GMM ESTIMATION OF LINEAR PANEL DATA MODELS WITH TIME-VARYING INDIVIDUAL EFFECTS

Seung Chan Ahn*
Department of Economics
Arizona State University, Tempe, AZ 85287, USA

Young Hoon Lee
Department of Economics
Hansung University, Seoul, South Korea

Peter Schmidt
Department of Economics
Michigan State University, E. Lansing, MI 48824, USA

Abstract

This paper considers models for panel data in which the individual effects vary over time. The temporal pattern of variation is arbitrary, but it is the same for all individuals. The model thus allows one to control for time-varying unobservables that are faced by all individuals (e.g., macro-economic events) and to which individuals may respond differently. A generalized within estimator is consistent under strong assumptions on the errors, but it is dominated by a generalized method of moments estimator. This is perhaps surprising, because the generalized within estimator is the MLE under normality. The efficiency gains from imposing second-moment error assumptions are evaluated; they are substantial when the regressors and effects are weakly correlated.

Key Words: Panel Data; Time-Varying Effects; Generalized Method of Moments; MLE.

JEL Classification Number: C23

---

*Corresponding Author. Department of Economics, Arizona State University, Tempe, AZ 85287-3806, USA. Phone: (480) 965-6574; FAX: (480) 965-0748; email: miniahn@asu.edu.

1. Introduction*

The literature on panel data often assumes observations on a large number of individuals with several observations on each individual. The fixed effects model has been widely adopted as a treatment of unobservable individual heterogeneity in such data. This model assumes that an unobservable time-invariant variable (or set of variables) is associated with each individual. (See, for example, Mundlak, 1961; MaCurdy, 1981; and Chamberlain, 1984.) The impact of the missing variable on the dependent variable is referred to as the individual effect, and one wishes to avoid the potential bias that occurs when the individual effect is correlated with the included explanatory variables. The conventional fixed-effects treatment of time-invariant individual effects has been the within estimator, which is equivalent to least squares after transformation of the data to deviations from individual means.

In this paper, we consider the case that the individual effects are time-varying. That is, we wish to accommodate time-varying unobservables that affect the dependent variable differently for different individuals, and that may be correlated with the included explanatory variables. This is possible under the assumption that the temporal pattern (though not necessarily the level) of the effects is the same for each individual.

More specifically, we will consider the model

*The first author gratefully acknowledges the financial support of the College of Business and Dean's Council of 100 at Arizona State University, the Economic Club of Phoenix, and the alumni of the College of Business. The second author was financially supported for this research by Hansung University in the year of 1999. The third author gratefully acknowledges the financial support of the National Science Foundation.

$$
\begin{aligned}
y_{it} &= X_{it}\beta + Z_i\gamma + u_{it}, \\
u_{it} &= \theta_t\alpha_i + \epsilon_{it}, \quad i = 1, \ldots, N; \quad t = 1, \ldots, T.
\end{aligned}
\tag{1}
$$

Here i denotes cross sectional unit (individual) and t denotes time. We assume that N is large while T is small, so that asymptotic properties of estimates will be derived assuming $N \to \infty$ with T fixed. The dependent variable is y_{it}, X_{it} is a 1×k vector of time-varying explanatory variables, and Z_i is a 1×g vector of time invariant regressors. The last entry of Z_i is one, so that the last parameter in γ denotes the overall intercept term. The ϵ_{it} are random noise with $E(\epsilon_{it}) = 0$. The α_i are unobservables, and θ_t is the parameter measuring the effect of α_i on y_{it} at time t. Clearly this specification implies that the temporal pattern of the effect of α_i on y_{it} is the same across different i. For identification of the model, we set $\theta_1 = 1$. (This does not involve any loss of generality, so long as $\theta_1 \neq 0$. Alternatively, we can follow Kiefer (1980a) who normalizes the model by setting $\Sigma_{t=1}^{T}\theta_t^2 = 1$.) This model includes the conventional fixed effects model as the special case that $\theta_t = 1$ for all t.

This model may be useful in several applications. First, in a model of earnings or hourly wages, the effect α_i may be the individual's unobservable talent or skill and the parameter θ_t would represent the price of the skill which is not necessarily constant over time.[1] Second, the specification (1) may be used to estimate stochastic frontier models with time-varying technical inefficiency. Lee and Schmidt (1993) have applied the model (1) to the frontier production function (efficiency measurement) problem, assuming that the temporal pattern of technical inefficiency is the same for all firms. Third, as noted above, a more common motivation would

[1]We would like to thank an Associate Editor for providing us this example.

be to control for unobservables that are not time-invariant, but whose temporal variation is the same for all individuals. A macro shock would have this feature: In a rational expectation model using panel data, the parameter θ_t could be thought of as a macro shock and the effect α_i as the effect of a shock on individual i (heterogeneous over individuals due to differences in individuals' information sets, preferences, etc). There is a large literature on estimation of rational expectations models using panel data. This literature dates back at least to Hall and Mishkin (1982) and Shapiro (1984); see Keane and Runkle (1992) for a more recent example. The equation that is estimated has an error that is interpretable as a forecast error, or macro shock, and which correspondingly cannot be averaged away over individuals. Thus the usual orthogonality conditions that reflect uncorrelatedness of the shock with variables in the information set at the time the forecast is made do not lead to consistent estimation as $N \to \infty$ with T fixed (see Hayashi, 1992). However, the model given above can accommodate such a macro shock. Further examples for the model (1) are given in Appendix A.

There are only a few previous studies of models which involve time-varying individual effects. Holtz-Eakin, Newey and Rosen (1988) consider a vector autoregressive model which allows for nonconstant individual effects. (See also Chamberlain, 1984.) They estimate the model by applying instrumental variables to the quasi-differenced autoregressive equations. Chamberlain (1992) considers a general random coefficient model which specifies only the conditional expectation of the dependent variable given exogenous variables and the individual effect. He derives an efficient generalized method of moments (GMM) estimator based on the moment conditions implied by the conditional mean specification.

This paper was motivated by the approach of Kiefer (1980a) and Lee (1991). They

considered estimation of (1) by least squares, treating both the θ_t and the α_i as fixed, and derived a concentrated least squares (CLS) estimator as the solution to an eigenvalue problem. If the ϵ_{it} are iid normal and the explanatory variables are independent of the ϵ_{it}, this estimator is the maximum likelihood estimator (MLE) and can also be interpreted as a conditional maximum likelihood estimator (CMLE). A nonstandard and noteable result is that the consistency of the CLS estimator depends on the second-moment assumptions on the ϵ_{it}. In particular, consistency requires that the ϵ_{it} be non-autocorrelated and have constant variance. Furthermore, given the strong assumptions needed for consistency, the CLS estimator turns out to be inefficient. We show in this paper that it is dominated by a GMM estimator that makes use of the first and second order moment conditions implied by exogeneity of the regressors and by nonautocorrelation and homoskedasticity of the ϵ_{it}. This is an unusual and interesting result because the "incidental parameters problem" is typically thought to be the potential inconsistency of the MLE. Neyman and Scott (1948) provide another example of a case in which the MLE is consistent but not efficient.

This paper differs from the work of Chamberlain (1992) primarily because we view the regression model as a projection model rather than one based on conditional expectations. Thus, for example, we impose conditions of no correlation between explanatory variables and errors, but we do not consider nonlinear functions of explanatory variables that may be relevant under the conditional expectation model. Instead, we focus on exploiting the moment conditions implied by restrictions on the variances and covariances of the errors, and we show that imposing these conditions can result in substantial efficiency gains. In keeping with the spirit of the projection model, we consider the unconditional variances and covariances of the errors, whereas

a generalization of the conditional expectations model would naturally focus on conditional variances and covariances, and again would generate additional nonlinear moment conditions.

The plan of the paper is as follows. Section 2 defines some notation and lists some assumptions. Section 3 considers GMM estimation based on assumptions of exogeneity of the regressors and several sets of assumptions about the second moments of the errors. Section 4 gives some results on the efficiency of these estimators. Section 5 examines the properties of the CLS estimator from a GMM perspective and shows why it is inefficient. Section 6 provides calculations of the relative asymptotic efficiency of the GMM and CLS estimators. Finally, our concluding remarks are given in Section 7.

2. Preliminaries

The model to be considered is as given in equation (1) above. Defining $\theta = (\theta_2, \ldots, \theta_T)'$ and $\xi = (1, \theta')'$, we can write the T observations for individual i as:

$$y_i = X_i\beta + e_T Z_i\gamma + u_i; \quad u_i = \xi\alpha_i + \epsilon_i, \tag{2}$$

where e_T is the T×1 vector of ones, $y_i = (y_{i1}, \ldots, y_{iT})'$, and X_i and ϵ_i are similarly defined. Denote the 1×(kT+g) vector of all of the values of the explanatory variables by $W_i = (X_{i1}, \ldots, X_{iT}, Z_i)$. We also define $\mu_w = E(W_i)$ and

$$\Sigma = E[(W_i, \alpha_i)'(W_i, \alpha_i)] = \begin{bmatrix} \Sigma_{ww} & \Sigma_{w\alpha} \\ \Sigma_{\alpha w} & \Sigma_{\alpha\alpha} \end{bmatrix}. \tag{3}$$

We first make the following **BASIC ASSUMPTIONS (BA)**:

$[W_i, \alpha_i, \epsilon_i']$ is independently and identically distributed (iid) over i. **(BA.1)**

ϵ_{it} has finite moments up to fourth order, and $E(\epsilon_{it}) = 0$. **(BA.2)**

The second moment matrix Σ is finite and nonsingular. **(BA.3)**

$E[W_i'(Z_i, \alpha_i)]$ is of full column rank. **(BA.4)**

$[W_i, \alpha_i]$ is uncorrelated with ϵ_i. **(BA.5)**

Although we treat α_i as random, our model is intended to capture the essential feature of a fixed effects model, in the usual sense that α_i is allowed to be correlated with W_i in an unrestricted way. The assumption that α_i is uncorrelated with ϵ_i is standard, and is typically motivated in the following way. If the α_i were taken to be literally fixed, we would plausibly assume a condition like "$\lim_{N\to\infty} N^{-1}\Sigma_i\alpha_i^2$ exists and is finite." But, given our assumptions on ϵ_i, this would imply that $\text{plim}_{N\to\infty} N^{-1} \Sigma_i\alpha_i\epsilon_i = 0$, which is the same result as is implied by our assumption that α_i and ϵ_i are uncorrelated.

Assumption BA.5 is clearly weaker than the assumption that $E(\epsilon_i|W_i,\alpha_i) = 0$. The implication of BA.5 is that the model (1) reflects a linear projection specification rather than a conditional expectation. More specifically, this assumption asserts that $\text{Proj}(y_{it}|W_i,\alpha_i) = X_{it}\beta + Z_i\gamma + \theta_t\alpha_i$, where $\text{Proj}(q|h)$ represents the population least squares projection of a variable q on a set of variables h. We will strengthen this assumption in some of our later analysis. Also, it is obvious that BA.5 implies that ϵ_{it} is uncorrelated with X_{is} for all t and s. In some applications, such as rational expectations models or sequential moment restriction models (Keane and Runkle, 1992; and Chamberlain, 1992), we may wish to assert only the weak exogeneity condition that ϵ_{it} is uncorrelated with X_{is} for $s \le t$; that is, $E(\epsilon_{it}X_{is}) = 0$ for $s \le t$, which is weaker

than the sequential conditional moment restriction, $E(\epsilon_{it}|X_{i1},...,X_{it}) = 0$, t = 1, ... , T. The weak exogeneity condition would lead to relatively straightforward modifications of our GMM estimators, since it would just reduce the set of available moment conditions. Chamberlain (1992) derives the efficiency bound for the model with sequential moment restrictions, and Hahn (1997) provides the efficient GMM estimator whose asymptotic covariance matrix coincides with the efficiency bound. For the model with the weak exogeneity assumption, the efficiency bound can be easily found. Since using more moment conditions never hurts the asymptotic efficiency of GMM, the GMM estimator utilizing all of the moment conditions implied by the weak exogeneity assumption should be efficient.

The coefficients (γ) of the time-invariant regressors (Z_i) are obviously not identified if ξ is constant; that is, if $\theta_2 = ... = \theta_T = 1$. Otherwise Assumption BA serves to identify the parameters β, γ and θ. Assumption BA.4 is made to ensure identification of θ. It is not hard to see that condition (BA.4) fails in the case that α_i is uncorrelated with every variable in W_i, and that in this case the overall intercept and θ_2, ... , θ_T are not identified. Intuitively, because the covariance matrix of ϵ_i is unrestricted, correlation between W_i and α_i is needed to distinguish the error component $\xi\alpha_i$ from ϵ_i. Further discussion of this identification condition is given in Appendix B.

Since the model (2) allows the coefficient (θ) of the effects α_i to vary over time, a natural generalization of the model would be the case in which the coefficients (γ) of the time-invariant regressors are allowed to be time dependent. However, unfortunately, the parameters in γ (say, γ_1, ... ,γ_T) cannot be separately identified. All we can identify are the reduced-from parameters, $(\gamma_2-\theta_2\gamma_1)$, ... , $(\gamma_T-\theta_T\gamma_1)$. Thus, unless there exist some prior restrictions on γ_1, it is not possible

to identify each of γ's. More detailed discussion of this problem is given in Appendix B.

In addition to our BASIC ASSUMPTIONS, we also consider the following **COVARIANCE ASSUMPTIONS (CA)** on the ϵ_{it}:

For any i, the ϵ_{it} are mutually uncorrelated. **(CA.1)**

For any i, var(ϵ_{it}) is the same for all t. **(CA.2)**

Thus, under CA, the covariance matrix of ϵ_i is of the form $\sigma_\epsilon^2 I_T$. Assumption CA.1 is implied by some versions of model (1), such as the weak exogeneity projection model that asserts that $\text{Proj}(y_{it}|W_i, \alpha_i, y_{i1}, ..., y_{i,t-1}) = X_{it}\beta + Z_i\gamma + \theta_t\alpha_i$. Assumption CA.2 is not typically implied by projection models (or conditional expectations models either), but it has often been assumed in the traditional fixed effects model. We note once more that CA.1-CA.2 are assumptions about unconditional second moments. Conditional expectations models such as the rational expectations model have implications for conditional second moments, and therefore may generate additional moment conditions.

3. GMM estimation under basic and covariance assumptions

We will first consider GMM estimation under our basic assumptions (BA) only. Under BA, it is apparent that the following moment conditions hold:

$$E(W_i' u_{it}) = \theta_t \Sigma_{w\alpha}, \quad t = 1, \ldots, T, \tag{4}$$

where $\theta_1 = 1$. We can eliminate the nuisance parameters $\Sigma_{w\alpha}$ from these moment conditions

without any loss of useful information on the parameters of interest. While there are many ways to do so, we choose the following (T-1)(Tk+g) moment conditions:

$$E[W_i^{/}(u_{it} - \theta_t u_{i1})] = 0, \quad t = 2, \ldots, T. \tag{5}$$

GMM based on the moment conditions (5) can be viewed as a natural extension of the within estimator for the usual fixed effects model. To understand this analogy, we should define the equivalents of the within and between transformations. Define $G = G(\theta) = (-\theta, I_{T-1})'$, where I_{T-1} is the (T-1)×(T-1) identity matrix. Since G and ξ are orthogonal, the matrix $[G, (1/\xi'\xi)\xi]'$ is a nonsingular nonstochastic matrix, and therefore, premultiplying both sides of equation (2) by this matrix does not cause any informational loss. This transformation gives us:

$$y_{it} = X_{it}\beta - X_{i1}[\theta_t\beta] + Z_i[(1-\theta_t)\gamma] + \theta_t y_{i1} + (u_{it} - \theta_t u_{i1}), \tag{6A}$$

$$(1/\xi'\xi)\xi' y_i = (1/\xi'\xi)\xi' X_i\beta + (1/\xi'\xi)(\xi' e_T) Z_i\gamma + (1/\xi'\xi)\xi' u_i, \tag{6B}$$

where equation (6A) holds for t = 2, ... , T. Equations (6A) are obtained simply by subtracting θ_t times the first equation from the last (T-1) equations of (1), a generalized within transformation. Chamberlain (1984, p. 1263) considers the same differencing method. Equation (6B) is simply the weighted sum over t of equation (1), a generalization of the between transformation. GMM based on the moment conditions (5) amounts to estimating the system (6A) by nonlinear three-stage least squares, using W_i as instruments for each of the equations. Equation (6B) is irrelevant for estimation under BA, because no legitimate instruments are available for it unless we make additional assumptions.

We now consider GMM estimation based on the moment conditions (5). Define the

vector of the parameters of interest by $\delta = (\beta', \gamma', \theta')'$. We will denote GMM based on the moment conditions (5) by GMM1, and the resulting estimator by $\hat{\delta}_1$, in order to distinguish it from other GMM estimators which we will consider later. The estimator $\hat{\delta}_1$ can be obtained by the usual GMM procedure. The (T-1)(Tk+g) moment conditions in (5) can be expressed in the following matrix form:

$$E[b_{1i}(\delta)] = E[W_i^{o\prime} G(\theta)' u_i(\beta,\gamma)] = 0, \tag{7}$$

where $b_{1i}(\delta)$ is the $\{(T-1)(Tk+g)\}\times 1$ moment vector, $W_i^o = (I_{T-1} \otimes W_i)$ and $u_i(\beta,\gamma) = y_i - X_i\beta - e_T Z_i \gamma$.[2] We define the sample average of the b_{1i} by:

$$b_{1N}(\delta) = \frac{1}{N}\sum_{i=1}^{N} b_{1i}(\delta). \tag{8}$$

Then, $\hat{\delta}_1$ solves the problem, $\min_\delta \; N b_{1N}(\delta)' V_{11}^{-1} b_{1N}(\delta)$, where $V_{11} = E[b_{1i} b_{1i}']$. Here V_{11} can be replaced by any consistent estimate, say $\hat{V}_{11}$. A natural candidate for $\hat{V}_{11}$ is $N^{-1}\Sigma_{i=1}^{N} b_{1i}(\hat{\delta}) b_{1i}(\hat{\delta})'$, where $\hat{\delta}$ is an initial consistent estimate of δ.

We now consider GMM estimation under the covariance assumptions (CA) as well as the basic assumptions (BA). Under CA.1 and CA.2, the T(T+1)/2 distinct elements of $E(u_i u_i')$ depend on θ and the two nuisance parameters σ_ϵ^2 and $\Sigma_{\alpha\alpha}$. Since θ is already identified under BA only, CA should imply T(T+1)/2 - 2 additional moment conditions which can be imposed in

---

[2]If the time-varying regressors X_{it} were only weakly exogenous (i.e., $E(\epsilon_{it} | X_{i1}, \ldots, X_{it}) = 0$), the moment conditions given in (7) can be modified as follows. For simplicity, assume the time-invariant regressors Z_i are uncorrelated with the ϵ_{it} for any t. Let $W_{it} = (X_{i1}, \ldots, X_{it}, Z_i)$. Define $W_i^* = \mathrm{diag}(W_{i1}, \ldots, W_{i,T-1})$. Then the moment conditions implied by the weak exogeneity assumption can be expressed as $E[W_i^{*\prime} G(\theta)' u_i(\beta,\gamma)] = 0$.

GMM estimation of δ. We can derive these moment conditions by essentially the same method used by Ahn and Schmidt (1995b) in the simple dynamic panel data model. Note that:

$$E(u_i u_i^{/}) = E\begin{bmatrix} u_{i1}^2 & u_{i1}u_{i2} & \cdots & u_{i1}u_{iT} \\ u_{i2}u_{i1} & u_{i2}^2 & \cdots & u_{i2}u_{iT} \\ \vdots & \vdots & & \vdots \\ u_{iT}u_{i1} & u_{iT}u_{i2} & \cdots & u_{iT}^2 \end{bmatrix} = \begin{bmatrix} \Sigma_{\alpha\alpha}+\sigma_\epsilon^2 & \theta_2\Sigma_{\alpha\alpha} & \cdots & \theta_T\Sigma_{\alpha\alpha} \\ \theta_2\Sigma_{\alpha\alpha} & \theta_2^2\Sigma_{\alpha\alpha}+\sigma_\epsilon^2 & \cdots & \theta_2\theta_T\Sigma_{\alpha\alpha} \\ \vdots & \vdots & & \vdots \\ \theta_T\Sigma_{\alpha\alpha} & \theta_T\theta_2\Sigma_{\alpha\alpha} & \cdots & \theta_T^2\Sigma_{\alpha\alpha}+\sigma_\epsilon^2 \end{bmatrix}. \qquad \textbf{(9)}$$

We can find two types of moment conditions from (9). First, noting that $\theta_1 = 1$, there are T(T-1)/2 - 1 distinct conditions that $E(u_{it}u_{is}) = \theta_s E(u_{it}u_{i1})$ for any $1 < s < t$. These conditions reflect restrictions on the off-diagonal elements of $E(u_i u_i')$. Second, there are T-1 conditions that $E[u_{it}(\Sigma_{s=1}^T \theta_s u_{is})] = \theta_t E[u_{i1}(\Sigma_{s=1}^T \theta_s u_{is})]$ for $t > 1$. They reflect restrictions on all of the elements on $E(u_i u_i')$. Adding the number of restrictions, we get T(T+1)/2 - 2.

One of the many equivalent ways to express these moment conditions is as follows:

$$\begin{aligned} E[u_{it}(u_{is} - \theta_s u_{i1})] &= 0, \quad \text{for } t > s > 1, \\ E[u_{it}(u_{iT} - \theta_T u_{i1})] &= 0, \quad \text{for } 1 < t < T; \end{aligned} \qquad \textbf{(10)}$$

$$E[(1/\Sigma_{s=1}^T \theta_s^2)(\Sigma_{s=1}^T \theta_s u_{is})(u_{it} - \theta_t u_{i1})] = E[(1/\xi^{/}\xi)(\xi^{/}u_i)(u_{it} - \theta_t u_{i1})] = 0, \quad \text{for } t > 1. \qquad \textbf{(11)}$$

The T(T-1)/2 - 1 moment conditions in (10) are valid as long as CA.1 holds, while the validity of the (T-1) moment conditions in (11) also requires CA.2. The conditions (11) imply that the errors in (6A) and (6B) are uncorrelated.

For a matrix representation of (10) and (11), define $U_{it}^o(\beta,\gamma) = [u_{it}(\beta,\gamma), u_{i,t+1}(\beta,\gamma), \ldots, u_{iT}(\beta,\gamma)]$ for $t < T$, $U_{iT}^o(\beta,\gamma) = [u_{i2}(\beta,\gamma), u_{i3}(\beta,\gamma), \ldots, u_{i,T-1}(\beta,\gamma)]$, where $u_{it}(\beta,\gamma)$ is the t'th entry of

$u_i(\beta,\gamma)$. We also define $U_i^o(\beta,\gamma) = \text{diag}[U_{i3}^o(\beta,\gamma), U_{i4}^o(\beta,\gamma), \ldots, U_{iT}^o(\beta,\gamma)]$. Then we can express the conditions (10)-(11) as

$$E[b_{2i}(\delta)] = E[U_i^o(\beta,\gamma)'G(\theta)'u_i(\beta,\gamma)] = 0; \qquad \textbf{(12)}$$

$$E[b_{3i}(\delta)] = E[(1/\xi'\xi)\xi'u_i(\beta,\gamma)G(\theta)'u_i(\beta,\gamma)] = 0, \qquad \textbf{(13)}$$

where $b_{2i}(\delta)$ and $b_{3i}(\delta)$ are the $\{T(T-1)/2 - 1\} \times 1$ and $(T-1)\times 1$ vectors, respectively.

From now on, the procedure GMM2 and resulting estimator $\hat{\delta}_2$ will denote GMM based on the moment conditions (7) and (12), and GMM3 and $\hat{\delta}_3$ will denote GMM based on the conditions (7), (12) and (13). If the basic assumptions (BA) and the covariance assumptions (CA) are satisfied, GMM3 is at least as efficient as GMM2 and GMM1, while GMM2 is at least as efficient as GMM1. This follows from the general principle that adding valid moment conditions cannot decrease asymptotic efficiency.

Since the moment conditions b_{1i}, b_{2i} and b_{3i} are nonlinear, the GMM estimates $\hat{\delta}_1$, $\hat{\delta}_2$ and $\hat{\delta}_3$ need to be calculated by an iterative procedures. Alternatively, following Newey (1985), we can use a linearized GMM estimator details can be found in an earlier version of this paper (available on request). We also note the following. Given the validity of the moment conditions (5) (or (7)) that follow from BA, the T(T-1)/2 - 1 moment conditions in (10) can be replaced by

$$\begin{aligned} E[y_{it}(u_{is} - \theta_s u_{i1})] &= 0, \quad \text{for } t > s > 1, \\ E[y_{it}(u_{iT} - \theta_T u_{i1})] &= 0, \quad \text{for } t = 2, 3, \ldots, T-1. \end{aligned} \qquad \textbf{(14)}$$

Given the conditions (7), the conditions (14) contain the same information as (12) (or (10)), but they are more nearly linear.

To obtain a matrix representation of (14), we define $Y_{it}^{o} = (y_{it}, y_{i,t+1}, \ldots, y_{iT})$ for t < T, and $Y_{iT}^{o} = (y_{i2}, y_{i3}, \ldots, y_{i,T-1})$. Let $Y_i^{o} = \mathrm{diag}(Y_{i3}^{o}, Y_{i4}^{o}, \ldots, Y_{iT}^{o})$. Then we can express (14) as

$$E[b_{2i}^{*}(\delta)] = E[Y_i^{o\prime} G(\theta)' u_i(\beta, \gamma)] = 0, \qquad \textbf{(15)}$$

where $b_{2i}^{*}(\delta)$ is the {T(T-1)/2 -1} × 1 vector. Then the GMM estimator based on b_{1i} and b_{2i}^{*} is asymptotically equivalent to GMM2, and the GMM estimator based on b_{1i}, b_{2i}^{*} and b_{3i} is asymptotically equivalent to GMM3.

Although it would perhaps be unusual, we could also consider GMM that assumes BA and CA.2 (homoskedasticity) but not CA.1 (nonautocorrelation). When CA.2 holds but CA.1 does not, the moment conditions b_{3i} based on (13) (or equivalently, (11)) do not hold. To obtain moment conditions implied by CA.2 without CA.1 holding, observe that the diagonal elements in (9) contain two nuisance parameters, $\Sigma_{\alpha\alpha}$ and σ_{ϵ}^2. Eliminating these parameters, we can obtain the following T -2 moment conditions:

$$E[(\theta_2^2 - \theta_t^2)u_{i1}^2 + (\theta_t^2 - 1)u_{i2}^2 + (1 - \theta_2^2)u_{it}] = 0, \quad \text{for } t > 2. \qquad \textbf{(16)}$$

Comparing (11) and (16), we can see that CA.2 itself implies (T-2) moment restrictions, while it generates an additional moment condition if CA.1 is also incorporated.

For a matrix representation of (16), define

$$G_*(\theta) = \begin{bmatrix} \theta_2^2-\theta_3^2 & \theta_2^2-\theta_4^2 & \cdots & \theta_2^2-\theta_T^2 \\ \theta_3^2-1 & \theta_4^2-1 & \cdots & \theta_T^2-1 \\ 1-\theta_2^2 & 0 & \cdots & 0 \\ 0 & 1-\theta_2^2 & \cdots & 0 \\ \vdots & \vdots & & 0 \\ 0 & 0 & \cdots & 1-\theta_2^2 \end{bmatrix}_{T\times(T-2)} , \qquad \textbf{(17)}$$

and $u_i^*(\beta,\gamma) = (u_{i1}(\beta,\gamma)^2, \ldots , u_{iT}(\beta,\gamma)^2)'$. Then we can express the (T-2) moment conditions implied by homoskedasticity only as:

$$E[b_{NH,i}(\delta)] = E[G_*(\theta)' u_i^*(\beta,\gamma)] = 0 . \qquad \textbf{(18)}$$

Then GMM imposing BA and CA.2 would be based on b_{1i} and $b_{NH,i}$. We will call this estimator GMM-NH or $\hat{\delta}_{NH}$, since it imposes BA plus no heteroskedasticity. This estimator would be (asymptotically) more efficient than GMM1 but less efficient than GMM3; it cannot be ranked relative to GMM2.

4. Some efficiency results

In this section we prove some efficiency results for the estimators $\hat{\delta}_1$ (GMM1), $\hat{\delta}_2$ (GMM2), $\hat{\delta}_{NH}$ (GMM-NH) and $\hat{\delta}_3$ (GMM3) of the previous section, under somewhat stronger assumptions than we made there. In particular, we show that under certain independence assumptions we can substantially reduce the number of moment conditions without a loss of efficiency.

We first consider the GMM1 estimator $\hat{\delta}_1$ based on the moment conditions (5) -- or equivalently, (7). The number [(T-1)(Tk+g)] of moment conditions in (5) could be very large. In this case, GMM using all of the moment conditions may be computationally burdensome, and may have poor finite sample properties. This raises the question of whether we can reduce the number of moment conditions without losing efficiency of estimation. This is possible if we make a stronger assumption than the uncorrelatedness condition BA.5. Specifically, suppose that we replace BA.5 with the following assumption (BI, for basic assumption with independence):

$[W_i, \alpha_i]$ is independent of ϵ_i. **(BI)**

Define $\Phi = E(\epsilon_i \epsilon_i')$ and $\Psi = G'\Phi G$, where G is as defined in the discussion above equation (6). Under BI, we can optimally weight the moment conditions (5) using Ψ^{-1}, decreasing the number of moment conditions that need to be imposed in GMM in order to attain the same efficiency as $\hat{\delta}_1$.

We first define the linear projection of α_i on W_i:

$$\alpha_i = Proj(\alpha_i | W_i) + \eta_i = W_i \pi + \eta_i, \qquad \textbf{(19)}$$

where $E(W_i' \eta_i) = 0$ and $\pi = (\Sigma_{ww})^{-1} \Sigma_{w\alpha}$. This leads to the model:

$$y_i = X_i \beta + e_T Z_i \gamma + \xi W_i \pi + e_i; \quad e_i = \xi \eta_i + \epsilon_i. \qquad \textbf{(20)}$$

This is essentially the same construction as was applied by Chamberlain (1984) to the usual fixed effects model. This model serves to identify the parameters β, γ, θ and π; for example,

following Chamberlain, consistent estimates of these parameters can be obtained by applying the minimum distance principle to the least squares estimates from (20).

Now consider the following moment conditions.

$$E[X_i' G(\theta)\Psi^{-1} G(\theta)' u_i(\beta,\gamma)] = 0, \tag{21A}$$

$$E[Z_i' e_T' G(\theta)\Psi^{-1} G(\theta)' u_i(\beta,\gamma)] = 0, \tag{21B}$$

$$E[(W_i\pi)J' G(\theta)\Psi^{-1} G(\theta)' u_i(\beta,\gamma)] = E[(W_i\pi)\Psi^{-1} G(\theta)' u_i(\beta,\gamma)] = 0, \tag{21C}$$

where $J = [0_{(T-1)\times 1}, I_{T-1}]'$ and $J'G(\theta) = I_{T-1}$. Notice that (21A), (21B) and (21C) imply the k, g, and T - 1 moment conditions, respectively. Accordingly, the total number of moment conditions in (21A)-(21C) is exactly the same as the number of parameters in δ. These conditions can be imposed as follows. Define $R_i(\theta,\pi) = [G(\theta)'X_i, G(\theta)'e_T Z_i, (W_i\pi)I_{T-1}]$. Let $\dot{\delta} = (\dot{\beta}', \dot{\gamma}', \dot{\theta}')'$ denote the GMM estimator based on the moment conditions (21A)-(21C). It solves:

$$\frac{1}{N}\sum_{i=1}^{N} R_i(\hat{\theta},\hat{\pi})'\hat{\Psi}^{-1} G(\dot{\theta})' u_i(\dot{\beta},\dot{\gamma}) = 0, \tag{22}$$

where $\hat{\theta}$, $\hat{\pi}$ and $\hat{\Psi}$ are initial consistent estimates.

To interpret the moment conditions (21A)-(21C), recall that $G(\theta)'u_i(\beta,\gamma)$ is a vector whose typical element is of the form $u_{it} - \theta_t u_{i1}$. These moment conditions say that some linear functions of W_i are uncorrelated with $u_{it} - \theta_t u_{i1}$. Therefore they are linear combinations of the moment conditions (5), which say that W_i is uncorrelated with $u_{it} - \theta_t u_{i1}$ for all t. As a result, $\hat{\delta}_1$ (GMM1) must be efficient relative to $\dot{\delta}$. However, when BA.5 is strengthened to BI, we have the following useful result.

Proposition 1: Under BA.1-BA.4 and BI, $\hat{\delta}_1$ and $\dot{\delta}$ are asymptotically equivalent.

Proof: See Appendix C.

It is worth noting that under BI both $\hat{\delta}_1$ and $\dot{\delta}$ are in general inefficient. We can construct a more efficient estimator than $\hat{\delta}_1$ as in Chamberlain (1992). Define $h_i = h(W_i) = E(\alpha_i|W_i)$. Then Proposition 3 of Chamberlain implies that $\hat{\delta}_1$ is dominated by a GMM estimator, say $\hat{\delta}_C$, which is obtained by exploiting the moment conditions (21A)-(21C), but with $(W_i\pi)$ replaced by h_i. The efficiency gain of $\hat{\delta}_C$ over $\hat{\delta}_1$ is simply due to the fact that the conditional mean h_i is a better approximation of α_i than its linear projection, $W_i\pi$. Of course, $\hat{\delta}_1$ is asymptotically equivalent to $\hat{\delta}_C$ if $h_i = W_i\pi$. A practical difficulty in applying Chamberlain's method is that the calculation of $\hat{\delta}_C$ requires nonparametric estimation of h_i. Also, under BI even $\hat{\delta}_C$ is generally inefficient, because it exploits the restrictions implied by $E(\epsilon_i|W_i,\alpha_i) = 0$, but not those implied by independence.

Independence between the explanatory variables and the ϵ_{it} involves higher-order moments which may be relevant for GMM. However, the following proposition states that under a normality assumption, $\hat{\delta}_1$ can be fully efficient.

Proposition 2: Assume that ϵ_i and η_i are stochastically independent of W_i and of each other. Assume that η_i and ϵ_i are normally distributed with variance σ_η^2 and covariance matrix Φ, respectively. Then $\hat{\delta}_1$ is asymptotically efficient.

The proof is given in Appendix C. Basically it shows that, under the assumptions made,

$\hat{\delta}_1$ is locally semiparametrically efficient. We can note that for the MLE, δ and the nuisance parameters σ_η^2, Φ and π are estimated jointly. A sidelight of Proposition 2 is that under the normality assumption GMM1 allows one to estimate δ efficiently without estimating the nuisance parameters.

We now turn to the estimators $\hat{\delta}_2$, $\hat{\delta}_{NH}$ and $\hat{\delta}_3$ that impose the basic assumptions (BA) and the covariance assumptions (CA). Again we want to know whether we can reduce the number of moment conditions without losing efficiency. To proceed, we need to strengthen the covariance assumptions. Specifically, we replace CA by the stronger assumption (CI) that the ϵ_{it} are iid:

For any i, the ϵ_{it} are mutually independent. **(CI.1)**

For any i, the ϵ_{it} are identically distributed. **(CI.2)**

While the estimators $\hat{\delta}_2$, $\hat{\delta}_{NH}$ and $\hat{\delta}_3$ are in general inefficient, they can be fully efficient under CI and a normality assumption.

<u>Proposition 3</u>: Assume that BA.1-BA.4 and BI hold. Assume that η_i is independent of W_i and ϵ_i. Assume that η_i and ϵ_i are normally distributed with variance σ_η^2 and covariance matrix Φ, respectively. If CI.1 holds, $\hat{\delta}_2$ has the same asymptotic distribution as the (restricted) MLE of δ in the model given in equation (20) above, imposing the restriction that Φ is diagonal. If CI.2 holds but CI.1 does not, the GMM estimator based on the moment conditions b_{1i} and $b_{NH,i}(\hat{\delta}_{NH})$ is asymptotically equivalent to the (restricted) MLE of δ, imposing the restriction that all the

diagonal entries of Φ are the same. If both CI.1 and CI.2 hold, then $\hat{\delta}_3$ has the same asymptotic distribution as the (restricted) MLE of δ, imposing the restriction that $\Phi = \sigma_\epsilon^2 I_T$.

Proof: See Appendix C.

Basically, this proposition states conditions under which the GMM estimators $\hat{\delta}_2$, $\hat{\delta}_3$ and $\hat{\delta}_{NH}$ are efficient in the class of estimators that make use of second-moment information. Like Proposition 2, it can be viewed as a local efficiency result. The basic assumptions (BA) and covariance assumptions (CA) are stated in terms of uncorrelatedness only, so that only second-moment information is relevant. When we replace uncorrelatedness by independence, additional moment conditions involving higher-order moments are potentially relevant. However, these additional moment conditions do not lead to gains in asymptotic efficiency when η_i and the ϵ_{it} are normal. This is because under normality only the first two moments matter.

We can also show that under the assumptions of Proposition 3, many moment conditions in (7), (12) and (13) become redundant in the sense that imposing them in GMM does not improve the efficiency of the estimator. Stated formally:

<u>Proposition 4</u>: Assume BA.1-BA.4, BI, CI.1, CI.2, and normality of the ϵ_{it}. Then $\hat{\delta}_3$ has the same asymptotic distribution as the GMM estimator based on the moment conditions (21A)-(21C) and (13).

Proof: See Appendix C.

Proposition 4 gives conditions under which we can considerably reduce the number of

moment conditions in GMM3 without losing asymptotic efficiency. It was already clear from Proposition 3 that such a reduction in relevant moment conditions is possible under some assumptions, but the assumptions of Proposition 4 are weaker than those of Proposition 3, because we do not need to assume normality of η_i.

GMM3 is based on the set of moment conditions (7), (12) and (13). Proposition 4 essentially says two things. First, it says that we can replace the largest set of moment conditions (7) by the much smaller set of moment conditions (21A)-(21C). This is the same conclusion as in Proposition 1. However, it does not follow as an implication of Proposition 1; it is a new result. As shown in Breusch et al. (1999), the fact that one set of moment conditions is redundant, given another set of moment conditions, can change when the other set of moment conditions is made larger. Second, the proposition says that we can omit the set of moment conditions (12), because these moment conditions are redundant given those in (7) and (13), or given those in (21A)-(21B) and (13). This redundancy result is the reason that we expressed the moment conditions b_{2i} and b_{3i} in the way we did. The redundant conditions b_{2i} given in (12) are those that reflect uncorrelatedness of the errors. This is a very useful result because it allows us to eliminate a large number of nonlinear moment conditions without losing efficiency. We note that, although the moment conditions (12) are redundant given those in (7) and (13), they are not uninformative in general. For example, the GMM2 estimator based on (7) and (12) is more efficient than the GMM1 estimator based on (7) only, even given the assumptions of Proposition 4, so that there is useful information in (12) that is not in (7). However, given the assumptions of Proposition 4, there is no useful information in (12) that is not in (7) plus (13).

5. Concentrated least squares

In this section, we shall consider the concentrated least squares (CLS) estimation of the model. The CLS estimator is defined as follows. We treat the α_i as parameters to be estimated (along with β, γ and θ), so that this is a true fixed-effects treatment of the model. We can consider the following least squares problem:

$$\min_{\delta,\alpha_i} \quad SSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - X_i\beta - e_T Z_i\gamma - \xi\alpha_i)'(y_i - X_i\beta - e_T Z_i\gamma - \xi\alpha_i). \tag{23}$$

Solving the first order conditions of the problem (23), we can express the α_i in terms of δ:

$$\alpha_i = \frac{1}{(\xi'\xi)}\xi' u_i(\beta,\gamma), \quad i = 1, 2, \ldots, N, \tag{24}$$

where as before $u_i(\beta,\gamma) = y_i - X_i\beta - e_T Z_i\gamma$. Substituting (24) into (23) gives us the concentrated SSE (CSSE):

$$CSSE = \sum_{i=1}^{N} u_i(\beta,\gamma)' M(\xi) u_i(\beta,\gamma), \tag{25}$$

where $M(\xi) = I_T - \xi(\xi'\xi)^{-1}\xi'$. Minimizing CSSE with respect to δ gives us the CLS estimator, $\hat{\delta}_{CLS}$. If the ϵ_{it} are iid normal and if X_i and Z_i are treated as fixed, this estimator is the MLE. Lee (1991) shows that it can also be derived as a conditional MLE estimator, where the conditional likelihood function is obtained by conditioning on $\xi'(y_i - X_i\beta - e_T Z_i\gamma)$, the sufficient statistic for α_i. Because the α_i are "incidental parameters" in the sense of Neyman and Scott (1948) -- that is,

the number of α_i grows with sample size N -- and because the sufficient statistic for α_i depends on the parameters, the usual results for the asymptotic properties of the MLE or conditional MLE do not apply, and the asymptotic properties of the CLS estimator need to be derived directly.[3] Further discussion of similar cases can be found in Cornwell and Schmidt (1987), Kiefer (1980b) and Chamberlain (1992).

The CLS estimator was first considered by Kiefer (1980a) and was considered later but independently by Lee (1991) and Lee and Schmidt (1993). They provide some interesting computational results. For example, they show that the CLS estimator can be computed by a specific iterative process. For a given value of ξ, the estimator of β and γ is obtained by regression of $M(\xi)y_i$ on $M(\xi)(X_i, e_T Z_i)$; conversely, for a given value of β and γ, the estimator of ξ is the eigenvector corresponding to the largest eigenvalue of $\Sigma_i u_i(\beta,\gamma)u_i(\beta,\gamma)'$.

Under the strong assumptions for which the CLS estimator is the MLE (iid normal ϵ's and fixed regressors), Lee (1991) rigorously considers the asymptotic properties of the CLS estimator. An interesting point that we will develop below is that the CLS estimator is consistent under weaker assumptions than Lee made, but consistency basically requires our covariance assumptions as well as the basic assumptions. An important finding of Lee's is that the efficiency of the MLE is in doubt. He shows that $E(\partial^2 CSSE/\partial\delta\partial\delta')$ fails to be proportional to $E[(\partial CSSE/\partial\delta)(\partial CSSE/\partial\delta')]$ even under his strong assumptions. Due to this fact, the asymptotic covariance matrix of $\sqrt{N}(\hat{\delta}_{CLS} - \delta)$ takes the form

[3]It might be worth mentioning that under the weak exogeneity assumption, $E(\epsilon_{is}X_{it}) = 0$ for $s \geq t$, the CLS estimator would become inconsistent, and modifying it to be consistent under weak exogeneity would fundamentally destroy its least squares or maximum likelihood interpretation.

$$\left[E\left(\frac{\partial^2 CSSE}{\partial\delta\partial\delta'}\right)\right]^{-1} E\left[\frac{\partial CSSE}{\delta}\frac{\partial CSSE}{\delta'}\right]\left[E\left(\frac{\partial^2 CSSE}{\partial\delta\partial\delta'}\right)\right]^{-1} \qquad \textbf{(26)}$$

which is similar in form to the covariance matrix of a quasi-ML estimator under distributional misspecification. This result casts doubt on the efficiency of CLS, and we will show that this doubt is justified, in the sense that the CLS estimator is dominated by a GMM estimator. This may be a surprising result because the incidental parameters problem is typically thought to cast doubt on the consistency of the MLE, not on its efficiency.

A simple example may be worth brief consideration before moving on to the general case. Let T = 2 and suppose that there are no explanatory variables in the model, so we simply have

$$y_{i1} = \alpha_i + \epsilon_{i1}, \quad y_{i2} = \alpha_i\theta + \epsilon_{i2}. \qquad \textbf{(27)}$$

We wish to estimate the scalar parameter θ and the nuisance parameters $\alpha_1, ..., \alpha_N$. The CLS estimator minimizes the criterion function CSSE = $\Sigma_i\ [(y_{i1}-\alpha_i)^2 + (y_{i2}-\alpha_i\theta)^2]$. Some algebra reveals that the CLS estimator of θ solves the equation

$$\sum_{i=1}^{N}(y_{i1} + \hat{\theta}y_{i2})(y_{i2} - \hat{\theta}y_{i1}) = 0. \qquad \textbf{(28)}$$

Obviously this corresponds to imposition of the moment condition

$$E[(y_{i1} + \theta y_{i2})(y_{i2} - \theta y_{i1})] = 0. \qquad \textbf{(29)}$$

Evaluating the expectation on the left hand side of (29), we obtain $\theta\text{var}(\epsilon_{i2}) - \theta\text{var}(\epsilon_{i1}) +$

$(1-\theta^2)\text{cov}(\epsilon_{i1},\epsilon_{i2})$, and this equals zero for all values of θ if and only if the covariance assumptions hold. Thus the consistency of the estimator follows from and requires these covariance assumptions. Furthermore, the inefficiency of the CLS estimator follows from its failure to exploit the moment condition $E(y_{i2} - \theta y_{i1}) = 0$, which is informative for θ so long as the mean of the α_i is non-zero.

We now return to the general model (1), and we consider the properties of the CLS estimator from the perspective of GMM. For expository convenience, we will do so under assumptions BA.1-BA.4, BI and CI. These assumptions are weaker than the assumptions made by Lee, but stronger than actually necessary, since BA and CA are sufficient for the consistency and asymptotic normality of the CLS estimator.

We previously defined the quasi-differencing matrix G, which has the property that, for any T dimensional vector d, the t'th element of $G'd$ equals $d_{t+1} - \theta_{t+1}d_1$. Using the fact that $M(\xi) = G(G'G)^{-1}G'$, the first-order conditions for the minimization of CSSE can be written as follows:

$$\frac{\partial CSSE}{\partial \delta} = \begin{bmatrix} \frac{\partial CSSE}{\partial \beta} \\ \frac{\partial CSSE}{\partial \gamma} \\ \frac{\partial CSSE}{\partial \theta} \end{bmatrix} = -2\sum_{i=1}^{N} \begin{bmatrix} X_i' G(G'G)^{-1}G'u_i(\beta,\gamma) \\ Z_i' e_T' G(G'G)^{-1}G'u_i(\beta,\gamma) \\ \frac{1}{\xi'\xi}[\xi' u_i(\beta,\theta)](G'G)^{-1}G'u_i(\beta,\gamma) \end{bmatrix} = 0. \qquad \textbf{(30)}$$

The consistency of the CLS estimator requires that the terms in the large brackets in equation (30) have expectation zero (k + g + (T-1) moment conditions). The first two terms have expectation zero under our basic assumptions; what is required is essentially that ϵ_i is uncorrelated with X_i and Z_i. However, for the last term to have expectation zero requires that

$\xi' u_i(\beta,\gamma)$ be uncorrelated with $G' u_i(\beta,\gamma)$, which requires the covariance assumptions. Furthermore, it is clear from (30) that the CLS estimator is equivalent to a GMM estimator based on the moment conditions (21A), (21B) and (13), with $\Psi = G'G$ in (21A) and (21B). Thus the CLS estimator exploits only a subset of the moment conditions that GMM3 exploits, which explains the inefficiency of CLS compared to GMM3. To make this point clear, consider the case in which the ϵ_{it} are iid normal. Then Proposition 4 implies that $\hat{\delta}_3$ is asymptotically equivalent to the GMM estimator based on the conditions (21A)-(21C) and (13), and the inefficiency of CLS results from the fact that it fails to exploit the moment conditions (21C).

6. Calculation of asymptotic covariance matrices

In this section we attempt to quantify the efficiency gains of using the moment conditions implied by the covariance assumptions. We consider five estimators, GMM1 ($\hat{\delta}_1$), GMM2 ($\hat{\delta}_2$), GMM-NH ($\hat{\delta}_{NH}$), GMM3 ($\hat{\delta}_3$) and CLS ($\hat{\delta}_{CLS}$). We will calculate and compare the asymptotic covariance matrices of these estimates, for various parameter values. A derivation of the asymptotic covariance matrices is given in Appendix C. The calculations will be made assuming that BA.1-BA.4, BI and CI hold. Under these assumptions, GMM3 is most efficient, while both GMM2 and GMM2-NH are more efficient than GMM1. However, *a priori* efficiency comparison of GMM2 with GMM-NH is not possible, because these estimates are based upon non-nested sets of moment conditions. By the same reasoning, the efficiency comparison of CLS with GMM1, GMM2 and GMM-NH is not possible.

The basis of our calculations is the model (1) with a single time-varying regressor and an intercept term:

$$y_{it} = \beta X_{it} + \gamma + \theta_t \alpha_i + \epsilon_{it} , \qquad \textbf{(31)}$$

where $\theta_1 = 1$ and $\theta_t = \theta_{t-1} + \rho$ for t = 2, ... , T. We assume that:

$$E\begin{bmatrix} X_{i1} \\ \vdots \\ X_{iT} \\ \alpha_i \end{bmatrix} = \begin{bmatrix} \mu_X e_T \\ \mu_\alpha \end{bmatrix} ; \; cov\begin{bmatrix} X_{i1} \\ \vdots \\ X_{iT} \\ \alpha_i \end{bmatrix} = \begin{bmatrix} \sigma_X^2 I_T & \sigma_{X\alpha} e_T \\ \sigma_{X\alpha} e_T^{/} & \sigma_\alpha^2 \end{bmatrix} ; \; cov\begin{bmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{i,T-1} \\ \epsilon_{iT} \end{bmatrix} = \sigma_\epsilon^2 I_T \; . \qquad \textbf{(32)}$$

With these assumptions, we have $\mu_w = (\mu_X e_T', 1)'$ and:

$$\Sigma_{ww} = \begin{bmatrix} \sigma_X^2 I_T + \mu_X^2 e_T e_T^{/} & \mu_X e_T \\ \mu_X e_T^{/} & 1 \end{bmatrix} ; \; \Sigma_{w\alpha} = \begin{bmatrix} (\sigma_{X\alpha} + \mu_X \mu_\alpha) e_T \\ \mu_\alpha \end{bmatrix} ; \; \Sigma_{\alpha\alpha} = \sigma_\alpha^2 + \mu_\alpha^2 \; . \qquad \textbf{(33)}$$

The relevant parameters are σ_X^2, ρ, σ_α^2, $\sigma_{X\alpha}$, μ_α, μ_X and T. We normalize the covariance matrices by setting $\sigma_\epsilon^2 = 1$. Our choices of parameter values ensure positive definiteness of the covariance matrices and identification of δ.

Table 1 reports some results for the case that the ϵ_{it} are normal. The numbers in the parentheses are the variances of the GMM1 estimates. The other results are the ratios of the asymptotic variances of the GMM2, GMM-NH, GMM3 and CLS estimates to the asymptotic variances of the GMM1 estimates [e.g, var($\hat{\beta}_2$)/var($\hat{\beta}_1$)]. A salient fact in Table 1 (and also the other tables) is that exploiting the covariance restrictions does not provide much efficiency gain for the estimates of β, while it substantially improves the efficiency of the estimates of γ and the θ_t. Apparently the moment conditions implied by the orthogonality between the errors and the

regressors are much more important in the estimation of the coefficients of time-varying regressors than those which are implied by the covariance restrictions.

In Table 1, with a few exceptions, the efficiency of GMM2, GMM-NH, GMM3 and CLS over GMM1 increases, as the time-variation of the individual effect (ρ), the mean of the individual effect (μ_α) and the variance of the individual effect (σ_α^2) increase, and as the correlation between the regressor and the individual effect ($\sigma_{X\alpha}$) decreases. In the cases with small values of $\sigma_{X\alpha}$, the efficiency gains from using GMM2, GMM-NH, GMM3 and CLS instead of GMM1 are very large. This is reasonable because GMM1 does not provide a consistent estimate of γ or of the θ_t when $\sigma_{X\alpha} = 0$. It is interesting to see that the CLS estimates in general are slightly better than both the GMM2 and GMM-NH estimates. The efficiency of the CLS estimates is in most cases almost equal to that of the GMM3 estimates. Thus, while GMM3 does dominate CLS, it is not very much better given the assumptions underlying our calculations. There are two exceptional cases in which CLS performs worse than other GMM estimators: the cases with $\sigma_\alpha^2 = 0.38$ and with $\sigma_{X\alpha} = 1.15$. These are parameter values near the boundary of the parameter subspace for which all covariance matrices are positive definite.

The efficiency gains of GMM-NH over GMM1 are in general greater than the efficiency gains of GMM2 over GMM1, but only by very small margins. This result indicates that the efficiency gains by exploiting the moment conditions that require homoskedasticity of the ϵ_{it} are only marginally greater than the efficiency gains from the moment conditions that require mutual zero-correlation among the ϵ_{it}. Further, the efficiency gains of GMM3 over both GMM2 and GMM-NH are, if any, only marginal. This means that the efficiency gains from using both sets of moment conditions (implied by both the zero-correlation and homoskedasticity assumptions)

are not too much greater than the gains from using only one set of moment conditions.

Table 2 reports results for different values of T. As T increases, the efficiency gains of GMM2, GMM-NH, GMM3 and CLS over GMM1 tend to become smaller and more nearly equal. This implies that marginal information in the covariance assumption is not very great for large values of T.

Table 3 and 4 report the results for some cases in which the ϵ_{it} are distributed as Chi-squared with one degree of freedom, but normalized so that $E(\epsilon_{it}) = 0$ and $var(\epsilon_{it}) = 1$. Note that the results reported for GMM1 and GMM2 are identical to those in Table 1 and 2. This is so because under assumptions BI and CI, the covariance matrices of GMM1 and GMM2 estimates are independent of the third and fourth moments of the ϵ_{it}. (See the forms of the covariance matrices given in Appendix C.) Most of the efficiency comparisons reported in Tables 3 and 4 are similar to those in Table 1 and 2.

Several general comments can be made on the results in Tables 1-4. First, covariance restrictions are particularly important when T is small. In this case, GMM exploiting the moment conditions implied by the covariance restrictions is worthwhile. Second, the efficiency gains of GMM2 over GMM1 or GMM-NH over GMM1 are often substantial, while the additional efficiency gains of GMM3 over GMM2 or GMM-NH are usually small. This implies that most of efficiency gains of exploiting covariance restrictions can be obtained by either imposing the moment conditions that require only the mutual zero-correlation among the ϵ_{it}, or by the moment conditions that only require homoskedasticity. Third, CLS performs well in general, but it can be worse than GMM1 in some cases. Furthermore, its consistency requires both nonautocorrelation and homoskedasticity of the ϵ_{it}. There seems to be little reason to prefer CLS to GMM2 or

GMM-NH, at least in terms of the asymptotic properties of the estimates.

7. Conclusions

In this paper we have considered a panel data model with time-varying individual effects, which is a generalization of the traditional fixed effects model. This model is of theoretical interest, and we believe that it is also empirically useful in a variety of settings. We have counted and expressed the first and second order moment conditions implied by alternative sets of assumptions on the model, and we have considered several GMM estimators and examined the conditions under which these estimators are efficient. While some of the available moment conditions are nonlinear, we can derive very simple linearized GMM estimators that are asymptotically as efficient as the nonlinear GMM estimators. These alternative GMM procedures are available from the authors upon request.

The main novelty in our GMM approach is the exploitation of possible covariance restrictions. Our calculations of asymptotic covariance matrices of the GMM estimators demonstrate that exploiting covariance restrictions can result in substantial efficiency gains, especially when the number of time-series observations on each individual is small.

We have shown that a concentrated least squares estimator, which is a natural generalization of the within estimator and which is the MLE when the errors are iid normal and the regressors are fixed, is consistent only when the errors are white noise. Furthermore, given the conditions needed for its consistency, this estimator is inefficient. This is an unusual and intriguing theoretical result, and it is reasonable to ask to what class of models it generalizes. This is an interesting topic for further research.

Another useful generalization of the model is to allow for multiple-component individual effects. That is, instead of a single component of the form $\theta_t\alpha_i$, we could consider multiple components of the form $\Sigma_{j=1}^{J} \theta_{jt}\alpha_{ji}$. For example, in some models one might wish to include a time-invariant individual effect (i.e., $\theta_{1t} = 1$ for all t) as well as a time-varying individual effect. As another example, in models in which the θ_t represent macro shocks, one might have two or more distinct macro shocks to which individuals react differently. Lee (1991) has considered conditional least squares estimation of multiple components models, and a GMM treatment should also be feasible.