Estimating Dynamic Panel Data Models with Measurement Errors with an Application to School Evaluation based on Student Test Scores

Jie Chen* Shawn Ni† Michael Podgursky ‡

Abstract

A key difficulty in drawing inference of school effect from student test score gains is the fact that test scores are noisy measurements of students' academic achievements. In this paper we examine competing inferential methods of dynamic panel data models using noisy data. In particular, we consider a score-level model where the score of the current period depends on the score of the previous period. We compare Monte Carlo simulation results of estimators of this model and estimates using the student test score data of Missouri.

KEY WORDS: value-added, dynamic panel data, noisy data

1. Primary Subhead

In recent years, voluminous research has been conducted to estimate the "value-added" by school teachers, based on the gains in student scores in standard tests. The ultimate goal of this line of inquiry is to help administrators evaluate teachers and schools, redirect resources, and create incentive schemes to improve the K-12 education. It is widely acknowledged (e.g., McCaffrey, et. al. 2003 monograph) that the current literature is still limited and the scholarly research is not reliable enough for high stake public policy making. A key difficulty in drawing inference of teacher effect from student test score gains is the fact that test scores are noisy measurements of students' academic achievements. The design of the tests, the randomness in student's performance on the day of the test, etc. Kane and Staiger (2002a) estimate that forty percent of the variations in test scores are noise. Chay et al. (2005) show that ignoring the noise in data can result in substantial bias in teacher evaluations.

In practice, the number of observations on each student is too small to justify the use of large sample (T) theory. The finite sample inference of such models is far from trivial. In the absence of measurement errors, it is well known that the serial correlation in the dependent variable (test score) induces finite sample bias of OLS in the autoregressive model(see e.g., MacKinnon and Smith 1998). For fixed effect dynamic panel data models, the OLS is inconsistent as the number of student goes to infinity (see, e.g., Nickell 1981 and surveys by Chamberlain 1984 and Arellano 2003). The measurement errors in the level of test scores further complicate the inferential problem. The measurement errors in the level render a spurious negative serial correlation in score gains and inconsistency of OLS of the regression using observed test scores (for large T). Finite sample properties of dynamic panel simulated with the focus on near unit root cases (e.g., Kiviet 1995 and Phillips and Sul 2003) show that the bias can be significant.

In this paper we examine competing inferential methods of dynamic panel data models using noisy data. In particular, we consider a score-level model where the score of the current period depends on the score of the previous period, and a score-gain model where the score gain of the current period depends on the score gain of the previous period. We compare Monte Carlo simulation results of four estimators of these models. As a benchmark, we first consider the OLS and GLS of pooled regression or fixed effects regression, ignoring the measurement errors.

We then consider Generalized Method of Moments (GMM) estimator of fixed effects dynamic panel model with measurement errors. Without making distributional assumptions on the model, the basic approach to measurement errors is through instrumental variables (see Altonji and Siow (1987) and Dynan (2000) for such approach to noisy data in income and consumption expenditure). The parameters of interest are estimated through two stage least squares or (GMM), where both depend on the instruments employed. There are several difficulties with the instrumental variables approach. First, in the regression with the lags of test score, the period t equation involves measurement errors of dependent in period $t-1$, and the instruments need to be twice lagged. For relatively short panels, using lagged variables to serve as instruments exacerbates the shortage of data. Second, instruments are often weak and lead to inefficient estimation. Third, the finite sample distributions of estimators are not analytical.

If the distribution of the error is Gaussian, then Maximum Likelihood Estimation (MLE) based on the Kalman filter will produce more efficient estimates than instrumental variables methods. The third approach to score model is writing it a state space model. Kane and Staiger (2001) discussed the motivation for filtering out the noise in test score and show the approach improves forecasting performance. They used Kalman filter to evaluate the likelihood and estimated the model using MLE. The Kalman filter approach to the state space model has several drawbacks. The resultant likelihood is complicated,

*Jie Chen, Hong Kong and Shanghai Banking Corporation.
†Shawn Ni, Department of Economics, University of Missouri-Columbia.
‡Michael Podgursky, Department of Economics, University of Missouri-Columbia.

which does not permit analytical finite sample inference. In addition, although a researcher can either assume the variance of measurement errors or estimate it, there is no flexibility to model the uncertainty about it.

The last approach we consider is a Bayesian approach, which has several advantages. (i) It is often the case that the researcher has some information on the nature of measurement error. Such information can be incorporated in the prior and improves the precision of the estimates. (ii) The fundamental difference between the Bayesian inference and the frequentist inference discussed above is that the latter is conditional on the true parameters and the former is conditional on the sample observed. The presence of the lagged dependent variable in regressors makes the finite sample frequentist inference difficult. Because the Bayesian inference is conditioning on the finite sample of data, the posterior of the regression coefficients is no more complicated than that of the simple regression model. In the Bayesian framework the measurement errors can be viewed as nuisance parameters and easily integrated out from the joint posterior. The joint posterior of the thousands of parameters in this study can be derived from the likelihood and prior but the marginal posterior of each parameter is drawn numerically through Monte Carlo simulations. If the marginal posteriors are not standard, we can draw conditional posteriors of parameters (fixing the data and the rest of the parameters) from standard distributions, the Markov Chain will converge to the joint posterior, and the marginal distributions of a parameter of interest follows from averaging out other parameters. (iii) Unlike the OLS or GMM estimation that take the first difference in data, the Bayesian inference makes use of the whole data set. The difference in data utilization is important if the sample period is short, as is the case for most applications concerning student test scores. (iv) A key computational advantage of the Bayesian simulation over the state space model with Kalman filter is that no nonlinear optimization is involved here. It is known that nonlinear optimization over a large number of parameters for MLE can result in local optimum and produce unstable estimates with a minor change in instruments or starting value of the optimization algorithm.

To compare the finite sample performance of these estimators, we generate 1000 sets of data from dynamic panel of short panels with $T = 5$ and different sample sizes N. We compare several types of OLS, GMM, as well as Bayesian estimates and find the Bayesian estimates show smaller bias and mean squared errors than the OLS and GMM estimates.

We also compare estimates discussed above using the student test score data of St Louis school district. We find substantial difference in estimated school effects under competing estimators. The disparity in the estimates is not surprising given the differences in estimates we report in the numerical examples. We conclude that it is useful for researchers to check robustness of estimates of value-added models to the estimation method when using data of noisy short panels.

2. Estimators of Dynamic Panels with Measurement Errors

We consider two types of models on value-added. The number of observations in the panels are: N students with T observations in score level. Let S_{it} be the score of unit i in period t, $(t = t_i, t_i + 1, \cdots, T_i)$ subject to measurement error τ_{it}, $\tau_{it} \sim N(0, \kappa^2)$. The observed score is:

$$S^*_{i,t} = S_{i,t} + \tau_{i,t}. \tag{1}$$

First, true period $t+1$ score depends on period t score and covariants. We start with the framework that the gain score of unit i is

$$S_{i,t} = \phi S_{i,t-1} + \mu_i + \mathbf{X}'_{i,t}\boldsymbol{\theta} + \epsilon_{i,t}, \tag{2}$$

or $\epsilon_{i,t} \sim N(0, \sigma_i^2)$. Here $\mathbf{X}_{i,t}$ is a p-vector of control variables, including student dermatographies, teacher effect, school effect; $\epsilon_{i,t}$ is a normal error with variance σ_i^2. μ_i is the student i fixed effect. Substituting (1) to (2), we have a regression model on observed variables

$$S^*_{i,t} = \phi S^*_{i,t-1} + \mu_i + \mathbf{X}'_{i,t}\boldsymbol{\theta} + \xi_{i,t}, \tag{3}$$

where $\xi_{i,t} = \epsilon_{i,t} - \phi\tau_{i,t-1} + \tau_{i,t}$.

Denote $n_i = T_i - t_i$,

$$\mathbf{S^*_i} = \begin{pmatrix} S^*_{i,t_i} \\ S^*_{i,t_{i+1}} \\ \vdots \\ S^*_{i,T_i} \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_N \end{pmatrix},$$

$$\mathbf{W}_i = \begin{pmatrix} S^*_{i,t_i-1} & X_{i,t_i} \\ S^*_{i,t_i} & X_{i,t_i+1} \\ \vdots & \vdots \\ S^*_{i,T_i-1} & X_{i,T_i} \end{pmatrix}, \ \mathbf{i}_j = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{j\times 1},$$

$$\mathbf{D} = \begin{pmatrix} \mathbf{i}_{n_1} & 0 & \cdots & 0 \\ 0 & \mathbf{i}_{n_2} & 0 & 0 \\ 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \mathbf{i}_{n_N} \end{pmatrix}, \ \boldsymbol{\xi}_i = \begin{pmatrix} \xi_{i,t_i} \\ \xi_{i,t+1_i} \\ \vdots \\ \xi_{i,T_i} \end{pmatrix}.$$

Stacking up the student-level data, we denote

$$\mathbf{S}^* = \begin{pmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_N \end{pmatrix}, \ \mathbf{W} = \begin{pmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_N \end{pmatrix}, \ \ \boldsymbol{\beta} = (\phi, \boldsymbol{\theta}), \boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1 \\ \vdots \\ \boldsymbol{\xi}_N \end{pmatrix}.$$

2.1 Ordinary Least Square Estimator

The OLS or GLS estimates of dynamic panel data model (3) with measurement errors are inconsistent, because the error term of period t, $\xi_{i,t} = \epsilon_{i,t} - \phi\tau_{i,t-1} + \tau_{i,t}$, is correlated with the regressor $S^*_{i,t-1}$. The extent of the inconsistency depends on the data-generating parameters. We will explore this quantitatively through numerical simulations.

We can rewrite the regression (3) in matrix form

$$\mathbf{S}^* = \mathbf{D}\boldsymbol{\mu} + \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\xi}. \tag{4}$$

Because the number of students is much larger than the number of schools, instead of estimating the above model with student fixed effect dummy, we can either estimate the partitioned regression (which is equivalent to subtract the score by the student-average) or take the first difference of the score. The latter creates serial correlation in the error and requires at least three observations from each student while the former requires at least two.

Taking the first difference in (3) to eliminate the student fixed effect, we have

$$\begin{aligned} S^*_{i,t} - S^*_{i,t-1} &= \phi(S^*_{i,t-1} - S^*_{i,t-2}) + (\mathbf{X}'_{i,t} - \mathbf{X}'_{i,t-1})\boldsymbol{\theta} \\ &\quad + \xi_{i,t} - \xi_{i,t-1}, \end{aligned}$$

Denote

$$\Delta\mathbf{S}^*_i = \begin{pmatrix} S^*_{i,t_i+1} - S^*_{i,t_i} \\ S^*_{i,t_i+2} - S^*_{i,t_i+1} \\ \vdots \\ S^*_{i,T_i} - S^*_{i,T_i-1} \end{pmatrix}_{(n_i-1)\times 1},$$

$$\Delta\mathbf{W}_i = \begin{pmatrix} S^*_{i,t_i} - S^*_{i,t_i-1} & X'_{i,t_i} - X'_{i,t_i-1} \\ S^*_{i,t_i+1} - S^*_{i,t_i} & X'_{i,t_i+1} - X'_{i,t_i} \\ \vdots & \vdots \\ S^*_{i,T_i-1} - S^*_{i,T_i-2} & X'_{i,T_i} - X'_{i,T_i-1} \end{pmatrix},$$

$$\Delta\boldsymbol{\xi}_i = \begin{pmatrix} \xi_{i,t_i} - \xi_{i,t_i-1} \\ \xi_{i,t+1_i} - \xi_{i,t_i} \\ \vdots \\ \xi_{i,T_i} - \xi_{i,T_i-1} \end{pmatrix}_{(n_i-1)\times 1}.$$

The matrix form regression on the first difference is

$$\Delta\mathbf{S}^* = (\Delta\mathbf{W})\boldsymbol{\beta} + \Delta\boldsymbol{\xi}. \tag{5}$$

The OLS estimator is given by

$$\hat{\boldsymbol{\beta}}_1 = [(\Delta\mathbf{W})'(\Delta\mathbf{W})]^{-1}(\Delta\mathbf{W})'(\Delta\mathbf{S}^*). \tag{6}$$

The OLS estimator of the first-differenced data is inconsistent because the error is correlated with the regressor for two reasons, the correlation between measurement error in $\xi_{i,t}$ and $S^*_{i,t-1}$ mentioned earlier and the correlation between $\Delta\xi_{i,t}$ and regressor $\Delta S^*_{i,t}$ induced by the first difference.

OLS estimate with student-fixed effect can be expressed as

$$\hat{\boldsymbol{\beta}}_2 = (\mathbf{W}'\mathbf{M}\mathbf{W})^{-1}\mathbf{W}'\mathbf{M}\mathbf{S}^*. \tag{7}$$

where

$$\begin{aligned} \mathbf{M} &= \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}' \\ &= \begin{pmatrix} \mathbf{I}_{n_1} - \frac{\mathbf{i}_{n_1}\mathbf{i}'_{n_1}}{n_1} & 0 & \cdots & 0 \\ 0 & \mathbf{I}_{n_2} - \frac{\mathbf{i}_{n_2}\mathbf{i}'_{n_2}}{n_2} & 0 & 0 \\ 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & \mathbf{I}_{n_N} - \frac{\mathbf{i}_{n_N}\mathbf{i}'_{n_N}}{n_N} \end{pmatrix} \end{aligned}$$

This OLS estimator is inconsistent as the number of student N goes to infinity as shown by Nickell (1982). The presence of measurement errors makes the estimator inconsistent with respect to N and time period T.

2.2 GMM Estimator

Note that $\xi_{i,t} - \xi_{i,t-1}$ is correlated with $S^*_{i,t-2}$. Valid instruments of the lags of $S^*_{i,t}$ must be $S^*_{i,t-j}$ with $j = 3, \cdots, t-3$. As in Arellano and Bond(1991), we apply instrument $Z_{i,t}$, the moment condition is

$$\mathbf{E}[Z_{i,t}(\xi_{i,t} - \xi_{i,t-1})] = 0, \tag{8}$$

for $i = 1, \cdots, N$ and $t = t_i + 3, \cdots, T_i$. The moment conditions can be written in a matrix form

$$\mathbf{E}[\mathbf{Z}'_i(\Delta\xi_i)] = 0.$$

$\mathbf{Z}_j$ is the jth of $n_i - 2$ rows, $(S^*_{i,t_i}, \cdots, S^*_{i,t_i+j-1})$ $(j = 1, .., T_i - t_i - 2)$. Stacking up the instruments we have a matrix $\mathbf{Z}$. With a weighting matrix $\mathbf{H}$, the GMM estimator $\boldsymbol{\beta}$ minimizes

$$\left(\frac{1}{N}\sum_i \mathbf{Z}_i\Delta\xi'_i\right)\mathbf{H}\left(\frac{1}{N}\sum_i \mathbf{Z}'_i\Delta\xi_i\right).$$

The estimator is given by

$$\begin{aligned} &\hat{\boldsymbol{\beta}}_{GMM1} \\ &= [\sum_{i=1}^{n}(\Delta\mathbf{W}'_i)\mathbf{Z}_i\mathbf{H}\mathbf{Z}'_i(\Delta\mathbf{W}_i)]^{-1}[\sum_{i=1}^{n}(\Delta\mathbf{W}'_i)\mathbf{Z}_i\mathbf{H}\mathbf{Z}'_i(\Delta\mathbf{S}^*_i)]. \end{aligned}$$

The weighting matrix that minimizes the asymptotic variance of the estimator is

$$\hat{\mathbf{H}} = [(1/N)\sum_{i=1}^{N}\mathbf{Z}'_i(\widehat{\Delta\xi_i})(\widehat{\Delta\xi_i})'\mathbf{Z}_i]^{-1}, \tag{9}$$

where $\widehat{\Delta\xi_i}$ is the first difference of the residuals given a consistent estimator $\hat{\boldsymbol{\beta}}$. In a widely-used two step procedure of estimation, we first set $\mathbf{H}$ as the identity matrix and derive the first-step $\hat{\boldsymbol{\beta}}_{GMM1}$ from (9). Then we plug $\hat{\boldsymbol{\beta}}_{GMM1}$ into (9) to calculate the optimal weighting matrix $\mathbf{H}$ and plug the new weighting matrix in (9) to derive the second-step GMM estimate of $\boldsymbol{\beta}$.

An alternative to first differencing GMM estimator is given by Arellano and Bover (1995) based on forward orthogonal deviations. Denote the number of observation of each unit by n. The estimator is obtained by premultiplying the matrix form regression (3) by

$$\begin{aligned} \mathbf{A} &= diag[\frac{n-1}{n}, \cdots, \frac{1}{2}]^{\frac{1}{2}} \\ &\times \begin{pmatrix} 1 & -\frac{1}{n-1} & -\frac{1}{n-1} & \cdots & -\frac{1}{n-1} & -\frac{1}{n-1} & -\frac{1}{n-1} \\ 0 & 1 & -\frac{1}{n-2} & \cdots & -\frac{1}{n-2} & -\frac{1}{n-2} & -\frac{1}{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{pmatrix} \end{aligned}$$

for, $i = 1, \cdots, N$

$$\mathbf{A}\mathbf{S}^*_i = \mathbf{A}\mathbf{D}\boldsymbol{\mu}_i + \mathbf{A}\mathbf{W}_i\boldsymbol{\beta} + \mathbf{A}\boldsymbol{\xi}_i. \tag{10}$$

By construction, elements of each row of $\mathbf{A}$ add to zero, hence the fixed effects are eliminated and we have

$$\mathbf{A}\mathbf{S}_i^* = \mathbf{A}\mathbf{W}_i\boldsymbol{\beta} + \mathbf{A}\boldsymbol{\xi}_i. \tag{11}$$

$\mathbf{A}\boldsymbol{\xi}_i$ is a vector of weighted errors, the tth element of which only depends on the errors of tth period and onward. The moment condition takes the form

$$\mathbf{E}(\mathbf{Z}_i'\mathbf{A}\boldsymbol{\xi}_i) = \mathbf{E}[\mathbf{Z}_i'\mathbf{A}(\mathbf{S}_i^* - \mathbf{W}_i\boldsymbol{\beta})] = 0.$$

The GMM estimator is

$$\hat{\boldsymbol{\beta}}_{GMM2} = [\sum_{i=1}^{N}(\mathbf{W}_i'\mathbf{A}'\mathbf{Z}_i)\mathbf{H}(\mathbf{Z}_i'\mathbf{A}\mathbf{W}_i)]^{-1}$$
$$\times[\sum_{i=1}^{N}(\mathbf{W}_i'\mathbf{A}'\mathbf{Z}_i)\mathbf{H}(\mathbf{Z}_i'\mathbf{A}\mathbf{S}_i^*)],$$

where the weighting matrix $\mathbf{H}$ is chosen in two steps. The one-step choice for $\mathbf{H}$ is

$$\mathbf{H} = \left\{\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{A}\mathbf{A}'\mathbf{Z}_i\right\}^{-1}.$$

A two-step choice is

$$\mathbf{H} = \left\{\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{A}\hat{\boldsymbol{\xi}}_i\hat{\boldsymbol{\xi}}_i'\mathbf{A}'\mathbf{Z}_i\right\}^{-1},$$

where $\hat{\boldsymbol{\xi}}_i$ is the step-one residual.

2.3 Bayesian Model

An advantage of the state-space and the following Bayesian approach is that all observations are utilized. Denote the observed data as the true data plus a noise τ:

$$S_{it}^* = S_{it} + \tau_{it},$$

where $\tau_{it} \sim N(0, \kappa^2)$. Our prior for κ^2 is $IG(p_\kappa, v_\kappa)$.

Denote

$$\boldsymbol{\tau}_i = \begin{pmatrix} \tau_{i,t_i-1} \\ \tau_{i,t_i} \\ \vdots \\ \tau_{i,T_i} \end{pmatrix}. \tag{12}$$

We assume

$$\boldsymbol{\tau}_i = \mathbf{S}_i^* - \mathbf{S}_i \sim N(\mathbf{0}, \kappa^{\mathbf{2}}\mathbf{I}_{\mathbf{n_i+1}}).$$

Define $\rho_{it}' = (\mu_i + \mathbf{X}_{it}'\boldsymbol{\theta})$. We rewrite model (2) as

$$\mathbf{B_i S_i} - \rho_{\mathbf{i}} \sim \mathbf{N}(\mathbf{0}, \sigma_{\mathbf{i}}^{\mathbf{2}}\mathbf{I}_{\mathbf{n_i}}),$$

with

$$\mathbf{B}_i = \begin{pmatrix} -\phi & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\phi & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddot{} & \ddot{} & \ddot{} & 0 \\ 0 & 0 & .. & 0 & 0 & -\phi & 1 \end{pmatrix}_{(n_i)\times(n_i+1)},$$

$$\boldsymbol{\rho}_i = \begin{pmatrix} \rho_{i,t_i} \\ \vdots \\ \rho_{i,T_i} \end{pmatrix}$$

The posterior is

$$\pi(\boldsymbol{\beta}, \kappa, \mu_i, \boldsymbol{\sigma}_i, \mathbf{S}_i(i = 1, \cdots, N) \mid \mathbf{D})$$
$$\propto \kappa^{-2(p_\kappa+1)}\exp(-\frac{v_\kappa}{\kappa^2})$$
$$exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})\right\}$$
$$\times\prod_{i=1}^{N}\Big[\sigma_i^{-2(p_\sigma+1)}\exp(-\frac{v_0}{\sigma_i^2})(\sigma_i^2)^{-\frac{n_i}{2}exp\left\{-\frac{(\mu_i-\bar{\mu}_i)^2}{2\sigma_\mu^2}\right\}}$$
$$exp\left\{-\frac{(\mathbf{S}_i^* - \mathbf{S}_i)'(\mathbf{S}_i^* - \mathbf{S}_i)}{2\kappa^2}\right\}$$
$$\times\exp\left\{-\frac{(\mathbf{B}_i\mathbf{S}_i - \boldsymbol{\rho}_i)'(\mathbf{B}_i\mathbf{S}_i - \boldsymbol{\rho}_i)}{2\sigma_i^2}\right\}\Big].$$

It follows that the conditional posterior of the true data is

$$\mathbf{S}_i \sim N(\hat{\mathbf{S}}_i, \mathbf{Q}_i), \tag{13}$$

where $\mathbf{Q}_i = (\kappa^{-2}\mathbf{I} + \mathbf{B}_i'\mathbf{B}_i\sigma_i^{-2})^{-1}$, $\hat{\mathbf{S}}_i = \mathbf{Q}_i(\kappa^{-2}\mathbf{S}_i{}^* + \mathbf{B}_i'\boldsymbol{\rho}_i\sigma_i^{-2})$.

Denote $h_{it} = S_{it} - \phi S_{it-1} - \mathbf{X}_{it}'\boldsymbol{\theta}$. The posterior for student fixed effects is given by

$$\pi(\mu_i \mid \boldsymbol{\beta}, \sigma_i, \mathbf{D}) \propto \exp\left\{-\sum_{t=t_i}^{T_i}\frac{(h_{it} - \mu_i)^2}{2{\sigma_i}^2}\right\}exp\left\{-\frac{(\mu_i - \bar{\mu}_i)^2}{2\sigma_\mu^2}\right\}$$
$$\propto N([\sigma_\mu^{-2} + {\sigma_i}^{-2}(n_i+1)]^{-1}[\sigma_\mu^{-2}\bar{\mu}_i + {\sigma_i}^{-2}\sum_{t=t_i}^{T_i}h_{it}],$$
$$[\sigma_\mu^{-2} + {\sigma_i}^{-2}(n_i+1)]^{-1}). \tag{14}$$

The posterior of the slope coefficients of the regression is

$$\pi(\boldsymbol{\beta} \mid \boldsymbol{\mu}, \sigma, \mathbf{D}) \propto exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})'\boldsymbol{\Omega}^{-1}(\boldsymbol{\beta} - \bar{\boldsymbol{\beta}})\right\}$$
$$\exp\left\{-\sum_{i=1}^{N}\sum_{t=t_i}^{T_i}\frac{(S_{it} - \mu_i - W_{it}'\boldsymbol{\beta})^2}{2{\sigma_i}^2}\right\}$$
$$\propto N([\boldsymbol{\Omega}^{-1} + \sum_{i=1}^{N}\sum_{t=t_i}^{T_i}{\sigma_i}^{-2}W_{it}'W_{it}]^{-1}[\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{\beta}}$$
$$+\sum_{i=1}^{N}\sum_{t=t_i}^{T_i}{\sigma_i}^{-2}W_{it}'(S_{it} - \mu_i)],\ [\boldsymbol{\Omega}^{-1} + \sum_{i=1}^{N}\sum_{t=t_i}^{T_i}{\sigma_i}^{-2}W_{it}'W_{it}]^{-1}). \tag{15}$$

The conditional posterior of the noise variance is

$$\kappa^2 \mid \mathbf{S}_i^*, \mathbf{S}_i; (i = 1, \cdots, N) \sim IG(p_\kappa + \sum_{i=1}^{N}\frac{n_i+1}{2},$$
$$v_\kappa + \sum_{i=1}^{N}\frac{(\mathbf{S}_i^* - \mathbf{S}_i)'(\mathbf{S}_i^* - \mathbf{S}_i)}{2}). \tag{16}$$

The conditional posterior of the regression error variance is

$$\pi(\sigma_i^2 \mid \beta, \mu_i, \boldsymbol{\theta}, \mathbf{S}_i, \mathbf{D}) \propto (\sigma_i^2)^{-(p_\sigma+1)}$$

$$exp(-\frac{v_0}{\sigma_i^2})(\sigma_i^2)^{-(n_i)} exp\left\{-\frac{(\mathbf{B}_i S_i - \boldsymbol{\rho}_i)'(\mathbf{B}_i S_i - \boldsymbol{\rho}_i)}{2\sigma_i^2}\right\}$$
$$\propto IG(p_\sigma + \frac{n_i}{2}, v_\sigma + \frac{(\mathbf{B}_i S_i - \boldsymbol{\rho}_i)'(\mathbf{B}_i S_i - \boldsymbol{\rho}_i)}{2}), \quad (17)$$

The conditional posteriors suggests a Gibbs sampling MCMC algorithm: In cycle k, with $(\mathbf{c}_i^{(k-1)}, \boldsymbol{\beta}_i^{(k-1)}, \boldsymbol{\theta}^{(k-1)}, \sigma_i^{(k-1)}, \boldsymbol{\kappa}^{(k-1)})$ drawn already,

(1) Draw $(\mathbf{S}_i^{(k)} \mid \sigma_i^{(k-1)}, \boldsymbol{\beta}^{(k-1)}, \mu_i, \mathbf{D})$ $(i = 1, \cdots, N))$ from the normal distribution (13).

(2) Draw $\mu_i{}^{(k)} | \mathbf{S}_i^{(k)}, \boldsymbol{\beta}^{(k-1)}, \sigma_i^{(k-1)}, \mathbf{D}$ $(i = 1, \cdots, N))$ from the normal distribution (14).

(3) Draw $(\boldsymbol{\beta}^{(k)} \mid \mathbf{S}^{(k)}, \boldsymbol{\mu}^{(k)}, \sigma^{(k-1)}, \mathbf{D})$ from the normal distribution (15).

(4) Draw $\kappa^{2^{(k)}} \mid \mathbf{S}^{(k)}, \sigma_i^{(k-1)}, (i = 1, \cdots, N), \mathbf{D})$ from (16).

(5) Draw $\sigma_i^{2^{(k)}} \mid \mathbf{S}^{(k)}, \boldsymbol{\beta}^{(k)}, \mu_i^{(k)} (i = 1, \cdots, N), \mathbf{D})$ from (17).

We use the posterior mean as the estimator.

3. Numerical Simulation

We simulate 1000 samples of panel data based on parameters similar to the estimates from the Missouri student test score data. Specifically, $N = 1000$, $t_i = 1, T_i = 5$, $p = 1$, $\mathbf{X}_{it} = cos(0.1 \times i \times t)$, $\theta = .1$, $\mu_i = sin(i)$, $\kappa^2 = 0.2$, $\sigma_i^2 = 0.1$, $\phi = 0.9$. We obtain estimates of the parameters using two types of OLS, two types of GMM, and the Bayesian posterior mean. For Bayesian estimation, we set the prior of parameters as the data generating parameters and prior variance as unity.

The following tables show that compared with the data generating parameters $\phi = 0.9$ all estimates show downward bias, especially the estimates based on first-differenced data (OLS1 and GMM1). GMM estimators do better than OLS, for differenced data (GMM1 vs. OLS1) or undifferenced data (GMM2 vs. OLS2). The Bayesian estimates show smallest bias and mean squared errors. Intuitively, the prior shrinks the Bayesian posterior towards the mean, reducing the error of estimates.

To investigate the effect of sample size N on the estimates obtained through different methods, we increase N from 100 to 1000, and keep the rest of the parameters in the data generating model unchanged. We report the average of the estimates of 1000 generated samples and plot the estimates for these generated samples. The tables show that increasing N does not substantially affect the bias in the estimates, but the plots show that the frequency distributions become tighter compared with the case with $N = 100$.

3.1 Estimates of Schools Effects Using St Louis Data

The data set consists of 109 schools and five consecutive years (from 2000 to 2004) of Terra Nova for math of 27000 students. We allow for student fixed effects and estimate

Table 1: Estimation Result I: **t=5**, N=100, $x = cos(i * t * 0.1)$ and $\mu(i) = sin(i)$. Prior for $\mu(i) \sim N(0, 1)$

	OLS1	OLS2	GMM1	GMM2	Bayes
ϕ	.400(.042)	.732(0.026)	.546(.065)	.705(.038)	.843(.026)
θ	.074(.026)	.092(0.024)	.080(.040)	.090(.033)	.098(.035)

The table reports the average of estimates for 1000 simulated samples (with standard deviations of the estimates cross samples in the parentheses).

Table 2: Mean Squared Error(MSE) of parameters

Method	OLS1	OLS2	GMM1	GMM2	Bayes
MSE of $\widehat{\phi}$	.252	.029	.129	.040	.004
MSE of $\widehat{\theta}$	.001	.001	.002	.001	.001

The table reports the mean squared errors of estimates for 1000 simulated samples.

Table 3: Estimation Result I: **t=5**, N=1000, $x = cos(i * t * 0.1)$ and $\mu(i) = sin(i)$. Prior for $\mu(i) \sim N(0, 1)$

	OLS1	OLS2	GMM1	GMM2	Bayes
ϕ	.393(.013)	.729(.008)	.540(.021)	.699(0.012)	.843(.008)
θ	.075(.008)	.092(.007)	.082(.013)	.090(0.011)	.097(.011)

The table reports the average of estimates for 1000 simulated samples (with standard deviations of the estimates cross samples in the parentheses).

Table 4: Mean Squared Error(MSE) of parameters

Method	OLS1	OLS2	GMM1	GMM2	Bayes
MSE of $\widehat{\phi}$	.257	.029	.130	.041	.003
MSE of $\widehat{\theta}$	.001	.000	.001	.000	.000

The table reports the mean squared errors of estimates for 1000 simulated samples.

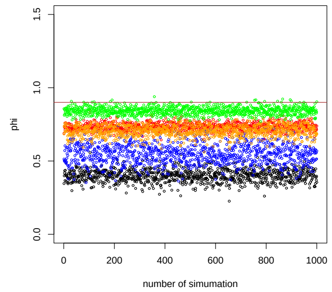

Figure 1: N=100. Data generating values $\phi = 0.9$. Black=OLS(1). Red = OLS(2). Blue= GMM(1). Orange = GMM(2). Green= Bayesian estimator.

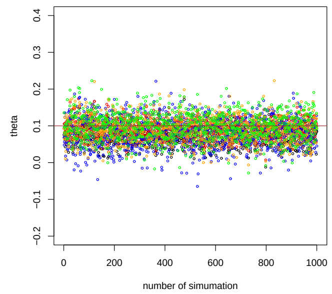

Figure 2: N=100. Data generating values $\theta = 0.1$. Black=OLS(1). Red = OLS(2). Blue= GMM(1). Orange = GMM(2). Green= Bayesian estimator.

school effects. We plot the estimated 108 school effects using OLS2, GMM2, and the Bayesian posterior mean. The plots show that the estimated school effects are quite sensitive to the estimation method. The range of the estimated school effects is about ten percent, and the range is slightly tighter under for the Bayesian estimates.

4. Summary and Comments

The Monte Carlo simulation results presented in this study shows that Bayesian estimates under plausible priors dominate OLS and GMM estimates in dynamic panel models when data are noisy and the panels are short. Estimated school effects using St Louis school district data with different estimation methods substantially differ. We conclude that when estimating value-added models with noisy test scores, it is advisable to check the robustness of the estimated school effects.

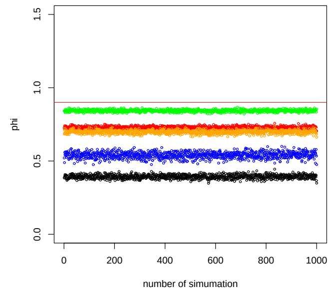

Figure 3: N=1000. Data generating values $\phi = 0.9$. Black=OLS(1). Red = OLS(2). Blue= GMM(1). Orange = GMM(2). Green= Bayesian estimator.

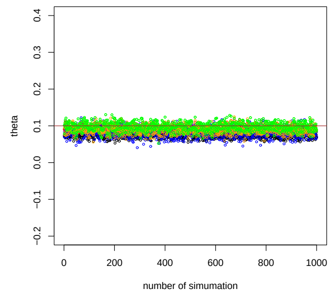

Figure 4: N=1000. Data generating values $\theta = 0.1$. Black=OLS(1). Red = OLS(2). Blue= GMM(1). Orange = GMM(2). Green= Bayesian estimator.

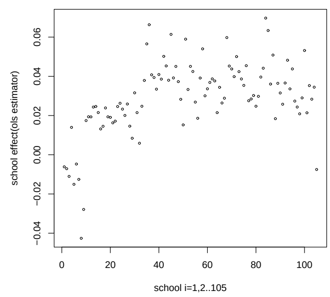

Figure 5: Estimated School Effects by OLS2 : dependent variable = log(score), 108 Schools

REFERENCES

Altonji, Joseph G. and Aloysius Siow (1987). "Testing the Response of Consumption to Income Changes with (Noisy) Panel Data," *Quarterly Journal of Economics*, 102, 293-328.

Arellano, Manuel (2003). *Panel Data Econometrics*, Oxford: Oxford University Press.

Arellano M. and O. Bover (1995). "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics*, 68, 29-51.

Baltagi, Badi. H. and Ping X. Wu (1999). "Unequally Spaced Panel Data Regressions with AR(1) Disturbance," *Econometric Theory*, 15, 814-823.

Chamberlain, Gary (1984). "Panel Data," *Handbook of Econo-*

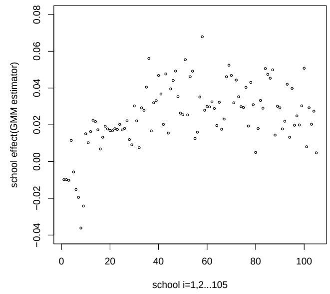

Figure 6: Estimated School Effects by GMM2 : dependent variable = log(score), 108 Schools

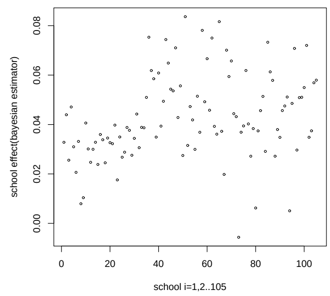

Figure 7: Estimated School Effect by Bayesian Estimates : dependent variable = log(score), 108 Schools

metrics, 11, Eds.: Zvi Griliches and Michael D. Intriligator. Amsterdam: Elsevier, 1247-1318.

Chay, Kenneth Y. ,Patrick J. McEwan, AND Miguel Urquiola (2005) "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools" *American Economic Review*, 95 1237-1258.

Dynan, Karen E. (2000). "Habit Formation in Consumer Preference: Evidence from Panel Data" *American Economic Review*, 90, 391-406.

Hamilton, James D. (1994). *Time Series Analysis,* Princeton University Press.

Kane, Thomas J. and Staiger, Douglas O. "Improving School Accountability Measures." National Bureau of Economic Research, Inc., NBER Working Papers: No. 8156, 2001.

Kane, Thomas J. and Staiger, Douglas O. (2002a) "The Promise and Pitfalls of Using Imprecise School Accountability Measures." Journal of Economic Perspectives, 16(4), pp. 91114.

Kane, Thomas J. and Staiger, Douglas O. (2002b) "Volatility in Test Scores: Implications for Test- Based Accountability Systems," in Diane Ravitch, ed., Brookings papers on education policy. Washington, DC: Brookings Institution Press, pp. 23583.

Kiviet, Jan F. (1995). "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics*, 68, 53-78.

MacKinnon James G. and Anthony A. Smith (1998). "Approximate bias correction in econometrics," *Journal of Econometrics*, 85, 205-230.

Nickell, Stephen (1981). "Biases in Dynamic Models with Fixed Effects," *Econometrica* 49, 1417-1426.

Phillips Peter C. B. and Sul, Donggyu (2003) "Dynamic panel estimation and homogeneity testing under cross section dependence," *Econometrics Journal*, 6, 217-259