EFFICIENT ESTIMATION OF MODELS FOR DYNAMIC PANEL DATA

Seung C. Ahn

Arizona State University

Peter Schmidt

Michigan State University

November, 1989

Revised July, 1990

Revised September, 1990

Revised February, 1992

Revised March, 1993

The second author gratefully acknowledges the financial support of the National Science Foundation. We also thank Trevor Breusch for his help with this paper, and several referees for their helpful comments.

REFERENCES

Abowd, J. and D. Card (1989), "On the Covariance Structure of Earnings and Hours Changes," Econometrica, 57, 411-446.

Ahn, S.C. (1990), "Three Essays on Share Contracts, Labor Supply, and the Estimation of Models for Dynamic Panel Data," unpublished Ph.D. dissertation, Michigan State University.

Ahn, S.C. and P. Schmidt (1990), "Efficient Estimation of Models for Dynamic Panel Data," unpublished manuscript, Michigan State University.

Amemiya, T. and T.E. MaCurdy (1986), "Instrumental Variable Estimation of an Error Components Model," Econometrica, 54, 869-881.

Anderson, T.W. and C. Hsiao (1981), "Estimation of Dynamic Models with Error Components," Journal of the American Statistical Association, 76, 598-606.

Arellano, M. and S. Bond (1991), "Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," Review of Economic Studies, 58, 277-97.

Arellano, M. and O. Bover (1990), "Another Look at the Instrumental Variables Estimation of Error-Component Models," unpublished manuscript, London School of Economics.

Blundell, R. and Smith (1991), R.S., "Conditions Initiales et Estimation Efficace dans les Modèles Dynamiques sur Donne'es de Panel," Anales D'Economie et de Statistique, 20-21, 109-123.

Bhargava, A. and J.D. Sargan (1983), "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods, Econometrica, 51, 1635-1659.

Breusch, T.S., G.E. Mizon and P. Schmidt (1989), "Efficient Estimation Using Panel Data," Econometrica, 57, 695-701.

Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data," Review of Economic Studies, 47, 225-238.

Chamberlain, G. (1982), "Multivariate Regression Models for Panel Data," Journal of Econometrics, 18, 5-46.

Chamberlain, G. (1984), "Panel Data," Chapter 22 in Handbook of Econometrics, Volume II, Z. Griliches and M. Intriligator, eds., North-Holland.

Hausman, J.A. and W.E. Taylor (1981), "Panel Data and Unobservable Individual Effects," Econometrica, 49, 1377-

1398.

Holtz-Eakin, D. (1988), "Testing for Individual Effects in Autoregressive Models," Journal of Econometrics, 39, 297-308.

Holtz-Eakin, D., W. Newey and H.S. Rosen (1988), "Estimating Vector Autoregressions with Panel Data," Econometrica, 56, 1371-1396.

Hsiao, C. (1986), Analysis of Panel Data, New York: Cambridge University Press.

Mundlak, Y. (1978), "On the pooling of Time-Series and Cross-Section Data," Econometrica, 46, 69-86.

Newey, W. (1985), "Generalized Method of Moments Specification Testing," Journal of Econometrics, 29, 229-256.

Newey, W. (1990), "Semiparametric Efficiency Bounds," Journal of Applied Econometrics, 5, 99-136.

Schmidt, P., S.C. Ahn and D. Wyhowski (1992), "Comment," Journal of Business and Economic Statistics, 10, 10-14.

TABLE 1

RATIOS OF ASYMPTOTIC VARIANCES

STATIONARITY ASSUMED (TO DETERMINE σ_{00} AND $\sigma_{0\alpha}$)

		Var(IV)/Var(GMM1)						Var(IV)/Var(GMM2)					
		$\sigma_{\alpha\alpha}/\sigma_{\varepsilon\varepsilon}$						$\sigma_{\alpha\alpha}/\sigma_{\varepsilon\varepsilon}$					
	δ	0	.25	.5	1	2	4	0	.25	.5	1	2	4
T=3	-.99	1.00	1.00	1.01	1.01	1.01	1.01	1.00	1.01	1.01	1.01	1.01	1.01
	-.9	1.04	1.05	1.05	1.06	1.07	1.10	1.05	1.05	1.06	1.07	1.09	1.12
	-.8	1.08	1.09	1.10	1.12	1.15	1.21	1.08	1.09	1.11	1.14	1.18	1.25
	-.5	1.14	1.19	1.23	1.31	1.43	1.61	1.14	1.20	1.26	1.35	1.51	1.72
	-.3	1.15	1.24	1.31	1.44	1.66	1.94	1.15	1.25	1.35	1.51	1.77	2.11
	0	1.14	1.29	1.42	1.67	2.04	2.52	1.14	1.32	1.49	1.80	2.27	2.86
	.3	1.11	1.33	1.54	1.91	2.49	3.20	1.11	1.39	1.67	2.18	2.95	3.91
	.5	1.08	1.35	1.62	2.10	2.82	3.71	1.08	1.45	1.84	2.53	3.58	4.87
	.8	1.03	1.41	1.78	2.42	3.37	4.55	1.03	1.64	2.26	3.34	4.96	6.97
	.9	1.02	1.45	1.85	2.54	3.57	4.85	1.02	1.76	2.49	3.75	5.62	7.96
	.99	1.00	1.48	1.92	2.65	3.76	5.13	1.00	1.94	2.78	4.21	6.35	9.03
T=4	-.99	1.00	1.00	1.00	1.00	1.01	1.01	1.00	1.00	1.00	1.01	1.01	1.01
	-.9	1.03	1.03	1.04	1.05	1.06	1.09	1.03	1.04	1.04	1.05	1.07	1.10
	-.8	1.06	1.07	1.08	1.10	1.14	1.19	1.06	1.07	1.09	1.11	1.15	1.21
	-.5	1.12	1.17	1.22	1.29	1.39	1.52	1.12	1.18	1.23	1.30	1.41	1.55
	-.3	1.15	1.24	1.32	1.44	1.60	1.79	1.15	1.25	1.34	1.46	1.64	1.84
	0	1.17	1.35	1.50	1.73	2.02	2.33	1.17	1.37	1.53	1.78	2.10	2.43
	.3	1.16	1.46	1.72	2.11	2.61	3.11	1.16	1.50	1.79	2.23	2.79	3.35
	.5	1.13	1.55	1.90	2.45	3.13	3.82	1.13	1.61	2.03	2.66	3.46	4.25
	.8	1.06	1.72	2.28	3.13	4.19	5.25	1.06	1.89	2.62	3.74	5.12	6.50
	.9	1.03	1.80	2.45	3.43	4.65	5.86	1.03	2.07	2.96	4.31	6.00	7.67
	.99	1.00	1.91	2.64	3.74	5.11	6.48	1.00	2.33	3.41	5.02	7.03	9.04
T=10	-.99	1.00	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.00	1.01
	-.9	1.02	1.03	1.03	1.03	1.04	1.05	1.02	1.03	1.03	1.03	1.04	1.05
	-.8	1.04	1.05	1.06	1.07	1.08	1.10	1.04	1.05	1.06	1.07	1.08	1.10
	-.5	1.08	1.12	1.15	1.18	1.21	1.24	1.08	1.12	1.15	1.18	1.21	1.25
	-.3	1.10	1.18	1.22	1.27	1.32	1.37	1.10	1.18	1.22	1.27	1.32	1.37
	0	1.15	1.30	1.38	1.46	1.55	1.61	1.15	1.30	1.38	1.47	1.55	1.62
	.3	1.19	1.49	1.64	1.80	1.94	2.04	1.19	1.49	1.65	1.81	1.95	2.05
	.5	1.21	1.70	1.95	2.21	2.41	2.56	1.21	1.71	1.96	2.22	2.43	2.57
	.8	1.17	2.37	3.00	3.61	4.11	4.44	1.17	2.42	3.06	3.71	4.22	4.56
	.9	1.11	2.86	3.76	4.67	5.38	5.86	1.11	2.96	3.92	4.89	5.65	6.16
	.99	1.01	3.65	4.98	6.30	7.36	8.07	1.01	3.94	5.42	6.89	8.07	8.86

TABLE 2

RATIOS OF ASYMPTOTIC VARIANCES

STATIONARITY NOT ASSUMED

	Var(IV)/Var(GMM1)						Var(IV)/Var(GMM2)					
$\sigma_{\alpha\alpha}$:	2	2	2	2	2	2	2	2	2	2	2	2
$\sigma_{0\alpha}$:	.5	.5	.5	0	.5	1	.5	.5	.5	0	.5	1
σ_{00}:	1	4	7	4	4	4	1	4	7	4	4	4
T:	3	3	3	3	3	3	3	3	3	3	3	3
δ = 0	4.52	1.61	1.33	1.49	1.61	1.73	5.20	1.72	1.39	1.58	1.72	1.87
.3	32.7	2.23	1.60	1.86	2.23	2.79	40.5	2.51	1.73	2.05	2.51	3.23
.5	57.3	3.34	2.07	2.39	3.34	5.54	73.6	3.97	2.36	2.76	3.97	6.85
.8	4.86	6.18	4.61	4.18	6.18	6.67	6.28	7.94	5.83	5.24	7.94	8.75
.9	3.37	5.00	4.94	4.49	5.00	3.91	4.31	6.45	6.36	5.73	6.45	5.06
1	2.58	3.51	3.70	4.00	3.51	2.58	3.26	4.49	4.75	5.14	4.49	3.26

	Var(IV)/Var(GMM1)						Var(IV)/Var(GMM2)					
$\sigma_{\alpha\alpha}$:	1	2	5	2	2	2	1	2	5	2	2	2
$\sigma_{0\alpha}$:	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
σ_{00}:	4	4	4	4	4	4	4	4	4	4	4	4
T:	3	3	3	3	4	10	3	3	3	3	4	10
δ = 0	1.31	1.61	2.56	1.61	1.75	1.62	1.38	1.72	2.72	1.72	1.80	1.62
.3	1.55	2.23	4.20	2.23	2.60	2.45	1.73	2.51	4.57	2.51	2.73	2.46
.5	2.06	3.34	5.71	3.34	3.91	3.98	2.48	3.97	6.33	3.97	4.19	4.00
.8	6.98	6.18	5.52	6.18	4.76	3.31	10.4	7.94	6.21	7.94	5.29	3.35
.9	5.64	5.00	4.66	5.00	3.63	2.17	8.56	6.45	5.25	6.45	4.04	2.20
1	2.93	3.51	3.83	3.51	2.67	1.59	4.29	4.49	4.30	4.49	2.96	1.61

1. INTRODUCTION

This paper considers a regression model for dynamic panel data. That is, we assume that we have observations on a large number of individuals, with several observations on each individual, and the model of interest is a regression model in which the lagged value of the dependent variable is one of the explanatory variables. The error in the model is assumed to contain a time-invariant individual effect as well as random noise.

We consider the commonly-assumed and empirically relevant case of a large number of individuals (N) and a small number of time series observations per individual (T), and so we study the asymptotic properties of our estimators as $N \to \infty$ for fixed T. Then the basic problem faced in the estimation of this model is that a fixed effects treatment leads to the within estimator (least squares after transformation to deviations from means), which is inconsistent because the within transformation induces a correlation of order 1/T between the lagged dependent variable and the error. The currently available response to this problem [e.g., Anderson and Hsiao (1981), Hsiao (1986), Holtz-Eakin (1988), Holtz-Eakin, Newey and Rosen (1988), Arellano and Bover (1990), Arellano and Bond (1991)] is to first difference the equation to remove the effects, and then estimate by instrumental variables (IV), using as instruments values of the dependent variable lagged two or more periods. This treatment leads to consistent estimates, but (under standard assumptions that we will list) not to efficient estimates, because it does not make use of all of the available moment conditions.

In this paper we find the additional moment conditions implied by a standard set of assumptions. Unlike the conditions currently exploited in the literature, some of these conditions are nonlinear. However, they can be easily imposed in a generalized method of moments (GMM) framework. We show that the resulting GMM estimator shares the efficiency properties of Chamberlain’s (1982, 1984) minimum distance (MD) estimator if the errors in the model are i.i.d (independently and identically distributed) across individuals. The GMM estimator is also compatible with a specification of the model in which the unobserved individual effects are independent across individuals but heteroskedastic. For some specific parameter values, we evaluate the relevant asymptotic covariance matrices to show that the additional moment conditions identified in this paper lead to non-trivial gains in asymptotic efficiency.

We also link the literature on the static panel data model [e.g., Mundlak (1978), Chamberlain (1980), Hausman and Taylor (1981), Amemiya and MaCurdy (1986), Breusch, Mizon and Schmidt (1989)] with the literature on the dynamic model by showing how to make efficient use of exogenous variables as instruments. There has been some confusion in the literature on this point, largely due to the effects of first-differencing on the errors. The assumption of strong exogeneity implies more orthogonality conditions than current estimators exploit. These additional moment conditions all lie in the "deviations from means" space and are irrelevant in the case of the static model, but they are relevant in the dynamic model. We show how to categorize and use these conditions.

The plan of the paper is as follows. Section 2 considers the simple dynamic model with no exogenous variables under a standard set of assumptions, and identifies the moment conditions available for estimation. Section 3 considers the same model under some alternative sets of assumptions. Section 4 considers the question of efficiency of estimation, and Section 5 gives some calculations of the gain from using the extra moment conditions we have found. Section 6 considers the model with exogenous regressors, and shows how to categorize the moment conditions that are available and relevant in the dynamic model. Finally, Section 7 contains some concluding remarks.

2. MOMENT CONDITIONS UNDER STANDARD ASSUMPTIONS

We will consider the simple dynamic model

(1) $$y_{it} = \delta y_{i,t-1} + u_{it} \; ,$$
$$u_{it} = \alpha_i + \varepsilon_{it} \; , \quad i = 1, \ldots, N \; ; \; t = 1, \ldots, T \; .$$

Here i denotes cross sectional unit (individual) and t denotes time. The error u_{it} contains a time invariant individual effect α_i and random noise ε_{it}.

Note that there are NT observations for the regression, and that the initial observed value of y (for individual i) is y_{i0}; assumptions about the generation of y_{i0} are important in this literature. We can also write the T observations for person i as

(2) $$y_i = \delta y_{i,-1} + u_i ,$$

where $y_i' = (y_{i1},...,y_{iT})$ and $y_{i,-1}$ and u_i are defined similarly.

The model (1) does not contain any explanatory variables beyond the lagged dependent variable. The issue raised by the presence of exogenous regressors will be discussed in Section 6.

We assume that α_i and ε_{it} have mean zero for all i and t. (Nonzero mean of α can be handled with an intercept.) We also assume that all observations are independent across individuals. Furthermore, we will begin with the following "standard assumptions" (SA):

(SA.1) For all i, ε_{it} is uncorrelated with y_{i0} for all t.

(SA.2) For all i, ε_{it} is uncorrelated with α_i for all t.

(SA.3) For all i, the ε_{it} are mutually uncorrelated.

We note that these assumptions are implied by a variety of simple models, such as Chamberlain's (1984) projection model, which asserts that $Proj(y_{it}|\alpha_i,y_{i0},...,y_{i,t-1}) = \alpha_i + \delta y_{i,t-1}$, where $Proj(z|x)$ represents the population least squares projection of a variable z on a set of variables x = $(x_1,...,x_k)$. We also note that our assumptions on the initial value y_{i0} are weaker than those often made in the literature. See, for example, Bhargava and Sargan (1983), Hsiao (1986) and Blundell and Smith (1991) for examples of stronger assumptions about the initial value y_{i0}.

Under SA, it is obvious that the following moment conditions hold:

(3) $$E(y_{is}\Delta u_{it}) = 0 \; ; \; t = 2, ..., T; \; s = 0, ..., t-2 .$$

There are T(T-1)/2 such conditions. These are the moment conditions that are currently exploited in the literature. However, under SA there are additional moment conditions beyond those in (3). In particular, the following T-2 moment conditions also hold:

(4) $$E(u_{iT}\Delta u_{it}) = 0 \; ; \; t = 2, ..., T-1 .$$

The conditions (3)-(4) are a set of T(T-1)/2 + (T-2) moment conditions that follow directly from the assumptions that the ε_{it} are mutually uncorrelated and uncorrelated with α_i and y_{i0}. Furthermore, they represent <u>all</u> of the moment conditions implied by these assumptions. An informal but instructive way to see this is as follows. We cannot observe α_i or ε_{it}, but $u_{it} = (\alpha_i+\varepsilon_{it})$ is "observable" in the sense that it can be written in terms of data and parameters (which is what is relevant for GMM). The implications of SA for the u_{it} are simply:

(5A) $E(y_{i0}u_{it})$ is the same for all t.

(5B) $E(u_{is}u_{it})$ is the same for all $s \neq t$.

This is also a set of T(T-1)/2 + (T-2) moment conditions, and it is easy to show that it is equivalent to the conditions (3)-(4). The main advantage of (3)-(4) over (5) is just that it maximizes the number of <u>linear</u> moment conditions.

A more formal proof of the number of restrictions implied by SA can be given as follows. We distinguish those things that we make assumptions about, namely $(\varepsilon_{i1},\ldots,\varepsilon_{iT},y_{i0},\alpha_i)$, from things that we make assumptions about and that are "observable" in the sense discussed above, namely $(\alpha_i+\varepsilon_{1i},\ldots,\alpha_i+\varepsilon_{iT},y_{i0})$. The covariance matrix of the things we make assumptions about will be denoted Σ, as follows:

$$(6)\quad \Sigma = \operatorname{cov}\begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{iT} \\ y_{i0} \\ \alpha_i \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1T} & \sigma_{10} & \sigma_{1\alpha} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2T} & \sigma_{20} & \sigma_{2\alpha} \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot & \cdot \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_{TT} & \sigma_{T0} & \sigma_{T\alpha} \\ \sigma_{01} & \sigma_{02} & \cdots & \sigma_{0T} & \sigma_{00} & \sigma_{0\alpha} \\ \sigma_{\alpha 1} & \sigma_{\alpha 2} & \cdots & \sigma_{\alpha T} & \sigma_{\alpha 0} & \sigma_{\alpha\alpha} \end{bmatrix}$$

On the other hand, the covariance matrix of the "observables" will be denoted Λ, as follows:

$$\begin{bmatrix} \alpha_i+\varepsilon_{i1} \end{bmatrix} \quad \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1T} & \lambda_{10} \end{bmatrix}$$

$$(7) \quad \Lambda = \operatorname{cov}\begin{bmatrix} \alpha_i+\varepsilon_{i2} \\ \cdot \\ \cdot \\ \cdot \\ \alpha_i+\varepsilon_{iT} \\ y_{i0} \end{bmatrix} = \begin{bmatrix} \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2T} & \lambda_{20} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \lambda_{T1} & \lambda_{T2} & \cdots & \lambda_{TT} & \lambda_{T0} \\ \lambda_{01} & \lambda_{02} & \cdots & \lambda_{0T} & \lambda_{00} \end{bmatrix}$$

$$= \begin{bmatrix} (\sigma_{\alpha\alpha}+\sigma_{11}+2\sigma_{\alpha 1}) & (\sigma_{\alpha\alpha}+\sigma_{12}+\sigma_{\alpha 1}+\sigma_{\alpha 2}) & \cdots & (\sigma_{\alpha\alpha}+\sigma_{1T}+\sigma_{\alpha 1}+\sigma_{\alpha T}) & (\sigma_{0\alpha}+\sigma_{01}) \\ (\sigma_{\alpha\alpha}+\sigma_{12}+\sigma_{\alpha 1}+\sigma_{\alpha 2}) & (\sigma_{\alpha\alpha}+\sigma_{22}+2\sigma_{\alpha 2}) & \cdots & (\sigma_{\alpha\alpha}+\sigma_{2T}+\sigma_{\alpha 2}+\sigma_{\alpha T}) & (\sigma_{0\alpha}+\sigma_{02}) \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ (\sigma_{\alpha\alpha}+\sigma_{1T}+\sigma_{\alpha 1}+\sigma_{\alpha T}) & (\sigma_{\alpha\alpha}+\sigma_{2T}+\sigma_{\alpha 2}+\sigma_{\alpha T}) & \cdots & (\sigma_{\alpha\alpha}+\sigma_{TT}+2\sigma_{\alpha T}) & (\sigma_{0\alpha}+\sigma_{0T}) \\ (\sigma_{0\alpha}+\sigma_{01}) & (\sigma_{0\alpha}+\sigma_{02}) & \cdots & (\sigma_{0\alpha}+\sigma_{0T}) & \sigma_{00} \end{bmatrix}$$

Under SA, we have $\sigma_{\alpha t} = 0$ for all t, $\sigma_{0t} = 0$ for all t, and $\sigma_{ts} = 0$ for all $t \neq s$. Then Λ simplifies as follows:

$$(8) \quad \Lambda = \begin{bmatrix} (\sigma_{\alpha\alpha}+\sigma_{11}) & \sigma_{\alpha\alpha} & \cdots & \sigma_{\alpha\alpha} & \sigma_{0\alpha} \\ \sigma_{\alpha\alpha} & (\sigma_{\alpha\alpha}+\sigma_{22}) & \cdots & \sigma_{\alpha\alpha} & \sigma_{0\alpha} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \sigma_{\alpha\alpha} & \sigma_{\alpha\alpha} & \cdots & (\sigma_{\alpha\alpha}+\sigma_{TT}) & \sigma_{0\alpha} \\ \sigma_{0\alpha} & \sigma_{0\alpha} & \cdots & \sigma_{0\alpha} & \sigma_{00} \end{bmatrix}$$

There are T-1 restrictions, that λ_{0t} is the same for t = 1,...,T; and T(T-1)/2-1 restrictions, that λ_{ts} is the same for t,s = 1,...,T, $t \neq s$. Adding the number of restrictions, we get T(T-1)/2 + (T-2).

An intuitive way to understand the extra moment conditions that we have identified and the efficiency gains that they lead to is as follows. We can rewrite the model (1) as follows:

$$(9A) \quad \Delta y_{it} = \delta \Delta y_{i,t-1} + \Delta u_{it}, \; t = 2,3,\ldots,T \,,$$

$$(9B) \quad y_{iT} = \delta y_{i,T-1} + u_{iT} \quad .$$

The usual IV estimator amounts to estimating the first-differenced equations (9A) by three-stage least squares, imposing the restriction that δ is the same in every equations, where the instrument set is y_{i0} for t=2; (y_{i0},y_{i1}) for t=3; ...; $(y_{i0},...,y_{i,T-2})$ for t=T. There are no legitimate observable instruments for the levels equation (9B), but this equation is still useful in estimation because of the covariance restrictions that u_{iT} is uncorrelated with Δu_{it}, t=2,...,T-1. These covariance restrictions are exactly the extra moment conditions identified in equation (4). Of course, there are additional covariance restrictions in (9A); for example, Δu_4 is uncorrelated with Δu_2. However, these additional covariance restrictions are not useful in estimation because, unlike the covariance restrictions in equation (4), they are implied by the basic moment conditions in equation (3).

An interesting sidelight is that the moment conditions (4)-(5) hold under weaker conditions than SA. Define the set of modified assumptions (MA):

(MA.1) For all i, $\text{cov}(\varepsilon_{it},y_{i0})$ is the same for all t.

(MA.2) For all i, $\text{cov}(\varepsilon_{it},\alpha_i)$ is the same for all t.

(MA.3) For all i, $\text{cov}(\varepsilon_{it},\varepsilon_{is})$ is the same for all $t \neq s$.

Note that MA assumes equal covariance where SA assumes zero covariance. Nevertheless, the moment conditions implied by SA are also implied by MA, as is easily checked in the same way as above. Thus we can test whether $\text{cov}(\varepsilon_{it},\alpha_i)$ is constant over t, but we cannot test whether it is zero. Similarly, we can test constancy of $\text{cov}(\varepsilon_{it},y_{i0})$ but not whether it is zero.

3. SOME ALTERNATIVE ASSUMPTIONS

In this section we briefly consider some alternative sets of assumptions. The first case we consider is the one in which (SA.1)-(SA.3) are augmented by the additional assumption that the ε_{it} are homoskedastic. That is, suppose that we add the assumption:

(AA.4) For all i, $\text{var}(\varepsilon_{it})$ is the same for all t.

(Here "AA" is short for "added assumption.") This assumption, when added to (SA.1)-(SA.3), generates the additional (T-1) moment conditions that

(10) $E(u_{it}^2)$ is the same for t = 1, ..., T.

(In terms of equation (8) above, λ_{tt} is the same for t = 1,...,T.) Therefore the total number of moment conditions becomes T(T-1)/2 + (2T-3). These moment conditions can be expressed as (3), (4) and (10). Alternatively, if we wish to maximize the number of linear moment conditions, Ahn and Schmidt (1990) show that these moment conditions can be expressed as (3) plus the additional conditions

(11A) $E(y_{it}\Delta u_{i,t+1} - y_{i,t+1}\Delta u_{i,t+2}) = 0$, t = 1, ..., T-2

(11B) $E(\bar{u}_i\Delta u_{i,t+1}) = 0$, t=1, ... ,T-1 ,

where $\bar{u}_i = T^{-1}\Sigma_{t=1}^{T}u_{it}$. Comparing this to the set of moment conditions without homoskedasticity [(3)-(4)], we see that homoskedasticity adds T-1 moment conditions <u>and</u> it allows T-2 previously nonlinear moment conditions to be expressed linearly.

Another possible assumption is the "stationarity" assumption of Arellano and Bover (1990). They assume that

(AA.5) $E(\alpha_i y_{it})$ is the same for all t.

This is equivalent in our framework to asserting that $\sigma_{\alpha\alpha} = (1-\delta)\sigma_{0\alpha}$. If SA and (AA.4) are maintained, (AA.5) adds one moment condition to the set consisting of (3) and (11). However, it also allows all of the moment conditions to be expressed linearly. In particular, given this condition, we can write the set of usable moment conditions as (4) plus

(12A) $E(u_{iT}\Delta y_{it}) = 0$, t = 1,...,T-1 ;

(12B) $E(u_{it}y_{it} - u_{i,t-1}y_{i,t-1}) = 0$, t = 2,...,T .

This is a set of T(T-1)/2 + (2T-2) moment conditions, all of which are linear in δ.

An interesting question that this paper will not address is how much our assumptions can be weakened without losing all moment conditions. Ahn and Schmidt (1990, Appendix) give a partial answer by counting the moment conditions implied by many possible combinations of (SA.1), (SA.2), (SA.3) and (A.4). The reader is referred to that paper for details.

4. ASYMPTOTIC EFFICIENCY

GMM based on the full set of moment conditions (3)-(4) makes efficient use of second moment information, and intuitively we should expect it to be asymptotically equivalent to other estimators that make efficient use of second moment information, notably Chamberlain's (1982, 1984) minimum distance estimator. We now proceed to prove this asymptotic equivalence.

To define some notation, let $\eta_i = (u_{i1},...,u_{iT},y_{i0})'$ be the vector of "observables," with $cov(\eta_i) = \Lambda$ as given in (7) above, and let $\xi_i = (y_{i0},y_{i1},...,y_{iT})'$ be the vector of observed values of y, with covariance matrix Ω. (Recall that η_i is "observable" in the sense that it can be written as a function of data and δ.) We assume that the η_i (or, equivalently, the ξ_i) are i.i.d across individuals. The assumption of homoskedasticity is required for the existence of a consistent minimum distance estimator.

The connection between η_i and ξ_i is linear: $\xi_i = D\eta_i$, $\eta_i = D^{-1}\xi_i$, where D is a (T+1)×(T+1) nonsingular matrix that depends on δ. Specifically, $D_{h,T+1} = \delta^{h-1}$, h = 1,...,T+1; for j = 1,...,T and h = 1,...,T+1, $D_{hj} = 0$ for $h\leq j$ and $D_{hj} = \delta^{|h-j-1|}$ for h>j. (If we rearrange η_i so that y_{i0} comes first instead of last, then D is lower triangular and $D_{hj} = \delta^{|h-j|}$ for $h\geq j$.) Thus $\Omega = D\Lambda D'$. Define

(13) $$S = N^{-1}\Sigma_i\xi_i\xi_i' \ , \ Z = N^{-1}\Sigma_i\eta_i\eta_i' \ ;$$

note that S = DZD'. Define the vectors of parameters

(14) $$\theta = (\sigma_{00},\sigma_{\alpha\alpha},\sigma_{0\alpha},\sigma_{11},...,\sigma_{TT})', \ \gamma = (\delta,\theta')' \ ,$$

so that Λ depends on θ, D depends on δ, and Ω depends on γ. Finally, define $\triangle = cov[vec(\eta_i\eta_i')]$, and note that $cov[vec(\xi_i\xi_i')] = \psi = (D\otimes D)\triangle(D\otimes D)'$.

In this notation, the basic results of Chamberlain (1984) can be stated as follows. The quasi-maximum likelihood (QML) estimator solves the problem

(15) $$\max_\gamma L = (N/2)\ln|\Omega^{-1}| - (N/2)trace(\Omega^{-1}S).$$

Its asymptotic covariance matrix is of the form

(16) $$C_{QML} = [G'H'(\Omega\otimes\Omega)^{-1}HG]^{-1}G'H'(\Omega\otimes\Omega)^{-1}\psi(\Omega\otimes\Omega)^{-1}HG[G'H'(\Omega\otimes\Omega)^{-1}HG]^{-1}.$$

Here $G = \partial v(\Omega)/\partial\gamma'$, where $v(\Omega)$ equals $vec(\Omega)$ with the above-diagonal elements of Ω deleted; and H is the "duplication matrix" [e.g., Magnus and Neudecker

(1988, p. 49)] such that $Hv(\Omega) = vec(\Omega)$. The minimum distance (MD) estimator based on the observables ξ_i solves the problem

(17) $\min_{\gamma} [v(S)-v(\Omega)]'A_N[v(S)-v(\Omega)]$

and its asymptotic covariance matrix is of the form

(18) $C_{MD} = (G'AG)^{-1}G'A(J\psi J')^{-1}AG(G'AG)^{-1}$

where $J = (H^+)' = (H'H)^{-1}H'$ is the "elimination matrix" such that $v(\Omega) = Jvec(\Omega)$, and where $A = \text{plim } A_N$. The QMLE is a MD estimator, with $A = H'(\Omega\otimes\Omega)^{-1}H$. This choice of A is suboptimal except under normality. The optimal minimum distance (OMD) estimator uses $A = (J\psi J')^{-1}$, and its asymptotic covariance matrix is of the form

(19) $C_{OMD} = [G'(J\psi J')^{-1}G]^{-1}$.

The OMD estimator is asymptotically efficient relative to the QML estimator, but they are equally efficient under normality.

Normal quasi-maximum likelihood estimation of the dynamic panel data model has been discussed by Bhargava and Sargan (1983) and Hsiao (1986, section 4.3.2) under a variety of assumptions about the generation of y_{i0}, including assumptions equivalent to ours (y_{i0} correlated with α_i and stationarity not imposed). Chamberlain's results indicate that these estimators are asymptotically dominated by estimators that make efficient use of second moment information but do not impose normality, such as the OMD estimator and (as we shall see) our GMM estimator.

Our GMM estimator is based upon the covariance matrix of the "observables" η_i instead of the observed ξ_i. Accordingly, we now consider OMD estimation based upon $\Lambda = cov(\eta_i)$ instead of upon $\Omega = cov(\xi_i)$, as above; that is, we now consider $Z = \Sigma_i\eta_i\eta_i'$ instead of $S = \Sigma_i\xi_i\xi_i'$. This alternative OMD estimator solves the problem

(20) $\min_{\gamma} [v(Z)-v(\Lambda)]'B_N[v(Z)-v(\Lambda)]$,

where $\text{plim } B_N = (J\triangle J')^{-1}$, and where it should be remembered that Z depends on δ while Λ depends on θ. Since $S = DZD'$ and $\Omega = D\Lambda D'$, the information content in $[v(S)-v(\Omega)]$ and $[v(Z)-v(\Lambda)]$ is the same. Thus it is reasonable that the OMD estimators based on (17) and on (20) should be asymptotically equivalent (though they may differ in finite samples due to the way that the weighting matrices are evaluated). We show in Appendix 1 that this is so. Thus it doesn't matter asymptotically whether OMD estimation is based on cov(η_i) or on cov(ξ_i).

We now wish to show that the OMD estimator based on (20) is asymptotically equivalent to our GMM estimator. Define $m(\gamma) = v[Z(\delta)-\Lambda(\theta)]$, so that the OMD estimator is a GMM estimator based on the moment condition $E[m(\gamma)] = 0$. Now consider a possible transformation of these moment conditions: $b(\gamma) = Fm(\gamma)$, where F is nonstochastic and nonsingular. Obviously GMM based on $b(\gamma)$ is the same as GMM based on $m(\gamma)$ since we have just taken a nonsingular linear combination of the moment conditions. (In fact, this is essentially the same result as just given in the previous paragraph.) In particular, as discussed in Appendix 1, there exists a matrix F such that

(21) $$b(\gamma) = Fm(\gamma) = [h(\delta)', (p(\delta)-\theta)']'$$

where the first subset of transformed moment conditions, $h(\delta)$, is the set of moment conditions exploited by our GMM estimator, and the second subset of transformed moment conditions, $p(\delta)-\theta$, just determines θ in terms of δ. It follows [see, e.g., Abowd and Card (1989, appendix)] that GMM based on $h(\delta)$ only yields the same estimate of δ as GMM based on the entire set of moment conditions $b(\gamma)$ [or $m(\gamma)$].

The fact that our GMM estimator based on the moment conditions (3)-(4) is asymptotically equivalent to Chamberlain's optimal minimum distance estimator confirms the intuition that the GMM estimator is efficient in the class of estimators that make use of second-moment information. In this regard, it should be noted explicitly that (SA.1)-(SA.3) are stated in terms of uncorrelatedness only, so that only second-moment information is relevant. If we replaced uncorrelatedness with independence, for example, additional moment conditions involving higher-order moments would (at least in principle) be potentially relevant. These additional moment conditions would not lead to gains in asymptotic efficiency when η_i is normal; because our model is linear in η_i, under normality only second moments matter.

The GMM/minimum distance estimator reaches the semiparametric efficiency bound discussed by Newey (1990). A sketch of the proof is given in Appendix 2. This result is intuitively reasonable. The unrestricted estimator of $v(\Omega)$, $v(S)$, reaches the semiparametric efficiency bound $J\psi J'$, using standard results of Chamberlain (1987) and Newey (1990) on the semiparametric efficiency of the sample mean. The minimum distance estimator efficiently imposes the restrictions on $v(\Omega)$, thus preserving semiparametric efficiency. Further discussion can be found in Chamberlain (1987).

Even though under SA the optimal minimum distance estimator shares the asymptotic properties of our GMM estimator, the latter may be preferred for two reasons. First, the GMM estimator of δ can be more easily calculated, because the nuisance parameters ($\sigma_{\alpha\alpha}$, σ_{00}, $\sigma_{0\alpha}$, σ_{11}, ..., σ_{TT}) need not be considered in the GMM estimation procedure. Second, the consistency and asymptotic normality of the GMM estimator require only SA and cross-sectional independence of η_i and ξ_i, while cross-sectional homoskedasticity is also required for the consistency of the minimum distance estimator.

5. CALCULATIONS OF ASYMPTOTIC VARIANCES

In this section we attempt to quantify the gains in asymptotic efficiency that arise from using the extra moment conditions identified in this paper. We consider three estimators. (i) Let $\hat{\delta}_{IV}$ be GMM using (3) only. This is the estimator suggested in the existing literature. Under homoskedasticity it can be calculated as linear IV on the first-differenced equation. (ii) Let $\hat{\delta}_{GMM1}$ be GMM using (3) and (4). This estimator exploits the extra nonlinear moment conditions that arise under SA. (iii) Let $\hat{\delta}_{GMM2}$ be GMM using (3) and (11). This estimator assumes homoskedasticity as well as SA. Asymptotically $\hat{\delta}_{IV}$ is least efficient and $\hat{\delta}_{GMM2}$ is most efficient.

We will quantify the differences in efficiency by calculating the asymptotic covariance matrices of the three estimates, for given parameter values. The asymptotic covariance matrices can be found in Ahn (1990) and Ahn and Schmidt (1990). The relevant parameter values are T, δ, σ_{00}, $\sigma_{\alpha\alpha}$, $\sigma_{0\alpha}$, σ_{11}, ..., σ_{TT}. However, in this section we will consider only the case that the ε_{it} are i.i.d., so that the T variances σ_{11}, ..., σ_{TT} can be replaced by a single variance, say $\sigma_{\varepsilon\varepsilon}$. Furthermore, because our results are invariant to scale, a normalization of the variances is possible, and we pick $\sigma_{\varepsilon\varepsilon} = 1$ as our normalization.

In Table 1 we report some results for the case in which y is stationary. Thus the errors are homoskedastic so that the extra moment conditions underlying $\hat{\delta}_{GMM2}$ are valid. Furthermore, we set σ_{00} and $\sigma_{0\alpha}$ according to the stationarity conditions $\sigma_{0\alpha} = \sigma_{\alpha\alpha}/(1-\delta)$, $\sigma_{00} = \sigma_{\alpha\alpha}/(1-\delta)^2 + \sigma_{\varepsilon\varepsilon}/(1-\delta^2)$. This leaves three relevant parameters: T, δ, and $\sigma_{\alpha\alpha}$. We present results for T = 3, 4 and 10; for δ ranging from -.99 to .99; and for $\sigma_{\alpha\alpha}$ (to be interpreted as $\sigma_{\alpha\alpha}/\sigma_{\varepsilon\varepsilon}$) ranging from zero to four. The results are the ratios of asymptotic variances [e.g., $var(\hat{\delta}_{IV})/var(\hat{\delta}_{GMM2})$].

The efficiency gains found in Table 1 from using additional moment conditions are certainly non-trivial. They increase as T increases, which is perhaps surprising, since the proportion of extra moment conditions to total moment conditions decreases with T. (The numbers of moment conditions for the three estimators are 3, 4 and 6 for T=3; 6, 8 and 11 for T=4; and 45, 53 and 62 for T=10.) The efficiency gains also increase as $\sigma_{\alpha\alpha}/\sigma_{\varepsilon\varepsilon}$ increases, which is not surprising. The efficiency gains increase as δ increases toward one, a finding predicted correctly by one of the referees of an earlier draft of this paper. When δ is close to unity, the "first stage regressions" in the calculation of $\hat{\delta}_{IV}$ are essentially regressions of differences on levels, and have little explanatory power. Finally, the efficiency gains of $\hat{\delta}_{GMM1}$ over $\hat{\delta}_{IV}$ are often substantial, while the additional efficiency gains of $\hat{\delta}_{GMM2}$ over $\hat{\delta}_{GMM1}$ are usually small. For example, for T = 4, δ = 0.9 and $\sigma_{\alpha\alpha}/\sigma_{\varepsilon\varepsilon} = 1$, the asymptotic variance of $\hat{\delta}_{IV}$ is 3.4 times as large as the asymptotic variance of $\hat{\delta}_{GMM1}$ and 4.3 times as large as the asymptotic variance of $\hat{\delta}_{GMM2}$, so the asymptotic variance of $\hat{\delta}_{GMM1}$ is only about 1.3 times as large as the asymptotic variance of $\hat{\delta}_{GMM2}$. Most of our efficiency gains come from imposing the extra moment conditions (4) that follow from SA, not from the moment conditions that follow from the additional assumption of homoskedasticity.

In Table 2 we report results for some cases in which we do not impose stationarity, though we still impose homoskedasticity (with $\sigma_{\varepsilon\varepsilon}$ normalized to unity). We will not discuss these results in detail, but we note that there are exceptions to some of the conclusions from the results in Table 1. However, the gains from our extra moment conditions are, if anything, larger than they were in the stationary case.

This model is heavily overidentifed, in the sense that there are many more moment conditions than parameters to estimate. It is a conventional wisdom that the asymptotic properties of estimators may be a poor guide to their finite sample properties in heavily overidentified models. While we did not conduct any simulations on our model, Arellano and Bond (1991, section 4) provide some relevant Monte Carlo evidence. Their model is the same as ours, except that it contains an exogenous regressor, which was used as a single instrument. Their estimators GMM1 and GMM2 correspond to the set of T(T-1)/2 moment conditions exploited by our $\hat{\delta}_{IV}$ (they differ in finite samples in the way in which the weighting matrix is evaluated), while their estimators AHd and AH1 follow Anderson and Hsiao (1981) in using only a single moment condition based on lagged y's (using instruments $\Delta y_{i,t-2}$ and $y_{i,t-2}$ respectively). Their results are for T=7 so the model is highly overidentified. They report (p. 284) that the extra instruments result in substantial finite sample efficiency gains, and also (p. 285) that the asymptotic standard errors were close to the finite sample standard deviations of the estimates. Their model and estimation procedures are close enough to ours to hope that their optimistic results would generalize.

6. Moment Conditions with Exogenous Variables

We now consider the regression model with explanatory variables:

(22) $$y_{it} = \delta y_{i,t-1} + X_{it}\beta + Z_i\gamma + u_{it}, \quad u_{it} = \alpha_i + \varepsilon_{it} \ .$$

Here X_{it} is a $1 \times k$ vector of time-varying explanatory variables and Z_i is a $1 \times g$ vector of time invariant explantory variables. We create data vectors and matrices by ordering observations first by individual and then by time. Thus, for example,

(23) $$y = (y_{11}, \ldots , y_{1T}; \ldots ; y_{N1} , \ldots , y_{NT})' .$$

With this convention the model in matrix form is

(24) $$y = y_{-1}\delta + X\beta + Z\gamma + u = W\xi + u ,$$

$$u = \alpha + \varepsilon, \; W = (y_{-1}, X, Z), \; \xi' = (\delta, \beta', \gamma').$$

We adopt the following standard notation for projections. For any matrix A, let P_A be the projection onto the column space of A, so that $P_A = A(A'A)^{-1}A'$ if the inverse exists; let $Q_A = I-P_A$. For any integer m, let e_m be an mx1 vector of ones, and define $P = e_T e_T'/T$ and $Q = I-P$. Then define

(25) $$V = I_N \otimes e_T \; ; \; P_V = I_N \otimes e_T e_T'/T = I_N \otimes P; \; Q_V = I_N - P_V = I_N \otimes Q \; ;$$

so that V is a matrix of individual dummy variables, P_V is the $NT \times NT$ idempotent matrix that converts an $NT \times 1$ vector ordered as in (23) above into a vector of individual means, and Q_V is the $NT \times NT$ idempotent matrix that converts an $NT \times 1$ vector into deviations from individual means. The spaces spanned by Q_V and P_V are orthogonal and exhaustive: $P_V Q_V = 0$, $P_V + Q_V = I_{NT}$.

We assume independence of all variables across individuals (different values of i). We also assume that the regressors X and Z are strongly exogenous with respect to ε; we will later distinguish variables that are or are not correlated with α. We will not make detailed assumptions of regularity conditions on X and Z, but rather simply assume that the relevant second moment matrices converge to nonsingular limits as $N\to\infty$.

In order to explain the efficient use of exogenous variables in the dynamic model, we first give a brief summary of existing results for the static model that does not contain the regressor y_{-1}, we consider the model

(26) $$y = X\beta + Z\gamma + u \; , \; u = \alpha + \varepsilon \; .$$

Under our assumptions about α and ε, the covariance structure of u is:

(27) $$\Omega^{-1} = [\mathrm{cov}(u)]^{-1} = Q_V + \theta^2 P_V \; , \; \Omega^{-1/2} = Q_V + \theta P_V \; ,$$

where Q_V and P_V are defined in (25), $\theta^2 = \sigma_{\varepsilon\varepsilon}/(\sigma_{\varepsilon\varepsilon}+T\sigma_{\alpha\alpha})$, and the equalities in (27) actually hold up to an irrelevant factor of proportionality. Now transform (26) by $\Omega^{-1/2}$:

(28) $$\Omega^{-1/2}y = \Omega^{-1/2}X\beta + \Omega^{-1/2}Z\gamma + \Omega^{-1/2}u$$

so that the error in (28) is whitened. The estimators that we will generalize are IV estimators of (28), using instruments of the form $G = [Q_VX,P_VJ]$, where J is to be defined. As noted by Breusch, Mizon and Schmidt, IV of (28) using instruments G of this form is equivalent to GMM based on the orthogonality condition $E(G'u) = 0$, because the optimal weighting matrix in GMM implicitly performs the $\Omega^{-1/2}$ transformation.

The precise form of J in the instrument set varies across authors, and its relationship to exogeneity assumptions is not entirely explicit in the literature. Following Hausman and Taylor, we partition X and Z: $X = [X_1,X_2]$, $Z = [Z_1,Z_2]$, where X_1 and Z_1 are uncorrelated with α but X_2 and Z_2 are correlated with α. As a matter of notation, let the column dimensions of X_1, X_2, Z_1 and Z_2 be k_1, k_2, g_1 and g_2, respectively. We also follow Breusch, Mizon and Schmidt in defining the following notation. For any (arbitrary) $NT \times h$ matrix S, with representative row S_{it} (of dimension $1 \times h$), define the $N \times ht$ matrix S^o as:

(24) $$S^o = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1T} \\ S_{11} & S_{12} & \cdots & S_{1T} \\ \vdots & \vdots & & \vdots \\ S_{11} & S_{12} & \cdots & S_{1T} \end{bmatrix}$$

and define the $NT \times hT$ matrix $S^* = S^o \otimes e_T$. (Note that all T values of S_{it} are separate values in S^*.) Then we have the following instrument sets (where HT stands for Hausman and Taylor (1981), BMS stands for Breusch, Mizon and Schmidt (1989), and AM stands for Amemiya and MaCurdy (1986)):

(30) $$J_{HT} = [X_1,Z_1]\ ,\ J_{AM} = [X_1*,Z_1],\ J_{BMS} = [X_1*,(Q_VX_2)*,Z_1]\ ,$$

As in Breusch, Mizon and Schmidt, $(Q_vX_2)^*$ indicates that any (T-1) independent deviations from means of each variable in X_2 are used separately as instruments.

In the set of instruments $G = [Q_vX, P_vJ]$, obviously P_vJ is in the space spanned by P_v while Q_vX is in the space spanned by Q_v. Hausman and Taylor actually defined their instrument set to be of the form $\bar{G} = [Q_v, J]$, so that all of Q_v was included in the instrument set, not just Q_vX. Breusch, Mizon and Schmidt showed that this did not matter, for the static model. Although the projections onto G and $\bar{G}$ are different, these instrument sets lead to identical estimators for the static model. This is so because, if we conceptually separate the regressors in (28) into Q_vX, P_vX and Z, instruments in Q_v space other than Q_vX have no explanatory power in the first-stage regressions. Therefore, <u>for the static model</u>, there is no point in looking for additional instruments in the space spanned by Q_v. However, <u>for the dynamic model</u>, Q_v is not a legitimate part of the instrument set; this is the reason the within estimator is inconsistent. Therefore functions of X and Z that lie in the space spanned by Q_v can be legitimate instruments and can make a difference to the estimator, by helping to better "explain" the regressor Q_vy_{-1}. As we will see, given strong exogeneity of X and Z with respect to ε, many such instruments exist, in addition to Q_vX.

In order to find such instruments, we now make explicit exogeneity assumptions, and make clear how they relate to the existing different treatments of the static model.

(E.1) For all i, X_{it} is uncorrelated with ε_{is} for all t and s.

(E.2) For all i, Z_i is uncorrelated with ε_{it} for all t.

(E.3) For all i, $X_{1,it}$ is uncorrelated with α_i for all t.

(E.4) For all i, $Z_{1,i}$ is uncorrelated with α_i.

(E.5) For all i, $E(X_{2,it},\alpha_i)$ is the same for all t.

Assumptions E.1 and E.2 say that X and Z are strongly exogenous with respect to the noise ε. Assumptions E.3 and E.4 say that X_1 and Z_1 are also uncorrelated with α, but they allow X_2 and Z_2 to be correlated with α.

Finally, E.5 is consistent with the view that the correlation between X_2 and α is due solely to a time-invariant component of X_2 that is correlated with α. The Breusch, Mizon and Schmidt estimator implicitly depends all five of these assumptions, while the Amemiya and MaCurdy estimator depends only on the first four. The Hausman and Taylor estimator depends on E.1, E.2, E.4, and a weakened version of E.3, namely, that $\bar{X}_{1,i}$ is uncorrelated with α_i for all i.

The exogeneity conditions E.1 - E.4 immediately imply the following usable moment conditions.

(M.1) $E(X_{1,it}u_{is}) = 0$ for all t and s.

(M.2) $E(Z_{1,i}u_{it}) = 0$ for all t.

(M.3) For t = 1,...,T, $E(X_{2,it}u_{is})$ is the same for all s. (It may depend on t.)

(M.4) $E(Z_{2,i}u_{it})$ is the same for all t.

In addition, if we add exogeneity condition E.5 we have the additional moment condition

(M.5) $E(X_{2,it}u_{is})$ is the same for all t.

Note that M.3 and M.5 together imply that $E(X_{2,it}u_{is})$ is the same for all t and s.

We now proceed to write these conditions in matrix form, and to show that the GMM estimator that imposes these conditions in the static model is equivalent to the Breusch, Mizon and Schmidt estimator. To do so, we define a little more notation. First, we define the T × (T-1) (first-differencing) matrix L as follows:

$$L = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \qquad (31)$$

Second, define the N × k_1T matrix X_1^o and the N × k_2T matrix X_2^o as in (29), and define the N × g_1 matrix Z_1^o and the N × g_2 matrix Z_2^o as follows:

(32) $$Z_1^o = \begin{bmatrix} z_{1,1} \\ \cdot \\ \cdot \\ \cdot \\ z_{1,N} \end{bmatrix} , \quad Z_2^o = \begin{bmatrix} z_{2,1} \\ \cdot \\ \cdot \\ \cdot \\ z_{2,N} \end{bmatrix}$$

Then, following Schmidt, Ahn and Wyhowski (1992), we can express the moment conditions M.1-M.4 as follows:

(33) $E(R_1'u) = 0$, $R_1 = [X_1^o \otimes I_T, X_2^o \otimes L, Z_1^o \otimes I_T, Z_2^o \otimes L]$.

Note that each variable in X_1 generates T^2 instruments; each variable in X_2 generates T(T-1) instruments; each variable in Z_1 generates T instruments; and each variable in Z_2 generates (T-1) instruments. Furthermore, if (under exogeneity assumption E.5) we add moment condition M.5, we can express moment conditions M.1-M.5 as follows:

(34) $E(R_2'u) = 0$, $R_2 = [R_1,(Q_V X_2)^*] = [X_1^o \otimes I_T, X_2^o \otimes L, Z_1^o \otimes I_T, Z_2^o \otimes L,(Q_V X_2)^*]$.

To compare GMM based on the instrument sets R_1 and R_2 to the Amemiya and MaCurdy and the Breusch, Mizon and Schmidt estimators, we use the following notation. We define the relationship "$\overset{P}{=}$" between two matrices to mean that they yield the same projection; that is, $A \overset{P}{=} B$ if $P_A = P_B$. Specifically, we recall the T × T idempotent matrix $Q = I_T - e_T e_T'/T$, and we note that $P_L \overset{P}{=} P_Q = Q$, $I_T \otimes L \overset{P}{=} Q_V$. We also define the NT × k_1T(T-1) matrix $X_1^{**} = X_1^o \otimes L$, with X_2^{**}, Z_1^{**} and Z_2^{**} defined similarly. Then we can state the following results:

(35A) $R_1 \overset{P}{=} [X_1^{**}, X_2^{**}, Z_1^{**}, Z_2^{**}; X_1^*, Z_1]$

(35B) $R_2 \overset{P}{=} [X_1^{**}, X_2^{**}, Z_1^{**}, Z_2^{**}; X_1^*, (Q_V X_2)^*, Z_1]$.

To understand these results, we consider the instruments $X_1^o \otimes I_T$ that appear in both R_1 and R_2, and which represent the moment conditions M.1. We can rewrite these as follows: $X_1^o \otimes I_T = Q_V(X_1^o \otimes I_T) + P_V(X_1^o \otimes I_T) \overset{P}{=} [X_1^o \otimes Q, X_1^*] \overset{P}{=} [X_1^o \otimes L, X_1^*] = [X_1^{**}, X_1^*]$. Note that there are k_1T^2 instruments, corresponding to the k_1T^2 moment conditions in M.1; k_1T(T-1) are in Q_V space and k_1T are in P_V space.

The remaining moment conditions are similar. The moment conditions M.2 are represented by the instruments $Z_1^o \otimes I_T$, which can be rewritten as follows: $Z_1^o \otimes I_T \overset{P}{=} [Z_1^o \otimes L, Z_1] = [Z_1^{**}, Z_1]$. There are g_1T moment conditions in M.2, and there are $g_1(T-1)$ instruments in Q_V space and g_1 instruments in P_V space. The moment conditions M.3 are represented by the instruments $X_2^o \otimes L = X_2^{**}$. There are $k_2T(T-1)$ instruments, all in Q_V space. The moment conditions M.4 are represented by the instruments $Z_2 \otimes L = Z_2^{**}$. There are $g_2(T-1)$ instruments, all in Q_V space. Finally, the moment conditions M.5 add the additional $k_2(T-1)$ instruments $(Q_vX_2)^*$ to R_2; all of these instruments are in P_v space.

From (35A) we can see that the instrument set R_1 is larger than the Amemiya and MaCurdy instrument set $G_{AM} = [Q_vX, X_1^*, Z_1]$ but smaller than the instrument set $\bar{G}_{AM} = [Q_v, X_1^*, Z_1]$. However, for the static model, R_1, G_{AM} and $\bar{G}_{AM}$ all lead to the same estimator, since additional instruments in Q_v space are irrelevant in the static model. Similarly, from (35B) we see that the instrument set R_2 is larger than the Breusch, Mizon and Schmidt instrument set $G_{BMS} = [Q_VX, X_1^*, (Q_VX_2)^*, Z_1]$, and not as large as the instrument set $\bar{G}_{BMS} = [Q_V, X_1^*, (Q_VX_2)^*, Z_1]$; however, for the static model, R_2, G_{BMS} and $\bar{G}_{BMS}$ all lead to the same estimator. Thus, for the static model, GMM based on the moment conditions M.1-M.4 is equivalent to the Amemiya-MaCurdy estimator, and GMM based on the moment conditions M.1-M.5 is equivalent to the Breusch, Mizon and Schmidt estimator.

We now return to the dynamic model (24) that contains y_{-1} as a regressor, as well as X and Z. Now the extra instruments that are in R_1 but not in G_{AM}, or in R_2 but not in G_{BMS}, are relevant because they help to "explain" the regressor (Q_Vy_{-1}). Thus, if we maintain the assumptions SA.1-SA.3 and E.1-E.4, the available moment conditions are given by (3), (4) and (33). If we add assumption E.5, the set of moment conditions becomes (3), (4) and (34).

Estimation of the model is a straightforward application of GMM. The moment conditions (3) and (33) [or (34)] are linear in the parameters, while the moment conditions (4) are nonlinear. A linearized GMM estimator is possible, along the line of Newey (1985). Some discussion of computational details is given in Ahn (1990) and Ahn and Schmidt (1990).

7. CONCLUDING REMARKS

In this paper, we have extended the existing literature on the dynamic panel data model in two directions. First, under standard assumptions, we have counted the additional moment conditions not used by existing estimators, and shown how to exploit these moment conditions efficiently. These extra moment conditions can lead to substantial improvements in the efficiency of estimation, leading to reductions in asymptotic variance on the order of two or three. As a result we expect these extra moment conditions to be of practical importance in empirical work. Second, when the model contains exogenous variables in addition to the lagged dependent variable, we have shown how to exploit the exogeneity assumptions about these variables efficiently. Moment conditions (or instruments) that do not increase the efficiency of estimation in the static model do increase efficiency in the dynamic model.

It is also possible that the analysis of this paper could be extended within the dynamic panel data model. Many of the "random effects" treatments of this model that are currently found in the literature rely on stronger assumptions than we have made, and they could profitably be recast in terms of the moment conditions that they imply. For example, stationarity of the y process has been considered by Arellano and Bover (1990). It implies that $E(\alpha_i y_{it})$ is the same for all t, and leads to imposable moment conditions. Stationarity also implies a restriction on the variance of y_{i0} that leads to an imposable moment condition not considered by Arellano and Bover. Presumably other specific assumptions about initial conditions would have similar implications.

APPENDIX 1

Equivalence of MD Estimators

We wish to show that the OMD estimator based on (17) is asymptotically equivalent to the OMD estimator based on (20). To do so we show that the minimands of the two estimators are asymptotically identical; that is, we show that

$$\text{(A.1)}\qquad [v(S)-v(\Omega)]'(J\psi J')^{-1}[v(S)-v(\Omega)] = [v(Z)-v(\Lambda)]'(J\Delta J')^{-1}[v(Z)-v(\Lambda)] \ .$$

To do so, note

$$\text{(A.2)}\qquad \begin{aligned} v(S)-v(\Omega) &= J[vec(S)-vec(\Omega)] \\ &= J[vec(DZD')-vec(D\Lambda D')] \\ &= J(D\otimes D)[vec(Z)-vec(\Lambda)] \\ &= J(D\otimes D)H[v(Z)-v(\Lambda)]. \end{aligned}$$

Similarly, $J\psi J' = J(D\otimes D)\Delta(D\otimes D)'J'$. Therefore the first expression in (A.1) becomes

$$\text{(A.3)}\qquad \begin{aligned} &[v(S)-v(\Omega)]'(J\psi J')^{-1}[v(S)-v(\Omega)] \\ &= [v(Z)-v(\Lambda)]'H'(D\otimes D)J'[J(D\otimes D)\Delta(D\otimes D)'J']^{-1} \\ &\qquad \cdot J(D\otimes D)H[v(Z)-v(\Lambda)]. \end{aligned}$$

To simplify this, use the fact [Magnus and Neudecker (1988, p. 50, equation (14)] that $(D\otimes D)HJ = HJ(D\otimes D)$ and the fact that $JH = I$, which imply that $J(D\otimes D) = JHJ(D\otimes D) = J(D\otimes D)HJ$. With this substitution, the inverse on the right hand side of (A.3) becomes

$$\text{(A.4)}\qquad [J(D\otimes D)HJ\Delta J'H'(D\otimes D)'J']^{-1} = [H'(D\otimes D)'J']^{-1}(J\Delta J')^{-1}[J(D\otimes D)H]^{-1}$$

and the right hand side of (A.3) simplifies to

$$\text{(A.5)}\qquad [v(Z)-v(\Lambda)]'(J\Delta J')^{-1}[v(Z)-v(\Lambda)]$$

which is the same as the right hand side of (A.1).

Explanation of Equation (21) of the Text

There are (T+1)(T+2)/2 moment conditions in $m(\gamma)$. The idea of equation (21) is to split these into the T(T-1)/2 + (T-2) moment conditions in $h(\delta)$ plus the additional (T+3) moment conditions in $p(\delta)-\theta$ that determine θ in terms of δ. Note that θ is as defined in equation (14) of the main text. The choice of $p(\delta)$ is not unique, and we pick $p(\delta) = N^{-1}\Sigma_i p_i(\delta)$, with

(A.6) $p_i(\delta) = [y_{i0}^2,\ u_{i1}u_{i2},\ u_{i1}y_{i0},\ u_{i1}^2-u_{i1}u_{i2},\ \ldots,\ u_{iT}^2-u_{i1}u_{i2}]$

Similarly, let $h(\delta) = N^{-1}\Sigma_i h_i(\delta)$, $m(\gamma) = N^{-1}\Sigma_i m_i(\gamma)$. Then we wish to find F such that $Fm_i(\gamma) = [h_i(\delta)', p_i(\delta)-\theta]$. To save space we exhibit this transformation for the simplest possible case, T = 3. The transformation for T>3 is a straightforward extension of the expression in (A.7).

$$
\text{(A.7)}\quad
\begin{bmatrix}
y_{i0}\Delta u_{i2} \\
y_{i0}\Delta u_{i3} \\
y_{i1}\Delta u_{i3} \\
u_{i3}\Delta u_{i2} \\
y_{i0}^2 - \sigma_{00} \\
u_{i1}u_{i2}-\sigma_{\alpha\alpha} \\
u_{i1}y_{i0}-\sigma_{0\alpha} \\
u_{i1}^2-u_{i1}u_{i2}-\sigma_{11} \\
u_{i2}^2-u_{i1}u_{i2}-\sigma_{22} \\
u_{i3}^2-u_{i1}u_{i2}-\sigma_{33}
\end{bmatrix}
=
\begin{bmatrix}
0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & -\delta & 0 & \delta & 0 \\
0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
u_{i1}^2-\sigma_{\alpha\alpha}-\sigma_{11} \\
u_{i1}u_{i2}-\sigma_{\alpha\alpha} \\
u_{i1}u_{i3}-\sigma_{\alpha\alpha} \\
u_{i1}y_{i0}-\sigma_{0\alpha} \\
u_{i2}^2-\sigma_{\alpha\alpha}-\sigma_{22} \\
u_{i2}u_{i3}-\sigma_{\alpha\alpha} \\
u_{i2}y_{i0}-\sigma_{0\alpha} \\
u_{i3}^2-\sigma_{\alpha\alpha}-\sigma_{33} \\
u_{i3}y_{i0}-\sigma_{0\alpha} \\
y_{i0}^2 - \sigma_{00}
\end{bmatrix}
$$

Note that the last T+3=6 moment conditions on the left hand side of (A.7) constitute $p_i(\delta)-\theta$. The first T(T-1)/2+(T-2)=4 moment conditions on the left hand side of (A.7) constitute $h_i(\delta)$, and are linear combinations of elements of $m_i(\gamma)$. For terms of the form $y_{is}\Delta u_{it}$, $s\geq 1$, the linear transformation depends on δ. However, as in the proof above of the equivalence of OMD estimators after a linear transformation involving δ, this does not matter asymptotically.

APPENDIX 2

We wish to show that the variance of the GMM/OMD estimator is equal to the semiparametric bound. Let ξ be a single observation of i.i.d. data. Define $x = v(\xi\xi')$ with $E(x) = v(\Omega) = \mu(\gamma)$ where the dimensions of μ and γ are p and q (p>q), repectively. Let $f(\xi|\rho)$ be a probability density function where ρ is an abstract (infinite-dimensional) parameter vector to which a parametric submodel corresponds. Then, $\mu(\gamma) = \int x f(\xi|\rho)d\xi$. Assuming that γ is differentiable function of ρ, say $\gamma(\rho)$, and defining $G=\partial\mu/\partial\gamma'$ as in the main text, we obtain:

(A.8) $\quad G\cdot\partial\gamma/\partial\rho' = \int xS_\rho' f(\xi|\rho)d\xi = E(xS_\rho')$,

where $S_\rho = \partial\ln(f)/\partial\rho$. Therefore, we have the following equality:

(A.9) $\quad \partial\gamma/\partial\rho' = E(dS_\rho')$,

where $d = (G'AG)^{-1}G'Ax$ and A is any nonsingular matrix.

Let $\hat{\delta}$ be an asymptotically linear estimator with influence function d. Then, by Theorem 2.2 of Newey (1990, p. 103), $\hat{\delta}$ is regular. Let $\{(S_\rho)_j\}$ be any sequence of all the possible scores for parametric submodels; and let s be a q × 1 vector such that there exists a sequence of matrices with q rows, $\{A_j\}$, with $\lim_{j\to\infty}E[\|s-A_j(S_\rho)_j\|^2] = 0$. We can define the tangent set $\mathcal{G}$ by:

(A.10) $\quad \mathcal{G} = \{s|\ E(s) = 0 \text{ and } \exists \text{ a matrix B with q rows} \ni E(xs') = GB\}$

This conjecture can be verified as follows. Consider any s^t which is a linear combination of the <u>true</u> score: $s^t = AS_\rho$. Then, it is obvious that $E(s^t) = 0$. Also, (A.8) implies that $E(xs^{t\prime}) = E(xS_\rho'A) = GB$. Therefore, any $s\in\mathcal{G}$ satisfies the restrictions on the scores implied by the semiparametric model.

Now, define $w = [G'(J\psi J)^{-1}G]^{-1}G'(J\psi J)^{-1}(x-\mu)$. Then, we can easily show that $w \in \mathcal{G}$. Note that

(A.11) $$d - w = \{(G'AG)^{-1}G'A - [G'(J\psi J')^{-1}G]^{-1}G'(J\psi J')^{-1}\}x + [G'(J\psi J')^{-1}G]^{-1}G'(J\psi J')^{-1}\mu.$$

Then, the characteristics of $\mathcal{G}$ imply that $E[(d-w)s'] = 0$ and therefore $E[(d-w)'s]$ for any $s\in\mathcal{G}$. By Theorem 3.1 of Newey (1990,

p. 106), the semiparametric bound is E(ww’), which is simply the variance of the OMD/GMM estimator. This completes the proof.