{\hat{e}'(Q + \theta_1 P_1 + \theta_2 P_2)\hat{e}} \right) \quad (9)$$

$$S_{\lambda}(\theta_1, \theta_2, \lambda) = J_{\lambda}(\lambda) - \frac{NT\hat{e}'_{\lambda}(Q + \theta_1 P_1 + \theta_2 P_2)\hat{e}}{\hat{e}'(Q + \theta_1 P_1 + \theta_2 P_2)\hat{e}} \quad (10)$$

Note that in the above derivation, we have used the result  $P_3\hat{e} = \frac{1}{\theta_3}P_3\Omega^{-1}\hat{e} = 0$ . It can be shown that these concentrated scores are also the partial derivatives (gradients) of  $\ell_{\max}(\theta_1, \theta_2, \lambda)$  with respect to  $\theta_1$ ,  $\theta_2$  and  $\lambda$ , respectively. Thus, the maximization of  $\ell_{\max}(\theta_1, \theta_2, \lambda)$  can be made more efficient with the use of these analytical gradients. Furthermore, those analytical gradients can be used to derive various LM tests for model specification, either jointly or individually. More discussions on this issue are given in the next section.

Under standard regularity conditions, the MLEs given above are consistent and asymptotically normal under the framework that  $T$  is fixed and  $N$  goes to infinity. See Hsiao (2003, p.41) for the case of regular two-way error components model. One of the regularity condition is that the true parameter values are an interior point of the parameter space. If these condition holds, one can simply estimate the standard errors of the parameter estimates using the negative inverse of the estimated Hessian matrix  $H(\hat{\beta}, \hat{\sigma}_v^2, \hat{\theta}_1, \hat{\theta}_2, \hat{\lambda})$ , where the elements of  $H(\beta, \sigma_v^2, \theta_1, \theta_2, \lambda)$  are given below,

$$\begin{aligned} H_{\beta\beta} &= -\frac{1}{\sigma_v^2}X'(\lambda)\Omega^{-1}X(\lambda) \\ H_{\beta\sigma_v^2} &= -\frac{1}{\sigma_v^4}X'(\lambda)\Omega^{-1}e \\ H_{\beta\theta_1} &= \frac{1}{\sigma_v^2}X'(\lambda)(P_1 + \frac{\partial\theta_3}{\partial\theta_1}P_3)e \\ H_{\beta\theta_2} &= \frac{1}{\sigma_v^2}X'(\lambda)(P_2 + \frac{\partial\theta_3}{\partial\theta_2}P_3)e \\ H_{\beta\lambda} &= \frac{1}{\sigma_v^2}[X'_{\lambda}(\lambda)\Omega^{-1}e + X'(\lambda)\Omega^{-1}e_{\lambda}] \\ H_{\sigma_v^2\sigma_v^2} &= \frac{NT}{2\sigma_v^4} - \frac{1}{\sigma_v^6}e'\Omega^{-1}e \\ H_{\sigma_v^2\theta_1} &= \frac{1}{2\sigma_v^4}e'(P_1 + \frac{\partial\theta_3}{\partial\theta_1}P_3)e \\ H_{\sigma_v^2\theta_2} &= \frac{1}{2\sigma_v^4}e'(P_2 + \frac{\partial\theta_3}{\partial\theta_2}P_3)e \end{aligned}$$

$$\begin{aligned}
H_{\sigma_v^2 \lambda} &= \frac{1}{\sigma_v^4} e'_\lambda \Omega^{-1} e \\
H_{\theta_1 \theta_1} &= -\frac{N}{2\theta_1^2} + \frac{(1 - \theta_2)^2}{2(\theta_1 + \theta_2 - \theta_1 \theta_2)^2} - \frac{1}{2\sigma_v^2} e' P_3 e \left( \frac{\partial^2 \theta_3}{\partial \theta_1^2} \right) \\
H_{\theta_1 \theta_2} &= \frac{1}{2(\theta_1 + \theta_2 - \theta_1 \theta_2)^2} - \frac{1}{2\sigma_v^2} e' P_3 e \left( \frac{\partial^2 \theta_3}{\partial \theta_1 \partial \theta_2} \right) \\
H_{\theta_1 \lambda} &= -\frac{1}{\sigma_v^2} e'_\lambda \left( P_1 + \frac{\partial \theta_3}{\partial \theta_1} P_3 \right) e \\
H_{\theta_2 \theta_2} &= -\frac{T}{2\theta_2^2} + \frac{(1 - \theta_1)^2}{2(\theta_1 + \theta_2 - \theta_1 \theta_2)^2} - \frac{1}{2\sigma_v^2} e' P_3 e \left( \frac{\partial^2 \theta_3}{\partial \theta_2^2} \right) \\
H_{\theta_2 \lambda} &= -\frac{1}{\sigma_v^2} e'_\lambda \left( P_2 + \frac{\partial \theta_3}{\partial \theta_2} P_3 \right) e \\
H_{\lambda \lambda} &= -\frac{1}{\sigma_v^2} (e'_{\lambda \lambda} \Omega^{-1} e + e'_\lambda \Omega^{-1} e_\lambda) + J_{\lambda \lambda}(\lambda).
\end{aligned}$$

Note that when evaluating the  $H$ -quantities above at the MLE of  $\beta$ , constrained given  $(\theta_1, \theta_2, \lambda)$ , or unconstrained, all the terms involving  $P_3 e$  vanish. For the Box-Cox power transformation,  $J_{\lambda \lambda}(\lambda) = 0$ .

Often, it is desirable to have the covariance estimate in the original parameterization. The Hessian under the original parameterization can be calculated using

$$H(\beta, \sigma_v^2, \sigma_\mu^2, \sigma_\eta^2, \lambda) = C(\beta, \sigma_v^2, \sigma_\mu^2, \sigma_\eta^2, \lambda) H(\beta, \sigma_v^2, \theta_1, \theta_2, \lambda) C'(\beta, \sigma_v^2, \sigma_\mu^2, \sigma_\eta^2, \lambda), \quad (11)$$

where

$$\begin{aligned}
C(\beta, \sigma_v^2, \sigma_\mu^2, \sigma_\eta^2, \lambda) &= \frac{\partial(\beta', \sigma_v^2, \theta_1, \theta_2, \lambda)}{\partial(\beta', \sigma_v^2, \sigma_\mu^2, \sigma_\eta^2, \lambda)'} \\
&= \begin{pmatrix} I_k & 0_k & 0_k & 0_k & 0_k \\ 0'_k & 1 & \frac{T\sigma_\mu^2}{(T\sigma_\mu^2 + \sigma_v^2)^2} & \frac{N\sigma_\eta^2}{(N\sigma_\eta^2 + \sigma_v^2)^2} & 0 \\ 0'_k & 0 & -\frac{T\sigma_v^2}{(T\sigma_\mu^2 + \sigma_v^2)^2} & 0 & 0 \\ 0'_k & 0 & 0 & -\frac{N\sigma_v^2}{(N\sigma_\eta^2 + \sigma_v^2)^2} & 0 \\ 0'_k & 0 & 0 & 0 & 1 \end{pmatrix}
\end{aligned}$$

with  $0_k$  being a  $k$ -vector of zeros. Similarly, the score function in the original parameterization can be obtained through the relationship

$$S(\beta, \sigma_v^2, \sigma_\mu^2, \sigma_\eta^2, \lambda) = C(\beta, \sigma_v^2, \sigma_\mu^2, \sigma_\eta^2, \lambda) S(\beta, \sigma_v^2, \theta_1, \theta_2, \lambda). \quad (12)$$



Some important remarks are in order:

**Remark 1:** Model (1) and the associated estimation procedures can easily be extended to allow the  $\lambda$  associated with each  $X$  variable to be different from each other and to be different from that associated with the response  $Y$ . See our first empirical application given in Section 4. However, for the simplicity of exposition, we use the same  $\lambda$  for all the variables to be transformed.

**Remark 2:** Our model and estimation method is not restricted to the Box-Cox transformation. Any monotonic transformation available in literature can be used.

**Remark 3:** In case of exact normality after transformation is in doubt, the MLE become quasi-MLE, and under certain regularity conditions they are still consistent and asymptotically normal, but with a different variance.

**Remark 4:** In estimating the traditional panel regression models, it is customary to separate the intercept from the rest of the regression coefficients, and it is popular to use the iterative MLE method for model estimation (see, e.g., Baltagi and Li, 1992; Baltagi, 2001; Hsiao, 2003). The reason for doing so is perhaps mainly the computing speed. With the growing power of even a personal computer, speed is no longer a problem and hence such procedures may no longer be necessary. In fact, it is algebraically simpler and computationally equivalent when putting the intercept together with the other regression coefficients, and then running a full MLE procedure based on (7)-(10). What is problematic for estimating the transformed random effects model is, however, the computer memory, especially when large panels are involved. Following sub-section gives a detailed discussion of this problem and a solution of it.

## 2.2. A computational devise

Many efforts were devoted to developing easier method for finding the parameters estimates to avoid the numerical optimization. This does not seem to be possible for the transformed random effects model. However, with the fast growing availability of the computing software, it is now considered a relative easy matter to maximize a function of three variables as the one given in (7). We use GAUSS/CO (constrained

optimization) procedure to realize this computing task. GAUSS/CO is a powerful procedure for minimizing a function of many variables under certain constraints on the variables. It can perform the minimization using numerical gradients and Hessian, but with the analytical gradients both efficiency and stability of the process can be improved.

However, difficulty arises when panels become large, i.e.,  $N$  and  $T$  become large. Notice that the matrices  $Q, P_1, P_2$ , and  $P_3$  are all  $NT \times NT$ , which can quickly suck out the computer memory. For example, for a medium-sized panel of 600 cross-sections ( $N = 600$ ) and 10 time periods ( $T = 10$ ), these matrices become  $6000 \times 6000$ , each containing 36 millions of numbers. Estimation of the model based panel data of this size would require a computer of 2.0 GB memory. Clearly, when the panel become something like  $10,000 \times 50$  which is common in micro econometrics, it would be impossible to use regular computer to estimate the model. This problem applies to both the transformed random effects models proposed in this paper and the regular random effects models, in both the processes of model estimation and the standard error calculation. A solution to this problem does not seem to exist in the literature.

We now present a simple solution to this problem. Notice the special structure of the four matrices  $Q, P_1, P_2$ , and  $P_3$ , and the type of memory demanding calculations:  $X(\lambda)' \Omega^{-1} X(\lambda)$ ,  $X(\lambda)' \Omega^{-1} h(Y, \lambda)$ ,  $\hat{e}' Q \hat{e}$ ,  $\hat{e}' P_1 \hat{e}$ ,  $\hat{e}'_\lambda P_1 \hat{e}$ , etc.. In order to avoid direct calculations of the matrices  $Q, P_1, P_2$ , and  $P_3$ , all it is necessary is to find an easy way to compute the vector quadratic forms

$$A'QB \quad \text{and} \quad A'P_iB, i = 1, 2, 3,$$

where  $A$  is  $NT \times m$  and  $B$  is  $NT \times l$ , representing various matrices or vectors involved in the estimation procedures such as  $X(\lambda)$ ,  $\hat{e}$ ,  $\hat{e}_\lambda$ , etc.

Recall that all the vectors or matrices are arranged with  $i$  being the slower running index and  $t$  being the faster running index. For each column  $A_j$  of  $A$ , reshape it into an  $N \times T$  matrix so that each row is a time series for a given cross section. Let  $\bar{R}_{Aj}$  be the  $N \times 1$  vector or row means, and  $\bar{C}_{Aj}$  be the  $T \times 1$  vector of column means. Let  $\bar{R}_A = \{\bar{R}_{Aj}\}_{N \times m}$ , and  $\bar{C}_A = \{\bar{C}_{Aj}\}_{T \times m}$ . Similarly, we have the  $N \times l$  matrix of

row means  $\bar{R}_B$ , and the  $T \times l$  matrix of column means  $\bar{C}_B$  for the matrix  $B$ . Let  $\bar{G}_A$  be  $m \times 1$  vector of column means of  $A$ , and  $\bar{G}_B$  be  $l \times 1$  vector of column means of  $B$ . It is easy to show that

$$\begin{aligned} A'P_1B &= T\bar{R}'_A\bar{R}_B - NT\bar{G}_A\bar{G}'_B, \\ A'P_2B &= N\bar{C}'_A\bar{C}_B - NT\bar{G}_A\bar{G}'_B, \\ A'P_3B &= NT\bar{G}_A\bar{G}'_B, \\ A'QB &= A'B - T\bar{R}'_A\bar{R}_B - N\bar{C}'_A\bar{C}_B + NT\bar{G}_A\bar{G}'_B. \end{aligned}$$

These calculations can easily be implemented by writing a GAUSS or MATLAB procedure, which is repeatedly called for by the maximization process and the process of calculating the Hessian matrix. Observing that  $\Omega^{-1} = Q + \theta_1 P_1 + \theta_2 P_2 + \theta_3 P_3$ , all the calculations of the  $NT \times NT$  matrices can be avoided. The gains in memory saving and computing speed in particular the former can be tremendous as exemplified in the numerical examples in Section 4. This method makes the fast calculation of MLEs for a large panel data possible.

With these computational devices, the largest matrices one needs the computer to handle is  $N \times k$  (assuming  $T < N$ ). Thus, a panel of  $N = 60,000$  individuals,  $T < 60,000$  time periods, and  $k = 50$  exogenous variables, can be comfortably handled by a regular desk top computer.

### 2.3. Reduced models

It is some times desirable to fit a transformed model with only individual random effects, or a model with only time random effects. This can be accomplished by simply setting (in the above estimation procedure)  $\theta_2 = 1$  or  $\theta_1 = 1$ . Nevertheless, it may be helpful to lay out the simplified estimation procedures for the reduced model.

For the model with only individual random effects, one maximizes

$$\ell_{\max}(\theta_1, \lambda) = -\frac{NT}{2} \log[\hat{e}'(Q^* + \theta_1 P^*)\hat{e}] + \frac{N}{2} \log \theta_1 + J(\lambda) \quad (13)$$

using the analytical gradients

$$S_{\theta_1} = \frac{N}{2\theta_1} - \frac{NT\hat{e}'(P^*)\hat{e}}{2\hat{e}'(Q^* + \theta_1 P^*)\hat{e}} \quad (14)$$

$$S_{\lambda} = J_{\lambda}(\lambda) - \frac{NT\hat{e}'_{\lambda}(Q^* + \theta_1 P^*)\hat{e}}{\hat{e}'(Q^* + \theta_1 P^*)\hat{e}}, \quad (15)$$

where  $Q^* = I_{NT} - P^*$  and  $P^* = \frac{1}{T}I_N \otimes J_T$ .

For the model with only time random effects, one maximizes

$$\ell_{\max}(\theta_2, \lambda) = -\frac{NT}{2} \log[\hat{e}'(Q^* + \theta_2 P^*)\hat{e}] + \frac{T}{2} \log \theta_2 + J(\lambda) \quad (16)$$

using the analytical gradients

$$S_{\theta_2} = \frac{T}{2\theta_2} - \frac{NT\hat{e}'(P^*)\hat{e}}{2\hat{e}'(Q^* + \theta_2 P^*)\hat{e}} \quad (17)$$

$$S_{\lambda} = J_{\lambda}(\lambda) - \frac{NT\hat{e}'_{\lambda}(Q^* + \theta_2 P^*)\hat{e}}{\hat{e}'(Q^* + \theta_2 P^*)\hat{e}}, \quad (18)$$

where  $Q^* = I_{NT} - P^*$  and  $P^* = \frac{1}{N}J_N \otimes I_T$ .

Of course, when  $\lambda$  is known, the model reduces to the regular two-way random effects model. In this case, one maximizes (7) using the likelihood equations (8) and (9), at the given value of  $\lambda$ .

### 3. Tests for Model Specification

Certain specific values of the parameters  $\sigma_{\mu}^2$ ,  $\sigma_{\eta}^2$  and  $\lambda$  constitute interesting tests for model specifications. For example,  $\sigma_{\mu}^2 = 0$  corresponds to a reduced model of no individual random effect;  $\sigma_{\eta}^2 = 0$  means that time effect does not exist in the model, and  $\lambda = 0$  or  $1$  corresponds to a model of log-linear or linear functional form. Standard likelihood ratio (LR) or Lagrange multipliers (LM) tests are two-sided. However, one-sided tests or a mixture of one-sided and two-sided tests are desirable as the parameters  $\sigma_{\mu}^2$  and  $\sigma_{\eta}^2$  cannot be negative, but  $\lambda$  has no restriction. We follow the methods of Silvapulle and Silvapulle (1995) to give LM and LR tests that take into account of the fact that the random effects parameters cannot be negative.<sup>3</sup>

---

<sup>3</sup>Other references dealing with the issues of one-sided tests include Rogers (1986), Gouriéroux *et al.* (1982), Self and Liang (1987) and Wolak (1991) and Verbeke and Molenberghs (2003).

### 3.1. Joint tests for random effects and functional form

Now, let  $S$  and  $H$  denote the score function and Hessian matrix in  $\sigma$ -parameterization. Let  $\xi = (\sigma_\mu^2, \sigma_\eta^2, \lambda)'$ ,  $S_\xi$  be the subvector of  $S$  corresponding to  $\xi$ , and  $D_{\xi\xi}$  be the submatrix of  $(-H)^{-1}$  corresponding to  $\xi$ . We are interested in testing a general hypothesis of the form

$$H_0^J : \xi = 0 \quad \text{versus} \quad H_a^J : \xi \in C$$

where  $C = \{\xi : \sigma_\mu^2 \geq 0, \sigma_\eta^2 \geq 0, l_1 \leq \lambda \leq l_2, \text{ with } \xi \neq 0, \text{ and } -\infty < l_1 < l_2 < \infty\}$ . Following Silvapulle and Silvapulle (1995), the LM test for  $H_0$  takes the form:

$$\text{LM}_J = \tilde{S}'_\xi \tilde{D}_{\xi\xi} \tilde{S} - \inf_b \{(\tilde{S}_\xi - b)' \tilde{D}_{\xi\xi} (\tilde{S}_\xi - b), b \in C\}, \quad (19)$$

where  $\tilde{S}_\xi$  and  $\tilde{D}_{\xi\xi}$  are  $S_\xi$  and  $D_{\xi\xi}$  evaluated at the constrained MLEs at  $H_0^J$ .

This joint test is fairly easy to calculate. Its asymptotic  $p$ -value for an observed value  $t_0$  of  $\text{LM}_J$  is shown by Silvapulle and Silvapulle (1995) to be within the following interval  $\{0.5P(\chi_1^2 \geq t_0), 0.5[P(\chi_2^2 \geq t_0) + P(\chi_3^2 \geq t_0)]\}$ . It may be of interest to conduct a test involving a nonzero value for  $\lambda$ , e.g.,  $\lambda = 1$  which means a linear functional form. This test can be converted back to the standard form as above by simply considering a reparameterization  $\delta = \lambda - 1$ .

Following the earlier notation and using the expression of the concentrated log likelihood (7), the joint LR test for  $H_0^J$  is

$$\text{LR}_J = 2[\ell_{\max}(\hat{\theta}_1, \hat{\theta}_2, \hat{\lambda}) - \ell_{\max}(1, 1, 0)], \quad (20)$$

which is asymptotically equivalent to  $\text{LM}_J$  and has the same limiting distribution.

### 3.2. Tests for random effects in the presence of data transformation

Once the joint test of  $H_0 : \xi = 0$  is rejected, one would be interested in knowing the cause of this rejection, simply because  $\xi \in C$  does not mean that all of its element are no zeros. There might still be a possibility that some element(s) of  $\xi$  are zeors. A relevant test is then a marginal test, i.e., a test concerning a certain components of  $\xi$ , allowing the presence of the remaining component(s) in the model. We now

consider the test of random effects in the presence of data transformation. Letting  $\sigma^2 = (\sigma_\mu^2, \sigma_\eta^2)'$ , we want to test

$$H_0^R : \sigma^2 = 0 \quad \text{versus} \quad H_a^R : \sigma^2 \in C_0$$

where  $C_0$  is a subset of  $C$  defined in Section 3.1 without the  $\lambda$  component. Let  $\tilde{S}_{\sigma^2}$  and  $\tilde{D}_{\sigma^2\sigma^2}$  be, respectively, the  $\sigma^2$  component of  $S$  and  $(-H)^{-1}$ , evaluated at the constrained MLEs at  $H_0$ . We have, similar to the earlier subsection,

$$LM_R = \tilde{S}_{\sigma^2}' \tilde{D}_{\sigma^2\sigma^2} \tilde{S}_{\sigma^2} - \inf_s \{(\tilde{S}_{\sigma^2} - s)' \tilde{D}_{\sigma^2\sigma^2} (\tilde{S}_{\sigma^2} - s)\}. \quad (21)$$

If the observed value for  $LM_R$  is  $t_0$ , then the asymptotic  $p$ -value for this test lies within the interval  $\{0.5P(\chi_1^2 \geq t_0), 0.5[P(\chi_1^2 \geq t_0) + P(\chi_2^2 \geq t_0)]\}$

Alternatively, the existing test for random effects in the random effects model without data transformation (Baltagi *et al.* 1992) can be extended to allow for the existence of data transformation. The extended test statistic takes the form,

$$LM_R^* = \begin{cases} \frac{NT}{2(T-1)} U_\mu^2(\tilde{\lambda}) + \frac{NT}{2(N-1)} U_\eta^2(\tilde{\lambda}), & \text{if } U_\mu(\tilde{\lambda}) > 0, U_\eta(\tilde{\lambda}) > 0 \\ \frac{NT}{2(T-1)} U_\mu^2(\tilde{\lambda}), & \text{if } U_\mu(\tilde{\lambda}) > 0, U_\eta(\tilde{\lambda}) \leq 0 \\ \frac{NT}{2(N-1)} U_\eta^2(\tilde{\lambda}), & \text{if } U_\mu(\tilde{\lambda}) \leq 0, U_\eta(\tilde{\lambda}) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (22)$$

where  $U_\mu(\tilde{\lambda}) = \tilde{e}'(J_N \otimes I_T)\tilde{e}/(\tilde{e}'\tilde{e}) - 1$ ,  $U_\eta(\tilde{\lambda}) = \tilde{e}'(I_N \otimes J_T)\tilde{e}/(\tilde{e}'\tilde{e}) - 1$ ,  $\tilde{e}$  are the OLS residuals from regressing  $Y(\tilde{\lambda})$  on  $X(\tilde{\lambda})$ , and  $\tilde{\lambda}$  is the constrained MLE of  $\lambda$  at  $H_0$ . The null asymptotic distribution of  $LM_R^*$  is a mixed chi-squared distribution:  $\frac{1}{4}\chi_0^2 + \frac{1}{2}\chi_1^2 + \frac{1}{4}\chi_2^2$ .<sup>4</sup> The test  $LM_R^*$  is interesting as it says that the existing test of Baltagi *et al.* (1992) can simply be extended by replacing a fixed data scale ( $\lambda$  assumed known to be, e.g., 1 or 0) by an estimated one without changing the asymptotic distribution of the test statistic. The proof of this is straightforward and hence omitted. Using the expression (7), the LR test for random effects is

$$LR_R = 2[\ell_{\max}(\hat{\theta}_1, \hat{\theta}_2, \hat{\lambda}) - \ell_{\max}(1, 1, \tilde{\lambda})], \quad (23)$$

---

<sup>4</sup>The critical values for this mixed chi-squared distribution are 2.952, 4.321 and 7.289, respectively, for  $\alpha = 0.1, 0.05$  and  $0.01$ . See Baltagi (2001, Ch. 4).

which has the same asymptotic distribution as  $LM_R$ .

### 3.3. Tests for functional form in the presence of random effects

Often, testing for functional form has his own unique important aspects as a suitable functional form characterizes more correctly the economic relationship. Let  $S_\lambda$  be the score function corresponding to  $\lambda$  and  $D_{\lambda\lambda}$  be the element of  $(-H)^{-1}$  corresponding to  $\lambda$ . We are interested in testing

$$H_0^F : \lambda = \lambda_0 \quad \text{versus} \quad H_a^F : \lambda \neq \lambda_0,$$

assuming the existence of the two-way random effects, i.e.,  $\sigma_\mu^2 > 0$  and  $\sigma_\eta^2 > 0$  (or  $0 < \theta_1 < 1$  and  $0 < \theta_2 < 1$ ). The LM test is simply

$$LM_F = \tilde{S}_\lambda \{\tilde{D}_{\lambda\lambda}\}^{\frac{1}{2}} \quad (24)$$

which has an asymptotic  $N(0, 1)$  distribution under the null hypothesis. The  $\sim$  notation indicates the quantity is evaluated at the parameter estimates at the null. The  $\lambda_0$  can be any convenient value, including 0, 1/2 and 1, representing the functional form of interest. When  $\sigma_\mu^2 = \sigma_\eta^2 = 0$  (or  $\theta_1 = \theta_2 = 1$ ), the test reduces to the Hessian-based LM test of functional form in a cross-sectional regression of Yang and Abeyasinghe (2003). The asymptotically equivalent LR test is

$$LR_F = 2[\ell_{\max}(\hat{\theta}_1, \hat{\theta}_2, \hat{\lambda}) - \ell_{\max}(\tilde{\theta}_1, \tilde{\theta}_2, \lambda_0)], \quad (25)$$

where  $\tilde{\theta}_1$  and  $\tilde{\theta}_2$  are the constrained MLEs of  $\theta_1$  and  $\theta_2$  at the null.

Besides the LM and LR tests, the Wald test can also be readily carried out as parameter estimates and their covariance matrix (the Hessian) have already been obtained. The three tests are asymptotically equivalent. The LM test requires only the estimation of the constrained model, the Wald test requires only the estimation of the full model, and the LR test requires both. Hence, computationally, the LM test is the simplest and this is the main reason for its popularity. However, the LM tests based on Hessian suffers from a drawback that the estimated variance of the score function is not guaranteed to be positive definite when the null values of

the parameters are too far away from their estimates. The  $LM_R^*$  given in (22) is an exception as it is essentially derived from the  $\lambda$ -given expected information matrix. Considering the fact that computing for this proposed model is not an issue, we recommend the LR test as it arises naturally from the model estimation process.

## 4. Empirical Applications

We consider two applications in this section to illustrate the models and estimation methods introduced above, concentrating on the functional form specification, random effects, etc.<sup>5</sup> Results strongly favor the use of general functional form instead of linear or log-linear. Also, the results from the first application reinforce the common perception that public capital has no significant linkage to the state production if the state-specific effects are properly controlled for. The results in the second application reinforce the early findings that a nonlinear relationship exists between age and earnings, and between tenure and earnings. We use the Box-Cox (BC) functional form in both applications

### 4.1. Public capital in private production

In investigating the productivity of public capital in private production, Baltagi and Pinnoi (1995), followed Munnell (1990), considered the following Cobb-Douglas production function:

$$\log Y = \beta_0 + \beta_1 \log K_1 + \beta_2 \log K_2 + \beta_3 \log L + \beta_4 U + u$$

where  $Y$  is gross state product,  $K_1$  is public capital,  $K_2$  is private capital,  $L$  is labor input and  $U$  is the state unemployment rate. The error term  $u$  follows a two-way error components process. Using a panel consisting of annual observations for 48

---

<sup>5</sup>We ignore the endogeneity possibly existed in the regressors, in particular in the second example, due to (i) technical difficulty for Box-Cox type model in handling the endogeneity issue, and (ii) dependence between some of the regressors and the errors may cause bias on the parameter estimates, but the strong evidence for the existence of the general Box-Cox functional form instead of traditional linear or loglinear form should still be informative.



contiguous states over 17 time periods (1970-1986), they arrived at conclusions that the coefficient of the public capital  $K_1$  is insignificantly different from zero, i.e., the public capital is not productive in state private production, if the state-specific effects are under control, but otherwise highly significant. The same conclusions were arrived by Holtz-Eakin (1994) based on a similar panel. Thus, the use of panel model which is able to control the individual effects completely changes the view on the role of public capital on private production. Further, Holtz-Eakin (1994, p. 18) raised concerns with the log linear functional form and argued that use of squares and cross-products of the right-hand-side variables yield essentially the same conclusions.

We consider several generalizations of the log linear functional form based on our transformed random effects model:

1. all the log functions are replaced by the BC transformation;
2. only the  $\log Y$  is replaced by  $(y^\lambda - 1)/\lambda$ ;
3. the log functions in  $Y$ ,  $K_1$  and  $K_2$  are replaced by the BC transformation, and
4. on top of 3., the log function in  $L$  is also replaced by a BC transformation but with a different transformation parameter.

These models are labeled as Model 1 - Model 4. We have also estimated the standard Box-Cox regression model, i.e., the Model 1 ignoring the random effects (denoted as Model BC), and the standard Cobb-Douglas production function (denoted as Model CD) using the MLE method. All the results are summarized in Table 1. Standard errors are in parentheses. The maximum of  $\ell_{\max}(\theta_1, \theta_2, \lambda)$  under various conditions is listed in the last row labeled by `loglik`, based on which the LR tests of model specifications can be carried out.

First, all four models show that the public capital is insignificant (unproductive) in the state's private production when the state-specific effects are controlled for. However, if these effects are not controlled, the coefficient of the public capital becomes highly significant even if a generalized production function is used (Model BC). These are consistent with the results of Baltagi and Pinnoi (1995) and Holtz-Eakin (1994)

who stress on the importance of controlling the unobserved, state-specific characteristics when investigating the productivity of public capital in private production.

The estimated transformation parameter is positive in all the four proposed models. The standard errors of the parameter estimates shows that the transformation parameter is significantly different from either the log or the linear transformation no matter which generalized Cobb-Douglas production function is used. This indicates that it is more appropriate to use the transformed random effects model instead of the standard one. In all the four models, both individual and time random effects are significant. This shows that none of the reduced models should be used for this data.

**Table 1.** Estimation of State Production Function

	Model CD	Model BC	Model 1	Model 2	Model 3	Model 4
$\beta_0$	2.4705 (0.1632)	1.5019 (0.1379)	3.8975 (0.6175)	1.7481 (0.3852)	1.9782 (0.3012)	5.4453 (1.1152)
$\beta_1$	0.0203 (0.0242)	0.1906 (0.0194)	-0.0051 (0.0274)	0.0338 (0.0306)	0.0313 (0.0251)	-0.0004 (0.0274)
$\beta_2$	0.2499 (0.0232)	0.2993 (0.0100)	0.2111 (0.0222)	0.3147 (0.0401)	0.2547 (0.0232)	0.2007 (0.0232)
$\beta_3$	0.7498 (0.0251)	0.9613 (0.0836)	1.7207 (0.1553)	0.9107 (0.0751)	0.9593 (0.0980)	1.5328 (0.1679)
$\beta_4$	-0.0044 (0.0011)	-0.0296 (0.0097)	-0.0368 (0.0127)	-0.0055 (0.0014)	-0.0060 (0.0016)	-0.0384 (0.0134)
$\sigma_v$	0.0012 (0.0001)	0.1306 (0.0658)	0.0952 (0.0439)	0.0018 (0.0003)	0.0020 (0.0004)	0.1014 (0.0475)
$\theta_1$	0.0085 (0.0020)	- - - -	0.0062 (0.0015)	0.0078 (0.0018)	0.0078 (0.0018)	0.0064 (0.0016)
$\theta_2$	0.0841 (0.0310)	- - - -	0.0634 (0.0234)	0.0808 (0.0297)	0.0809 (0.0297)	0.0602 (0.0225)
$\lambda$	- - - -	0.1365 (0.0238)	0.2146 (0.0219)	0.0202 (0.0078)	0.0255 (0.0101)	0.2178 (0.0223)
$\lambda^*$	- - - -	- - - -	- - - -	- - - -	- - - -	0.2359 (0.0250)
loglik	-8701.93	-9309.32	-8654.98	-8698.58	-8698.70	-8653.19

State production data for 48 states and 17 time periods (1970-86), from Munnell (1990), and is given as `Produc.prn` on the Wiley web site associated with Baltagi (2001).

Standard errors are in parentheses.  $\ell_{\max}(1, 1, 0) = -9325.79$  and  $\ell_{\max}(1, 1, 1) = -9892.76$ .

Model 3 is nested within Model 4. The likelihood ratio test for  $H_0 : \lambda^* = 0$  has

statistic value of 91.02, which is  $\chi_1^2$  under  $H_0$ . Thus, model 3 is rejected in favor of Model 4. Model 1 is also nested within Model 4, but the likelihood ratio test of  $H_0 : \lambda = \lambda^*$  is not significant as the statistic value 3.58 lies below the 95% value of  $\chi_1^2$  which is 3.841. This shows that although Model 4 is more general than Model 1, it is not significantly better. Comparing Model CD and Model BC with Model 1 (both are nested within Model 1), the likelihood ratio tests strongly reject both. Thus, we conclude that Model 1 turns out to be the best model and it greatly improves upon the commonly used model with log linear functional specification.

Finally, joint or conditional one-sided LM tests for model specification/reduction devised in Section 3 are also carried out against model 1. As discussed at the end of Section 3,  $LM_J$  and  $LM_R$  failed.  $LM_R^* = 1421.10$  compared with  $LR_R = 1221.48$ , and  $LM_F = 98.74$  compared with  $LR_F = 93.90$  and  $Wald_F = 96.02$ . Other Wald tests have similar magnitudes as the corresponding LR tests. All the hypotheses discussed in Section 3 are soundly rejected. These show that Model 1 cannot be further reduced within the framework of the Box-Cox functional family.

## 4.2. Wage distribution of U.S. male workers

This application concerns the wage distribution of the U.S. males workers. We use the data from the Panel Study of Income Dynamics (PSID), extracted and analyzed by Polachek and Yoon (1996). The data consist of 13,408 observations from 838 white male workers over 16 years (1969-1984). The endogenous variable is  $Y$  = wage in 1967 constant dollars, and the exogenous variables are  $X_1$  = education in years,  $X_2$  = experience in years,  $X_3$  = tenure in months, and possibly the quadratic terms for experience and tenure.

We fit four models to the wage variable, which enters into model in either log form (Columns 1 & 2 in Table 2) or Box-Cox form (columns 3 & 4). The endogenous variables enter into the model in either linear form or quadratic form. Our analyses concentrate on the choice of functional form, wage distribution, existence of unobservable individual and time effects, etc. Our model allows for testing on the conventional wisdom on wage distribution, firm-size distribution, etc., stemmed from the Gibrat's

Law, stating that the distributions of wages, firm sizes, etc. are lognormal.<sup>6</sup> The results are summarized in Table 2. The estimated standard errors are in parentheses. The values in the last row again denote the maximized  $\ell_{\max}(\theta_1, \theta_2, \lambda)$  under various model specifications. Further,  $\ell_{\max}(1, 1, \tilde{\lambda}) = -71012.23$  where  $\tilde{\lambda} = -0.1455$  is the constrained MLE of  $\lambda$  under  $H_0^R$ , and  $\ell_{\max}(1, 1, 0) = -71087.55$ .

From the results of Table 2, we first note that the two quadratic terms are both highly significant in both the model with log functional form and the model with Box-Cox functional form, thus should be included in the model. The likelihood ratio test for the joint significance of the two quadratic terms gives a value of 390.28 based on the first two models, and 394.58 based on the last two models. Both values should be referred to the 95%  $\chi_2^2$  cutoff value which is 5.991, showing that the two quadratic terms are significant. Second, we note that both the individual and time random effects are significant no matter which functional form we use. It is rather surprising to note that the estimates of these random effects, in particular that of the individual random effect, are quite robust against the functional form specification. Third, the data strongly favor the Box-Cox functional form to the log form. The asymptotic  $t$ -ratio for the  $\lambda$ -estimate is 4.79 based on the BC linear model, and 5.23 based on the BC quadratic model, both showing that  $\lambda$  is significantly different from zero, i.e., the log functional form. Further, the likelihood ratio test of  $H_0 : \lambda = 0$  gives a value of 23.22 based on log linear and BC linear models, and 27.52 based on the log quadratic and BC quadratic models. Both values should be compared with 3.841, the 95%  $\chi_1^2$  cutoff value. Thus, log linear model is rejected in favor of BC linear, and log quadratic is rejected in favor of BC quadratic. Hence, the final model of our choice should be the BC quadratic model. Finally, we conclude that log normality for the wage distribution is soundly rejected based on this panel data.

The various LM, LR and Wald tests of model specification discussed in Section 3 are also carried out, and the (available) results show that the BC quadratic model cannot be further reduced. For example,  $LM_R^* = 12721$ ,  $LR_R = 9124$ , and  $LR_J =$

---

<sup>6</sup>See Sutton (1997) for a survey of work on Gibrat's Law

9275. Once again, the  $LM_J$  and  $LM_R$  failed as the tested parameter values are too far from their estimates, resulting a negative variance estimate. Polachek and Yoon (1996) analyzed these data using a two-tiered frontier model with log-quadratic functional form specification. Our estimates based on the random effects model of log-quadratic functional form are quite comparable with theirs. However, with the BC-quadratic functional form, our model shows that the wage-experience and wage-tenure relationships are generally nonlinear instead of simply log-quadratic as concluded by Polachek and Yoon (1996).

**Table 2.** Estimation of Wage Distribution

	log linear	log quadratic	BC linear	BC quadratic
Constant	0.28207163 (.07170620)	0.01347445 (.06984960)	0.31060998 (.06680705)	0.06280708 (.06508298)
Education	0.07123583 (.00420110)	0.06953602 (.00409380)	0.06595067 (.00403634)	0.06396318 (.00390758)
Experience	0.00869209 (.00123180)	0.03508301 (.00182534)	0.00799320 (.00115359)	0.03248744 (.00175437)
Tenure	-0.00000592 (.00004009)	0.00041455 (.00008824)	-0.00000331 (.00003730)	0.00038785 (.00008178)
Experience <sup>2</sup>	- - - -	-0.00053445 (.00002896)	- - - -	-0.00049707 (.00002765)
Tenure <sup>2</sup>	- - - -	-0.00000103 (.00000025)	- - - -	-0.00000095 (.00000023)
$\sigma_v^2$	0.08077117 (.00101947)	0.07852969 (.00099117)	0.06993177 (.00225966)	0.06718576 (.00215992)
$\theta_1$	0.04149796 (.00209503)	0.04190114 (.00211514)	0.04183684 (.00211351)	0.04229244 (.00213643)
$\theta_2$	0.05568133 (.01994550)	0.07624718 (.02744946)	0.05479457 (.01962507)	0.07484133 (.02693748)
$\lambda$	- - - -	- - - -	-0.05188155 (.01084216)	-0.05611794 (.01077966)
loglik	-66658.9419	-66463.7995	-66647.3309	-66450.0418

PSID data for 838 white male workers over 16 time periods, 1969-84, given in Polachek and Yoon (1996). Standard errors are in parentheses.

We conclude this section by offering some remarks on computing. The panel used in our first application is quite small ( $NT = 48 \times 17 = 816$ ), it can be comfortably handled by a personal computer with or without using the computing technique

presented in Section 2.2. However, for second application, the panel becomes  $NT = 838 \times 16 = 13,408$ , much larger than the first one. Using a PC with 512 MB memory, the Gauss program without using the computational device quickly stops, showing a message of “insufficient memory”. In contrast, with the use of the computational device, the Gauss program returns the desired results in just a few seconds (depending on the initial values). Testing on much larger (simulated) panels ( $N = 60,000, T < N$ ) further prove the value of the proposed computational technique. Gauss programs (with or without use of the computational device) for the computations of the two applications are available from the authors upon request.

## 5. Monte Carlo Simulations

Monte Carlo experiments are conducted to check the finite sample performance of the MLEs of model parameters and subsequence statistical inferences. The data generating process (DGP) used in the Monte Carlo experiments is as follows.

$$h(Y, \lambda) = \beta_0 + \beta_1 X_1 + \beta_2 h(X_2, \lambda) + Z_\mu \mu + Z_\eta \eta + v$$

where  $h$  is the Box-Cox power transformation with  $\lambda = 0.1$ ,  $X_1$  is generated from  $U(0, 5)$ ,  $X_2$  from  $\exp[N(0, 1)]$ ,  $\beta = (20, 5, 1)'$ ,  $\sigma_\mu = \{0.2, 0.6, 1.2\}$ ,  $\sigma_\eta = \{0.2, 0.6, 1.2\}$ , and  $\sigma_v = \{0.5, 1.0, 1.5\}$ . 10,000 samples are generated for each Monte Carlo experiment corresponding each parameter configuration.

To generate error components  $\{\mu_i\}$ ,  $\{\eta_t\}$  and  $\{v_{it}\}$ , we consider three distributions: (i) normal, (ii) normal-mixture, and (iii) normal-gamma mixture, all standardized to have zero mean and unit variance. The standardized normal-mixture random variates are generated according to

$$W_i = ((1 - \xi_i)Z_i + \xi_i \tau Z_i) / (1 - p + p * \tau^2)^{0.5},$$

where  $\xi$  is a Bernoulli random variable with probability of success  $p$  and  $Z_i$  is standard normal independent of  $\xi$ . The parameter  $p$  in this case also represents the proportion of mixing the two normal populations. In our experiments, we choose  $p = 0.05$

or 0.10, meaning that 95% or 90% of the random variates are generated from the standard normal and the remaining 5% or 10% are from another normal population with standard deviation  $\tau$ . We choose  $\tau = 2$  to simulate the situations where there are mild departure from normality in the form of excess kurtosis. Similarly, the standardized normal-gamma mixture random variates are generated according to

$$W_i = ((1 - \xi_i)Z_i + \xi_i(V_i - \alpha))/(1 - p + p * \alpha)^{0.5},$$

where  $V_i$  is a gamma random variable with scale parameter 1 and shape parameter  $\alpha$ , and is independent of  $Z_i$  and  $\xi_i$ . The other quantities are the same as in the definition of normal-mixture. We choose  $p = 0.05$  or  $0.10$ , and  $\alpha = 1$ . Again, this represents a situation where there is a mild departure from normality, but in the form of both excess kurtosis and skewness.

Table 3 presents results for normal-mixture and Table 4 presents results for normal-gamma mixture, with  $p = 0.0, 0.05$ , and  $0.1$ . Note that when  $p = 0.0$ , the errors are exactly normal. The reported results include bias in percentage (% bias) of the parameter estimates, root mean squared error (rmse), and the empirical coverage probability for a nominal 95% confidence interval (95% CI) for each of the model parameters.

From the results, we see that the finite sample performance of the maximum likelihood estimation and inference is very good. Some general observations are in order: (i) the bias and rmse decrease quickly as  $N$  and  $T$  increase; (ii) the empirical coverage probability gets closer to its nominal level in general as  $N$  and  $T$  increase; (iii) as  $(\sigma_\mu, \sigma_\eta, \sigma_v)$  increases, the coverage probability decreases, and (iv) the Gaussian likelihood-based inference is quite robust against mild departure from normality.

Some details are as follows. The estimators  $\hat{\theta}_1$  and  $\hat{\theta}_2$  are much more biased than the others when sample sizes are small. The CI for  $\sigma_v^2$  has empirical coverage significantly lower than the nominal level when sample sizes are (10,10), but improves drastically when sample sizes are (20,20). Monte Carlo experiments are repeated under some other  $\lambda$  values, and the results (not reported for brevity) are similar.

**Table 3.** Bias, RMSE and Empirical Coverage for 95% CI: Normal Mixture,  $\lambda = .1$

Par	$p = 0.0$			$p = 0.05$			$p = 0.1$		
	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (10, 10, 0.2, 0.2, .5)$									
$\beta_0$	0.1387	1.0786	0.9407	0.0843	1.0799	0.9384	0.1318	1.1020	0.9374
$\beta_1$	0.9363	0.6838	0.9398	0.7564	0.6758	0.9359	0.8009	0.6757	0.9356
$\beta_2$	0.9712	0.1451	0.9366	0.6451	0.1413	0.9353	0.7101	0.1389	0.9353
$\sigma_v^2$	0.1975	0.0791	0.9038	-0.4362	0.0840	0.8853	-0.3082	0.0854	0.8821
$\theta_1$	27.5955	0.2580	0.9510	27.6838	0.2611	0.9491	28.5370	0.2646	0.9485
$\theta_2$	28.4713	0.2604	0.9546	28.0268	0.2624	0.9494	27.5481	0.2607	0.9517
$\lambda$	0.0155	0.0093	0.9401	-0.0897	0.0092	0.9383	-0.0553	0.0093	0.9387
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (10, 10, 0.6, 0.6, 1.0)$									
$\beta_0$	0.3914	2.2191	0.9368	0.4597	2.0883	0.9332	0.7083	2.3271	0.9259
$\beta_1$	3.3761	1.4619	0.9330	2.7383	1.2736	0.9286	4.2572	1.5373	0.9220
$\beta_2$	3.0534	0.3047	0.9288	2.1200	0.2616	0.9251	4.0519	0.3236	0.9169
$\sigma_v^2$	11.4578	0.7050	0.8805	7.6880	0.5972	0.8783	14.0145	0.7760	0.8700
$\theta_1$	30.5338	0.1777	0.9506	30.6975	0.1834	0.9489	32.1472	0.1865	0.9446
$\theta_2$	32.1446	0.1814	0.9525	32.6730	0.1870	0.9497	33.5013	0.1907	0.9491
$\lambda$	-0.3308	0.0190	0.9400	-0.1683	0.0171	0.9362	0.0715	0.0198	0.9266
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (10, 10, 1.2, 1.2, 1.5)$									
$\beta_0$	0.6160	2.6908	0.9382	0.5637	2.7076	0.9320	0.7908	2.8557	0.9243
$\beta_1$	4.2078	1.6677	0.9270	4.1096	1.6729	0.9206	5.5891	1.8700	0.9165
$\beta_2$	3.9903	0.3654	0.9254	4.7390	0.3901	0.9177	5.1455	0.4082	0.9073
$\sigma_v^2$	15.0132	1.8365	0.8757	15.8959	1.9653	0.8600	20.9714	2.2679	0.8517
$\theta_1$	32.4313	0.1224	0.9466	32.1654	0.1228	0.9471	33.2427	0.1281	0.9472
$\theta_2$	31.4464	0.1195	0.9515	33.6320	0.1249	0.9486	32.4591	0.1253	0.9451
$\lambda$	-0.4843	0.0218	0.9463	-0.6000	0.0220	0.9368	-0.2426	0.0237	0.9305
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 20, 0.2, 0.2, .5)$									
$\beta_0$	0.0242	0.5047	0.9496	-0.0047	0.5076	0.9463	0.0682	0.5167	0.9469
$\beta_1$	0.1708	0.3132	0.9486	0.1038	0.3177	0.9479	0.2762	0.3161	0.9451
$\beta_2$	0.1557	0.0664	0.9476	0.0715	0.0676	0.9455	0.2688	0.0685	0.9489
$\sigma_v^2$	-0.0763	0.0362	0.9380	-0.1082	0.0391	0.9260	0.1715	0.0394	0.9194
$\theta_1$	14.3707	0.1075	0.9576	14.5408	0.1066	0.9601	14.0182	0.1070	0.9517
$\theta_2$	13.9131	0.1044	0.9599	14.5817	0.1064	0.9572	15.4529	0.1094	0.9575
$\lambda$	-0.0154	0.0043	0.9511	-0.0653	0.0044	0.9473	0.0533	0.0044	0.9444
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 20, 0.6, 0.6, 1.0)$									
$\beta_0$	0.0841	0.9646	0.9472	0.0508	1.0184	0.9384	0.1496	1.0176	0.9406
$\beta_1$	0.5885	0.5802	0.9437	0.5639	0.6193	0.9366	0.8088	0.6231	0.9366
$\beta_2$	0.7558	0.1287	0.9451	0.5289	0.1357	0.9365	0.8501	0.1323	0.9382
$\sigma_v^2$	1.7311	0.2482	0.9305	1.8025	0.2704	0.9177	2.2933	0.2724	0.9176
$\theta_1$	13.4402	0.0533	0.9591	14.7901	0.0556	0.9553	14.1468	0.0547	0.9551
$\theta_2$	14.5459	0.0543	0.9622	14.2944	0.0549	0.9559	13.6030	0.0550	0.9519
$\lambda$	-0.0507	0.0080	0.9476	-0.1289	0.0086	0.9370	0.0313	0.0086	0.9411



**Table 3.** Cont'd

	$p = 0.0$			$p = 0.05$			$p = 0.1$		
	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 20, 1.2, 1.2, 1.5)$									
$\beta_0$	0.1419	1.3197	0.9423	0.2822	1.3729	0.9369	0.1761	1.3418	0.9281
$\beta_1$	0.9224	0.7744	0.9433	1.3734	0.8229	0.9297	1.1933	0.8043	0.9241
$\beta_2$	0.8876	0.1744	0.9431	1.3244	0.1828	0.9324	1.2382	0.1824	0.9263
$\sigma_v^2$	3.3825	0.7427	0.9251	4.4983	0.8124	0.9178	4.0706	0.8004	0.9076
$\theta_1$	13.5800	0.0317	0.9594	14.0906	0.0329	0.9522	13.5103	0.0323	0.9546
$\theta_2$	14.1420	0.0325	0.9549	13.4636	0.0322	0.9508	13.4448	0.0321	0.9546
$\lambda$	-0.1585	0.0106	0.9467	0.0650	0.0112	0.9333	-0.0318	0.0109	0.9236
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 20, 0.2, 0.2, .5)$									
$\beta_0$	-0.0084	0.3335	0.9484	-0.0108	0.3261	0.9463	0.0364	0.3275	0.9488
$\beta_1$	0.0321	0.2036	0.9468	0.0164	0.1996	0.9464	0.1426	0.2022	0.9495
$\beta_2$	0.0333	0.0439	0.9475	-0.0066	0.0434	0.9463	0.1657	0.0429	0.9492
$\sigma_v^2$	-0.0976	0.0233	0.9471	-0.1865	0.0243	0.9277	0.1461	0.0250	0.9285
$\theta_1$	5.0909	0.0550	0.9512	5.0433	0.0552	0.9516	4.8053	0.0560	0.9492
$\theta_2$	14.8764	0.0488	0.9644	15.7106	0.0496	0.9598	15.8558	0.0519	0.9563
$\lambda$	-0.0338	0.0028	0.9493	-0.0455	0.0028	0.9457	0.0423	0.0028	0.9483
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 20, 0.6, 0.6, 1.0)$									
$\beta_0$	0.0214	0.6100	0.9469	0.0213	0.6188	0.9387	0.0448	0.6511	0.9355
$\beta_1$	0.1798	0.3634	0.9472	0.2302	0.3725	0.9430	0.2986	0.3902	0.9344
$\beta_2$	0.2115	0.0812	0.9474	0.2408	0.0814	0.9468	0.3214	0.0837	0.9373
$\sigma_v^2$	0.5588	0.1518	0.9443	0.7380	0.1592	0.9343	0.8820	0.1692	0.9230
$\theta_1$	4.9682	0.0283	0.9577	4.6957	0.0284	0.9489	4.7624	0.0286	0.9494
$\theta_2$	15.8645	0.0238	0.9614	16.0105	0.0238	0.9639	15.3871	0.0239	0.9613
$\lambda$	-0.0543	0.0050	0.9488	-0.0301	0.0051	0.9421	-0.0004	0.0054	0.9365
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 20, 1.2, 1.2, 1.5)$									
$\beta_0$	0.0448	0.8144	0.9485	0.1341	0.8702	0.9343	0.1195	0.8796	0.9358
$\beta_1$	0.3876	0.4598	0.9484	0.6571	0.4965	0.9334	0.5960	0.5081	0.9297
$\beta_2$	0.4060	0.1032	0.9491	0.6484	0.1082	0.9367	0.5329	0.1113	0.9318
$\sigma_v^2$	1.2218	0.4291	0.9428	1.8830	0.4743	0.9268	1.8882	0.4881	0.9197
$\theta_1$	4.8254	0.0168	0.9522	4.1764	0.0167	0.9508	4.6862	0.0169	0.9499
$\theta_2$	14.8253	0.0135	0.9627	15.2566	0.0136	0.9578	15.8393	0.0137	0.9587
$\lambda$	-0.0225	0.0063	0.9493	0.1227	0.0068	0.9312	0.0622	0.0070	0.9309
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 50, 0.2, 0.2, .5)$									
$\beta_0$	0.0112	0.3238	0.9516	0.0239	0.3247	0.9452	0.0266	0.3392	0.9485
$\beta_1$	0.0790	0.1975	0.9513	0.1012	0.1989	0.9456	0.1116	0.2089	0.9484
$\beta_2$	0.0878	0.0423	0.9495	0.0765	0.0426	0.9451	0.1036	0.0446	0.9463
$\sigma_v^2$	-0.0701	0.0229	0.9443	-0.0168	0.0242	0.9314	-0.0104	0.0256	0.9257
$\theta_1$	15.8208	0.0503	0.9629	16.0525	0.0520	0.9589	15.7823	0.0500	0.9571
$\theta_2$	4.5343	0.0543	0.9547	4.7456	0.0548	0.9557	4.6460	0.0551	0.9526
$\lambda$	0.0017	0.0027	0.9505	0.0173	0.0028	0.9460	0.0154	0.0029	0.9468

**Table 3.** Cont'd

	$p = 0.0$			$p = 0.05$			$p = 0.1$		
	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 50, 0.6, 0.6, 1.0)$									
$\beta_0$	-0.0134	0.6142	0.9438	0.0569	0.6270	0.9423	0.0414	0.6486	0.9309
$\beta_1$	0.1067	0.3616	0.9436	0.2987	0.3770	0.9425	0.3071	0.3924	0.9328
$\beta_2$	0.1254	0.0781	0.9439	0.3105	0.0810	0.9434	0.3886	0.0853	0.9371
$\sigma_v^2$	0.4081	0.1527	0.9393	0.8606	0.1624	0.9337	0.9318	0.1702	0.9218
$\theta_1$	14.8354	0.0229	0.9616	14.6790	0.0231	0.9603	15.8934	0.0235	0.9633
$\theta_2$	4.5703	0.0279	0.9547	5.1709	0.0285	0.9520	4.7445	0.0280	0.9543
$\lambda$	-0.1059	0.0050	0.9442	0.0172	0.0052	0.9419	0.0015	0.0054	0.9318
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 50, 1.2, 1.2, 1.5)$									
$\beta_0$	-0.0115	0.8184	0.9486	0.0448	0.8516	0.9387	0.1505	0.8959	0.9336
$\beta_1$	0.2847	0.4630	0.9495	0.4167	0.4901	0.9351	0.6891	0.5189	0.9297
$\beta_2$	0.1781	0.1011	0.9483	0.3984	0.1072	0.9406	0.6231	0.1122	0.9315
$\sigma_v^2$	1.0174	0.4334	0.9408	1.4785	0.4664	0.9259	2.1340	0.4961	0.9212
$\theta_1$	15.6651	0.0139	0.9563	15.2617	0.0134	0.9598	15.4448	0.0137	0.9579
$\theta_2$	4.4971	0.0167	0.9504	5.0520	0.0170	0.9518	4.5319	0.0166	0.9508
$\lambda$	-0.1023	0.0064	0.9507	-0.0346	0.0067	0.9372	0.1172	0.0071	0.9273
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 50, 0.2, 0.2, .5)$									
$\beta_0$	-0.0028	0.1985	0.9512	0.0009	0.2008	0.9479	0.0235	0.2052	0.9445
$\beta_1$	0.0117	0.1200	0.9511	0.0192	0.1218	0.9459	0.0676	0.1234	0.9437
$\beta_2$	-0.0069	0.0255	0.9500	0.0140	0.0261	0.9468	0.0647	0.0260	0.9440
$\sigma_v^2$	0.0226	0.0140	0.9483	0.0430	0.0151	0.9296	0.1313	0.0155	0.9202
$\theta_1$	5.2749	0.0254	0.9514	5.5840	0.0262	0.9504	5.6152	0.0260	0.9527
$\theta_2$	5.2092	0.0255	0.9556	5.3008	0.0255	0.9527	5.4988	0.0254	0.9534
$\lambda$	-0.0126	0.0017	0.9506	-0.0062	0.0017	0.9469	0.0261	0.0017	0.9427
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 50, 0.6, 0.6, 1.0)$									
$\beta_0$	-0.0073	0.3755	0.9509	0.0213	0.3935	0.9436	0.0413	0.3915	0.9387
$\beta_1$	0.0425	0.2202	0.9489	0.1372	0.2331	0.9428	0.1646	0.2309	0.9382
$\beta_2$	0.0394	0.0477	0.9490	0.1001	0.0508	0.9418	0.1977	0.0499	0.9405
$\sigma_v^2$	0.1118	0.0928	0.9467	0.3985	0.0995	0.9374	0.4366	0.1002	0.9295
$\theta_1$	5.0095	0.0120	0.9527	5.6816	0.0122	0.9526	5.2457	0.0122	0.9501
$\theta_2$	5.0367	0.0121	0.9511	5.1018	0.0120	0.9555	4.7491	0.0120	0.9545
$\lambda$	-0.0351	0.0030	0.9495	0.0211	0.0032	0.9423	0.0407	0.0032	0.9376
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 50, 1.2, 1.2, 1.5)$									
$\beta_0$	-0.0193	0.5257	0.9511	0.0364	0.5536	0.9424	0.1483	0.5564	0.9372
$\beta_1$	0.0610	0.2869	0.9502	0.2051	0.3064	0.9368	0.4187	0.3169	0.9285
$\beta_2$	0.0243	0.0643	0.9506	0.1661	0.0690	0.9380	0.4209	0.0700	0.9334
$\sigma_v^2$	0.3188	0.2646	0.9478	0.6211	0.2856	0.9324	1.1417	0.3001	0.9230
$\theta_1$	5.4190	0.0071	0.9491	5.0763	0.0069	0.9527	5.2458	0.0070	0.9501
$\theta_2$	5.0622	0.0069	0.9515	5.2553	0.0070	0.9546	5.0317	0.0070	0.9484
$\lambda$	-0.0664	0.0040	0.9520	0.0157	0.0043	0.9368	0.1552	0.0044	0.9270

**Table 4.** Bias, RMSE and Empirical Coverage for 95% CI: Normal-Gamma Mix.,  $\lambda = .1$

	$p = 0.0$			$p = 0.05$			$p = 0.1$		
	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (10, 10, 0.2, 0.2, .5)$									
$\beta_0$	0.1387	1.0786	0.9407	0.0559	1.0833	0.9377	0.0954	1.0947	0.9341
$\beta_1$	0.9363	0.6838	0.9398	0.6970	0.6756	0.9369	0.7096	0.6680	0.9343
$\beta_2$	0.9712	0.1451	0.9366	0.5765	0.1407	0.9353	0.6100	0.1369	0.9363
$\sigma_v^2$	0.1975	0.0791	0.9038	-0.4202	0.0801	0.8946	-0.4909	0.0805	0.8955
$\theta_1$	27.5955	0.2580	0.9510	27.9857	0.2599	0.9528	28.6888	0.2642	0.9544
$\theta_2$	28.4713	0.2604	0.9546	28.2862	0.2616	0.9500	27.7482	0.2593	0.9538
$\lambda$	0.0155	0.0093	0.9401	-0.1364	0.0092	0.9395	-0.1082	0.0092	0.9358
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (10, 10, 0.6, 0.6, 1.0)$									
$\beta_0$	0.3914	2.2191	0.9368	0.3272	2.0506	0.9381	0.4917	2.2470	0.9355
$\beta_1$	3.3761	1.4619	0.9330	2.3040	1.2406	0.9323	3.4600	1.4611	0.9307
$\beta_2$	3.0534	0.3047	0.9288	1.8359	0.2581	0.9273	3.2947	0.3087	0.9267
$\sigma_v^2$	11.4578	0.7050	0.8805	7.0317	0.5770	0.8889	11.4402	0.7025	0.8778
$\theta_1$	30.5338	0.1777	0.9506	31.5011	0.1836	0.9531	32.7323	0.1879	0.9492
$\theta_2$	32.1446	0.1814	0.9525	33.2769	0.1881	0.9514	32.9169	0.1850	0.9551
$\lambda$	-0.3308	0.0190	0.9400	-0.3787	0.0168	0.9407	-0.2388	0.0190	0.9381
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (10, 10, 1.2, 1.2, 1.5)$									
$\beta_0$	0.6160	2.6908	0.9382	0.1624	2.6783	0.9324	-0.0561	2.7250	0.9313
$\beta_1$	4.2078	1.6677	0.9270	2.9737	1.6182	0.9216	2.9673	1.7171	0.9127
$\beta_2$	3.9903	0.3654	0.9254	3.3636	0.3745	0.9154	2.4886	0.3759	0.9117
$\sigma_v^2$	15.0132	1.8365	0.8757	12.4124	1.7937	0.8572	13.3607	1.8927	0.8519
$\theta_1$	32.4313	0.1224	0.9466	31.8950	0.1219	0.9481	32.9125	0.1235	0.9499
$\theta_2$	31.4464	0.1195	0.9515	33.6525	0.1244	0.9536	32.4661	0.1233	0.9515
$\lambda$	-0.4843	0.0218	0.9463	-1.2666	0.0218	0.9416	-1.6475	0.0229	0.9398
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 20, 0.2, 0.2, .5)$									
$\beta_0$	0.0242	0.5047	0.9496	-0.0276	0.5021	0.9474	0.0452	0.5150	0.9451
$\beta_1$	0.1708	0.3132	0.9486	0.0411	0.3137	0.9473	0.2153	0.3144	0.9441
$\beta_2$	0.1557	0.0664	0.9476	0.0013	0.0667	0.9454	0.2200	0.0684	0.9456
$\sigma_v^2$	-0.0763	0.0362	0.9380	-0.2168	0.0371	0.9337	0.0042	0.0375	0.9335
$\theta_1$	14.3707	0.1075	0.9576	14.5538	0.1059	0.9581	13.8887	0.1058	0.9519
$\theta_2$	13.9131	0.1044	0.9599	14.8473	0.1065	0.9571	15.4706	0.1082	0.9548
$\lambda$	-0.0154	0.0043	0.9511	-0.1050	0.0043	0.9476	0.0124	0.0044	0.9450
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 20, 0.6, 0.6, 1.0)$									
$\beta_0$	0.0841	0.9646	0.9472	-0.1501	0.9971	0.9419	-0.1839	0.9850	0.9448
$\beta_1$	0.5885	0.5802	0.9437	0.0402	0.5996	0.9387	-0.0517	0.5930	0.9430
$\beta_2$	0.7558	0.1287	0.9451	-0.0090	0.1318	0.9416	0.0058	0.1272	0.9402
$\sigma_v^2$	1.7311	0.2482	0.9305	0.5224	0.2541	0.9192	0.3674	0.2498	0.9256
$\theta_1$	13.4402	0.0533	0.9591	14.6529	0.0550	0.9565	14.3654	0.0546	0.9599
$\theta_2$	14.5459	0.0543	0.9622	14.1006	0.0545	0.9570	13.4086	0.0542	0.9524
$\lambda$	-0.0507	0.0080	0.9476	-0.4676	0.0084	0.9454	-0.5290	0.0083	0.9475

**Table 4.** Cont'd

	$p = 0.0$			$p = 0.05$			$p = 0.1$		
	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 20, 1.2, 1.2, 1.5)$									
$\beta_0$	0.1419	1.3197	0.9423	-0.1545	1.3293	0.9403	-0.6950	1.2960	0.9300
$\beta_1$	0.9224	0.7744	0.9433	0.2119	0.7785	0.9360	-0.9928	0.7484	0.9268
$\beta_2$	0.8876	0.1744	0.9431	0.2289	0.1758	0.9342	-1.0777	0.1710	0.9236
$\sigma_v^2$	3.3825	0.7427	0.9251	1.8803	0.7383	0.9185	-0.7022	0.7012	0.9047
$\theta_1$	13.5800	0.0317	0.9594	14.1314	0.0327	0.9541	13.3984	0.0319	0.9559
$\theta_2$	14.1420	0.0325	0.9549	13.6103	0.0324	0.9542	13.5376	0.0320	0.9580
$\lambda$	-0.1585	0.0106	0.9467	-0.6699	0.0108	0.9437	-1.4885	0.0106	0.9360
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 20, 0.2, 0.2, .5)$									
$\beta_0$	-0.0084	0.3335	0.9484	-0.0415	0.3234	0.9450	-0.0114	0.3264	0.9491
$\beta_1$	0.0321	0.2036	0.9468	-0.0592	0.1972	0.9470	0.0223	0.1995	0.9501
$\beta_2$	0.0333	0.0439	0.9475	-0.0902	0.0431	0.9444	0.0468	0.0425	0.9470
$\sigma_v^2$	-0.0976	0.0233	0.9471	-0.3297	0.0230	0.9410	-0.1212	0.0236	0.9379
$\theta_1$	5.0909	0.0550	0.9512	5.0823	0.0554	0.9542	4.8888	0.0558	0.9526
$\theta_2$	14.8764	0.0488	0.9644	15.7478	0.0497	0.9618	15.7914	0.0514	0.9587
$\lambda$	-0.0338	0.0028	0.9493	-0.0975	0.0027	0.9471	-0.0405	0.0028	0.9489
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 20, 0.6, 0.6, 1.0)$									
$\beta_0$	0.0214	0.6100	0.9469	-0.1485	0.6078	0.9426	-0.3103	0.6340	0.9395
$\beta_1$	0.1798	0.3634	0.9472	-0.1962	0.3616	0.9451	-0.6008	0.3715	0.9409
$\beta_2$	0.2115	0.0812	0.9474	-0.2107	0.0792	0.9456	-0.5686	0.0805	0.9425
$\sigma_v^2$	0.5588	0.1518	0.9443	-0.1682	0.1516	0.9357	-1.0510	0.1551	0.9304
$\theta_1$	4.9682	0.0283	0.9577	4.5788	0.0281	0.9521	4.6142	0.0281	0.9516
$\theta_2$	15.8645	0.0238	0.9614	16.0569	0.0239	0.9620	15.0444	0.0234	0.9631
$\lambda$	-0.0543	0.0050	0.9488	-0.3187	0.0050	0.9454	-0.6093	0.0052	0.9441
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 20, 1.2, 1.2, 1.5)$									
$\beta_0$	0.0448	0.8144	0.9485	-0.3091	0.8365	0.9444	-0.7377	0.8563	0.9306
$\beta_1$	0.3876	0.4598	0.9484	-0.4622	0.4680	0.9412	-1.5513	0.4748	0.9290
$\beta_2$	0.4060	0.1032	0.9491	-0.4251	0.1029	0.9426	-1.5859	0.1052	0.9327
$\sigma_v^2$	1.2218	0.4291	0.9428	-0.4216	0.4340	0.9339	-2.5331	0.4352	0.9126
$\theta_1$	4.8254	0.0168	0.9522	4.2923	0.0167	0.9524	4.7270	0.0168	0.9531
$\theta_2$	14.8253	0.0135	0.9627	15.4184	0.0136	0.9588	15.8453	0.0136	0.9583
$\lambda$	-0.0225	0.0063	0.9493	-0.6249	0.0066	0.9457	-1.4049	0.0068	0.9361
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 50, 0.2, 0.2, .5)$									
$\beta_0$	0.0112	0.3238	0.9516	-0.0009	0.3228	0.9485	-0.0210	0.3398	0.9468
$\beta_1$	0.0790	0.1975	0.9513	0.0372	0.1971	0.9469	-0.0111	0.2082	0.9455
$\beta_2$	0.0878	0.0423	0.9495	0.0231	0.0421	0.9505	-0.0216	0.0445	0.9455
$\sigma_v^2$	-0.0701	0.0229	0.9443	-0.1689	0.0231	0.9397	-0.2060	0.0247	0.9341
$\theta_1$	15.8208	0.0503	0.9629	15.9845	0.0515	0.9593	15.8088	0.0499	0.9610
$\theta_2$	4.5343	0.0543	0.9547	4.6887	0.0543	0.9545	4.8332	0.0550	0.9520
$\lambda$	0.0017	0.0027	0.9505	-0.0260	0.0027	0.9479	-0.0695	0.0029	0.9446

**Table 4.** Cont'd

	$p = 0.0$			$p = 0.05$			$p = 0.1$		
	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI	% Bias	RMSE	95% CI
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 50, 0.6, 0.6, 1.0)$									
$\beta_0$	-0.0134	0.6142	0.9438	-0.1277	0.6176	0.9458	-0.3243	0.6333	0.9383
$\beta_1$	0.1067	0.3616	0.9436	-0.1674	0.3675	0.9445	-0.6227	0.3767	0.9364
$\beta_2$	0.1254	0.0781	0.9439	-0.1876	0.0793	0.9449	-0.5389	0.0823	0.9390
$\sigma_v^2$	0.4081	0.1527	0.9393	-0.0376	0.1547	0.9361	-0.9950	0.1582	0.9254
$\theta_1$	14.8354	0.0229	0.9616	14.7925	0.0231	0.9606	15.7662	0.0233	0.9647
$\theta_2$	4.5703	0.0279	0.9547	5.1820	0.0284	0.9499	4.6815	0.0278	0.9558
$\lambda$	-0.1059	0.0050	0.9442	-0.2997	0.0051	0.9467	-0.6302	0.0053	0.9401
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (20, 50, 1.2, 1.2, 1.5)$									
$\beta_0$	-0.0115	0.8184	0.9486	-0.3627	0.8216	0.9426	-0.7198	0.8681	0.9343
$\beta_1$	0.2847	0.4630	0.9495	-0.6191	0.4612	0.9405	-1.4844	0.4850	0.9290
$\beta_2$	0.1781	0.1011	0.9483	-0.5931	0.1016	0.9462	-1.5009	0.1066	0.9296
$\sigma_v^2$	1.0174	0.4334	0.9408	-0.7036	0.4277	0.9296	-2.3781	0.4429	0.9161
$\theta_1$	15.6651	0.0139	0.9563	15.3102	0.0133	0.9615	15.4852	0.0136	0.9576
$\theta_2$	4.4971	0.0167	0.9504	4.9993	0.0169	0.9526	4.5576	0.0166	0.9522
$\lambda$	-0.1023	0.0064	0.9507	-0.7247	0.0065	0.9482	-1.3650	0.0069	0.9381
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 50, 0.2, 0.2, .5)$									
$\beta_0$	-0.0028	0.1985	0.9512	-0.0207	0.2005	0.9475	-0.0387	0.2030	0.9481
$\beta_1$	0.0117	0.1200	0.9511	-0.0359	0.1212	0.9469	-0.0851	0.1219	0.9482
$\beta_2$	-0.0069	0.0255	0.9500	-0.0432	0.0260	0.9483	-0.0774	0.0257	0.9473
$\sigma_v^2$	0.0226	0.0140	0.9483	-0.0643	0.0144	0.9377	-0.2095	0.0145	0.9397
$\theta_1$	5.2749	0.0254	0.9514	5.6266	0.0261	0.9500	5.6490	0.0260	0.9518
$\theta_2$	5.2092	0.0255	0.9556	5.4026	0.0255	0.9562	5.4178	0.0254	0.9552
$\lambda$	-0.0126	0.0017	0.9506	-0.0439	0.0017	0.9471	-0.0804	0.0017	0.9472
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 50, 0.6, 0.6, 1.0)$									
$\beta_0$	-0.0073	0.3755	0.9509	-0.1349	0.3879	0.9458	-0.2528	0.3866	0.9399
$\beta_1$	0.0425	0.2202	0.9489	-0.2558	0.2272	0.9476	-0.5692	0.2236	0.9407
$\beta_2$	0.0394	0.0477	0.9490	-0.2838	0.0498	0.9430	-0.5486	0.0485	0.9422
$\sigma_v^2$	0.1118	0.0928	0.9467	-0.4334	0.0945	0.9436	-1.0840	0.0935	0.9333
$\theta_1$	5.0095	0.0120	0.9527	5.6869	0.0121	0.9532	5.2776	0.0122	0.9521
$\theta_2$	5.0367	0.0121	0.9511	5.1328	0.0121	0.9541	4.7199	0.0120	0.9543
$\lambda$	-0.0351	0.0030	0.9495	-0.2486	0.0032	0.9480	-0.4651	0.0031	0.9430
$(N, T, \sigma_\mu, \sigma_\eta, \sigma_v) = (50, 50, 1.2, 1.2, 1.5)$									
$\beta_0$	-0.0193	0.5257	0.9511	-0.3972	0.5366	0.9436	-0.7238	0.5484	0.9311
$\beta_1$	0.0610	0.2869	0.9502	-0.8712	0.2912	0.9417	-1.7549	0.3040	0.9226
$\beta_2$	0.0243	0.0643	0.9506	-0.9005	0.0659	0.9420	-1.7896	0.0672	0.9263
$\sigma_v^2$	0.3188	0.2646	0.9478	-1.5364	0.2650	0.9347	-3.2674	0.2758	0.9070
$\theta_1$	5.4190	0.0071	0.9491	5.1391	0.0069	0.9546	5.1740	0.0070	0.9498
$\theta_2$	5.0622	0.0069	0.9515	5.2646	0.0070	0.9558	4.9899	0.0070	0.9497
$\lambda$	-0.0664	0.0040	0.9520	-0.7234	0.0041	0.9471	-1.3479	0.0043	0.9275

## 6. Conclusions

A flexible random effects model is developed that is shown to be more suitable in modeling the private production and workers' wage than the traditional models with linear or loglinear functional form. Clearly, this model can also be applied to model other economic activities such as firm size, health expenditure, consumption, demand, etc. A simple computational device is given, which makes the handling of large panel data feasible using a personal computer. More appropriate one-sided LM tests are given for testing the random effects and functional form jointly or individually.

Further extensions of the model in specification and estimation are both possible and interesting, such as the inclusion of serial correlation and/or spatial correlation, use of quasi-maximum likelihood estimation (QMLE) method, etc. The QMLE should be useful in the context that exact normality cannot be achieved by transformation, which is typical when using the Box-Cox transformation.

## References

- [1] Abrevaya, J. (1999). Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics*, **93**, 203-228.
- [2] Arellano, M. (2003). *Panel Data Econometrics*. Oxford: Oxford University Press.
- [3] Baltagi, B. H. (1997). Testing linear and log-linear error components regression against Box-Cox alternatives. *Statistics & Probability Letters* **33**, 63-68.
- [4] Baltagi, B. H. (2001). *Econometric Analysis of Panel Data*. New York: John Wiley & Sons, Ltd.
- [5] Baltagi, B. H. and Li, Q. (1992). A monotonic property for iterative GLS in the two-way random effects model. *Journal of Econometrics* **53**, 45-51.
- [6] Baltagi, B. H. and Pinnoi, N. (1995). Public capital stock and state productivity growth: Further evidence from an error components model. *Empirical Economics*, **20**, 251-239.
- [7] Baltagi, B. H., Chang, Y. J. and Li, Q. (1992). Monte Carlo results on several new and existing tests for the error component model. *Journal of Econometrics* **54**, 95-120.
- [8] Box, G.E.P. and Cox, D.R. (1964). An Analysis of Transformations (with discussion). *J. R. Statist. Soc. B* **26**, 211-46.
- [9] Breusch, T. S. (1987). Maximum likelihood estimation of random effects model. *Journal of Econometrics* **36**, 383-389.
- [10] Breusch, T. S. and Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies* **47**, 239-253.
- [11] Davidson, R. and MacKinnon, J. G. (1985). Testing linear and log-linear regressions against Box-Cox alternatives. *Canadian Journal of Economics* **18**, 499-517.
- [12] Davidson, R. and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford: Oxford University Press.

- [13] Frees, E. W. (2004). *Longitudinal and Panel Data*. Cambridge: Cambridge University Press.
- [14] Giannakas, K., Tran, K. C., and Tzouvelekas, V. (2003). On the choice of functional form in stochastic frontier modeling. *Empirical Economics*, **28**, 75-100.
- [15] Greene, W. H. (2000). *Econometric Analysis, 4th ed.* Singapore: Prentice-Hall Pte Ltd.
- [16] Gourieroux, C., Holly, A. and Monfor, A. (1982). Likelihood ratio test, Wald test and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica* **50**, 63-80.
- [17] Holtz-Eakin, D. (1994). Public-sector capital and the productivity puzzle. *Review of Economics and Statistics*. **76**, 12-21.
- [18] Hsiao, C. (2003). *Analysis of Panel Data*. Cambridge: Cambridge University Press.
- [19] Polachek, S. W. and Yoon, B. J. (1996). Panel estimates of a two-tiered earnings frontier. *Journal of Applied Econometrics*. **11**, 169-178.
- [20] Rogers, A. J. (1986). Modified Lagrange multiplier tests for problems with one-sided alternatives. *Journal of Econometrics* **31**, 341-361.
- [21] Munnell, A. (1990). Why has productivity growth declined? Productivity and public investment. *New England Economic Review*, **Jan./Feb.**, 3-22.
- [22] Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605-610.
- [23] Silvapulle, M. J. and Sivapulle, P. (1995). A score test against one-sided alternatives. *Journal of the American Statistical Association* **90**, 342-349.
- [24] Sutton, J. (1997). Gibrat's legacy. *Journal of Economic Literature*, **35**, 40-59.
- [25] Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, **59**, 254-262.



- [26] Wolak, F. (1991). The local nature of hypothesis tests involving inequality constraints in nonlinear models. *Econometrica* **59**, 981-995.
- [27] Yang Z. L. and Abeysinghe, T. (2003). A score test for Box-Cox functional form. *Economics Letters* **79**, 107-115.