ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 30 — 1993

A Bilinear Model for Heteroskedastic Panel Data

John G. CRAGG *

ABSTRACT. – A model is developed in which an unobserved variable, the same for all individuals but varying over time, interacts with observed variables and unknown coefficients for each individual. Different individual residual variances are allowed. Statistical inference, including estimation of the unobserved variable, presents some difficulties. Procedures to overcome these problems are suggested and analyzed for estimation of the parameters and testing some hypotheses. The procedures are applied to an illustrative example for the connection between the earnings and assets of corporations with results suggesting that the model may be appropriate and that rates of return are related to the size of assets of companies.

Un modèle bilinéaire pour données longitudinales hétéroscédastiques

RÉSUMÉ. – Un modèle est élaboré en incluant une variable inobservée, identique pour tous les individus mais variant d'une période à l'autre, qui interagit avec les variables observées et les coefficients inconnus de chaque individu. On admet la possibilité que la variance des résidus soit différente d'un individu à l'autre. L'inférence statistique, comprenant l'estimation de la variable inobservée, présente quelques difficultés. Des procédures permettant de surmonter les difficultés posées par l'estimation des paramètres et certains tests d'hypothèses sont suggérées et analysées. Ces procédures sont illustrées dans le cadre d'un exemple empirique. Celui-ci étudie le lien entre les bénéfices et les actifs des compagnies commerciales. Les résultats suggèrent que le modèle est approprié et que les taux de rendement sont reliés au volume des actifs des entreprises.

* John G. CRAGG: Department of Economics, University of British Columbia. Research supported in part by a Leave Fellowship and Research Grants 410-87-0778 and 410-90-0627 provided by SSHRCC. Generous provision of facilities by University College London and INSEE are gratefully acknowledged. Comments by participants at a seminar at the LSE and sessions of the Canadian Econometrics Study Group and the Econometric Society Winter Meetings, particularly by Chris Nichol and James MacKinnon, and by Stephen Donald strengthened the paper. I have benefitted greatly from the comments of two anonymous referees and the editor.

1 Introduction

This paper is concerned with a class of models which are bilinear in the critical unknowns. Such models arise in analyzing cross-section/time-series data of the sort encountered when records of companies are available for several periods. The model is applied to an example using this sort of data. It may be applied to other types of panel data as well.

The heuristic background has two elements. First, there are major differences in size among individuals that should affect various aspects of the data, including the variances of residuals. These differences are not known to be measured by some observable variable and there is therefore no obvious method of deflation to remove the heteroskedasticity. It is presumed that a logarithmic (or other prespecified) transformation will not resolve these problems adequately. Second, general economic (or other) conditions that vary over time impinge on the performances of the individual units to differing extents. Differences in these effects are related to differences in the size of the individual units and to the magnitude of relationships among variables.

The model differs in two ways from the more standard panel data models well summarized by JUDGE *et al.* [1980], CHAMBERLAIN [1984], or HSIAO [1985]. The first is the manner in which the time-series and cross-section effects are introduced. These parameters (which in some applications may actually represent substantive, but unobserved, variables) enter bilinearly, multiplied by observable variables. By contrast, the usual approach involves additive coefficients and effects. The problem with the standard specification which this overcomes is that time-effects of the same magnitude for all entities may not be appropriate when the sizes of the entities are radically different as occurs in data on corporations. The second difference from the standard model is specification of different variances for different individuals. This heteroskedasticity combines with the unknown time-parameters to produce many of the statistical problems encountered. It means that maximum-likelihood estimation presents difficulties. A least-squares estimator will be proposed instead.

The model is applied to a simple equation relating corporation earnings to assets which provides a useful example. Most applications are likely to be more complicated, but the results obtained are hopeful for the usefulness of the approach.

2 The Model

Consider the model

$$(1)\qquad y_{jt}=\phi_t x'_{jt}\beta_j+x'_{jt}\gamma_j+\varepsilon_{jt};\qquad t=1,\ \ldots,\ \mathrm{T};\qquad j=1,\ \ldots,\ \mathrm{J}.$$

Here the y_{jt} are dependent variables observed for J individual units in T periods of time. The x_{jt} are K × 1 vectors of observed variables which may include a unit for the constant term. The scalars ϕ_t as well as the vectors β_j and γ_j are unobserved sets of parameters. In practice, the ϕ_t may be regarded as some relevant unobserved variable. When necessary, we distinguish the values of the parameters actually generating the data by ϕ_t^0, β_t^0, γ_t^0, etc. The ε_{jt} are random disturbances assumed to be independent across j with

$$(2)\qquad \mathrm{E}(\varepsilon_{jt})=0,$$

and

$$(3)\qquad \begin{cases} \mathrm{E}(\varepsilon_{jt}^2)=\sigma_j^2, \\ \mathrm{E}(\varepsilon_{jt}\varepsilon_{ks})=0, \quad j\neq k, \quad \text{or} \quad t\neq s. \end{cases}$$

For some purposes, it will also be necessary to assume that the ε_{jt} are normally distributed.

The problem is to estimate the unknown parameters β_j, γ_j and especially ϕ_t using data on the observable variables x_{jt} and y_{jt} and to test hypotheses about them and about the specification of the model. The ϕ_t, β_j and γ_j are not identified without imposing some restrictions on them since the scale of ϕ can be altered with the reciprocal change being made in all β_j. Similarly, a constant can be added to ϕ provided that the product of that constant and β_j is subtracted from γ_j. We shall identify the parameters by imposing the constraints:

$$(4)\qquad \begin{aligned} \sum_{t=1}^{\mathrm{T}} \phi_t &= 0, \\ \sum_{t=1}^{\mathrm{T}-1} \phi_t^2 &= \mathrm{T}-1, \end{aligned}$$

when both β_j and γ_j are present, and by imposing the single constraint

$$(5)\qquad \sum_{t=1}^{\mathrm{T}} \phi_t^2=\mathrm{T}$$

when $\gamma_j=0$. Normalization (5) also applies when only some elements of γ_j are specified to be zero, since then the only unidentified aspect of ϕ is its scale. There is still ambiguity about the sign of the ϕ and β_j vectors which may be resolved by specifying that some particular $\phi_i > 0$. The easiest such normalization is to interpret (4) as specifying that $\left(\phi_T = -\sum_{t=1}^{T-1} \phi_t\right.$ and $\phi_{T-1} = \left[(T-1) - \sum_{t=1}^{T-2} \phi_t^2\right]^{1/2}$, while (5) specifies that $\phi_t = \left[T - \sum_{t=1}^{T-1} \phi_t^2\right]^{1/2}$

These specifications, which are adopted for computational convenience, mean that the ϕ's are not comparable under the two situations. They may be made so by using instead of (4) the alternative identifying restrictions $\sum_{t=1}^{T} \phi_t^2 = T$ and, for some k, $\sum_{j=1}^{J} \gamma_{jk} = 0$.

The restrictions (4) make sense only when the time period is given. In cases where different periods are considered or for analyses concerning the effect of changing the size of T, another normalization would be appropriate, possibly interpreting conditions (4) as applying for some fixed subperiod. The advantage of these restrictions is that they are easy to impose in computation since they readily allow treating ϕ_{T-1} and ϕ_T as functions of the other ϕ_t. Indeed, bearing in mind that some such arbitrary normalization is necessary, we can define the problem in terms of $(T-2)$ parameters ψ_t from which we can define the ϕ_t parameters as

$$\phi_t = \psi_t, \qquad t = 1, \ldots, T-2;$$
$$\phi_{T-1} = \left[(T-1) - \sum_{t=1}^{T-2} \psi_t^2\right]^{1/2}; \tag{6}$$
$$\phi_T = -\sum_{t=1}^{T-1} \phi_t.$$

When all $\gamma_j = 0$, this becomes

$$\phi_t = \psi_t, \qquad t = 1, \ldots, T-1;$$
$$\phi_T = \left[T - \sum_{t=1}^{T-1} \psi_t^2\right]^{1/2}. \tag{7}$$

These parametrizations not only ease computation, but also aid analysis. The ψ parameters are identified provided that at least T of the $\beta_j \neq 0$. We shall presume that most $\beta_j \neq 0$ and that the number of such non-zero vectors increases *pari passu* with J. We shall be making assumptions about the x_{jt} to ensure that they do not produce further identification problems. We shall also presume that T is comparatively small while J is fairly large, especially in the sense that the interesting asymptotic results are those obtained as $J \to \infty$ holding T fixed.

2.1. Estimation of ϕ

Maximum-likelihood estimation of the parameters of (1) assuming that the ε_{jt} are normally and independently distributed appears, deceptively, to be simple. Let

$$
(8)\qquad
\begin{aligned}
z_{jt} &= \{ \phi_t' x_{jt}' \;\; x_{jt}' \}, \\
Z_j' &= [z_{j1} \ldots z_{jT}], \\
\delta_j' &= \{ \beta_j' \, \gamma_j' \}, \qquad \text{and} \\
y_j' &= \{ y_{j1} \ldots y_{jT} \}.
\end{aligned}
$$

The logarithm of the likelihood function is

$$
(9)\qquad L = \kappa - .5\,T \sum_{j=1}^{J} \ln(\sigma_j^2) - .5 \sum_{j=1}^{J} (y_j - Z_j \delta_j)'(y_j - Z_j \delta_j)/\sigma_j^2 .
$$

The solution to the first-order conditions for a maximum can be expressed implicitly as

$$
(10)\qquad
\begin{aligned}
\hat{\delta}_j^{M} &= (\hat{Z}_j^{M\prime} \hat{Z}_j^{M})^{-1} \hat{Z}_j^{M\prime} y_j; \\
\hat{\sigma}_j^{2\,M} &= (y_j - \hat{Z}_j \hat{\delta}_j^{M})' (y_j - \hat{Z}_j^{M} \hat{\delta}_j^{M})/T; \\
\hat{\phi}_t^{M} &= \left(\sum_{j=1}^{J} (y_{jt} - x_{jt}' \hat{\gamma}_j^{M}) x_{jt}' \hat{\beta}_j^{M} / \hat{\sigma}_j^{2\,M} \right) \Big/ \left(\sum_{j=1}^{J} \hat{\beta}_j^{M\prime} x_{jt} x_{jt}' \hat{\beta}_j^{M} / \hat{\sigma}_j^{2\,M} \right).
\end{aligned}
$$

Here $\hat{Z}_j^M$ are the Z_j matrices obtained when $\phi_t = \hat{\phi}_t^M$ in (8). Conditions (10) need to be supplemented by the identifying restrictions, equations (4), since otherwise equations (10) do not have a unique solution even within a neighbourhood of a local maximum. [1]

Equations (10) appear to define the maximum-likelihood estimates for the δ_j and σ_j^2 parameters as the OLS ones obtained when $\phi_t = \hat{\phi}_t^M$. Similarly, the $\hat{\phi}_t^M$ are the GLS estimates that would arise if $\delta_j = \hat{\delta}_j^M$ and $\sigma_j^2 = \hat{\sigma}_j^{2\,M}$. However, while values jointly satisfying (10) would provide a stationary point of the likelihood function (9), they do not provide the global maximum.

The problem of maximizing (9) can instead be solved by finding for some individual unit j a vector, δ_j^m, and values ϕ_t^m, $t = 1, \ldots.\, T$, such that $y_{jt} = z_{jt}^{m\prime} \delta_j^m$. Such values (indeed an infinite number of them when $K > 1$) exist for each j. For any such values, take the values of the other parameters except σ_j^2 as those given by (10). Then as $\sigma_j^2 \to 0$, $L \to \infty$. Thus, maximizing (9) does not provide a sensible criterion, since the usual algorithms provide no guarantee that they will not try to find one of these absurd maxima rather than an interior one. This is quite apart from the

1. The minimization of (9) is subject to (4). The constraints, being identifying ones, are not binding so their Lagrange multipliers are zero at the constrained maximum. The first-order conditions yield equations (10) and (4), the latter expressed in ϕ^M.

usual problem of choosing among possible different local maxima that may exist. [2] In consequence, a more practical estimator is needed.

The obvious alternative to maximum-likelihood is least-squares; that is, to choose estimates $\hat{\delta}_j$ and $\hat{\phi}_t$ to minimize

$$S=\sum_{j=1}^{J}(y_j-Z_j\delta_j)'(y_j-Z_j\delta_j)/J,$$

again subject to (4). The first-order conditions now produce

$$\hat{\delta}_j=(\hat{Z}_j'\hat{Z}_j)^{-1}\hat{Z}_j'y_j, \tag{11}$$

while

$$\hat{\phi}_t=\sum_{j=1}^{J}(y_{jt}-x_{jt}'\hat{\gamma})x_{jt}'\hat{\beta}_j\Big/\sum_{j=1}^{J}\hat{\beta}_j'x_{jt}x_{jt}'\hat{\beta}_j. \tag{12}$$

Note that the difference from (10) arises because $\hat{\phi}_t$ is not now a weighted least-squares estimator.

For ease of calculation and analysis we can use the definition of $\hat{\delta}_j$ in (11) to concentrate the average of the objective function. Let

$$\delta_j^*=(Z_j'Z_j)^{-1}Z_j'y_j,$$

the value of δ_j that minimizes S given the ϕ vector contained in Z. The objective function to be minimized with respect to ϕ, (subject to (4)) is

$$\begin{aligned}S^*&=\sum_{j=1}^{J}(y_j-Z_j\delta_j^*)'(y_j-Z_j\delta_j^*)/J\\&=\sum_{j=1}^{J}y_j'[I-Z_j(Z_j'Z_j)^{-1}Z_j']y_j/J.\end{aligned} \tag{13}$$

This minimization can be done using any of a number standard algorithms. [3]

2. Failure of the global maximum of the likelihood function to provide sensible estimates, which does not disappear as $J\to\infty$, does not contradict standard results for the consistency and efficiency of maximum likelihood since the number of parameters increases *pari passu* with J, which is (cf. NEYMAN and SCOTT 1948) a classic circumstance violating the usual conditions for the optimality of maximum-likelihood estimation. The problem is that the maximizing vector is not included in the parameter space while the information used to estimate σ_j^2 does not increase as J does, making the usual propositions irrelevant. *Cf.* WALD [1949].

3. The calculations reported later began by computing the β_j presuming that $\gamma_j=0$ and the ϕ_t are constant, then calculating ϕ_t, still presuming that $\gamma_j=0$, and then calculating ϕ_t and δ_j recursively using (11) and (12) (imposing the identifying restrictions at each iteration). This procedure will converge to a local minimum, but after a few iterations the calculations switched to the Newton-Raphson algorithm which in trial problems produced faster convergence. In using this algorithm, it is easiest to substitute constraints (4) into the problem so that the minimization actually involves only the ψ parameters.

Asymptotic analysis of the concentrated sum-of-squares function (13) as $J \rightarrow \infty$ is fairly straightforward when we treat the X_j as random variables [4] independent of ε_j and of themselves across j. This is developed in the appendix using standard assumptions about the existence of moments. The three main conclusions are:

a) For fixed ψ, S^* converges in probability to a non-stochastic function of ψ which reaches its minimum at ψ^0.

b) $\hat{\psi}$, the estimate of ψ minimizing S^*, is a consistent, asymptotically normally distributed estimator of ψ^0.

c) The covariance matrix of the asymptotic distribution is consistently estimated using the Hessian and the outer product of the gradient of the individual terms of S^* in direct analogy to the heteroskedasticity-consistent covariance matrices proposed by WHITE [1980]. [5]

The critical part of the specification producing these results is lack of heteroskedasticity and autocorrelation of the ε_{jt} for the separate j. In this regard, this assumption is more critical than it is in usual regression or panel data models where it affects the distribution of estimates, but not the values they are implicitly estimating. [6]

2.2. Testing for a variable ϕ vector

These results provide a basis for estimating the ψ, and so the ϕ, vectors and making inferences about their components. Unfortunately, they do not deal directly with the major question inviting inference, namely whether a ϕ vector with different elements is appropriate. The difficulty of testing this in the full model is that under the null hypothesis the ψ vector is not identified and so the asymptotic distribution does not apply. We suggest two different approaches to this problem.

The first approach starts by presuming initially the hypothesis that all $\gamma_j = 0$. This is a legitimate parameterization under the null hypothesis and renders ϕ identifiable up to the scalar normalization provided by (5). We can now use the $(T-1)$-element vector ψ defined in (7) which under H^0 is the identity vector ι. As noted in the appendix, $\sqrt{J}\, \partial S^* / \partial \psi^0 \overset{A}{\sim} N(0, V)$ where a consistent estimate $\hat{V}$ of V can be obtained from the outer product

4. The X_t (or components of them) could equally well be treated as fixed provided that all needed averages of cross-products converge appropriately as $J \rightarrow \infty$. We treat the ϕ_t, σ_j^2 and the δ_j as fixed, finite parameters to be estimated.

5. As in WHITE [1980], these results are obtained even though the number of parameters increases with J. The "trick" is that consistent estimates of δ_j are not needed in order to obtain consistent and asymptotically normally-distributed estimates of the others.

6. The crucial thing is that $\operatorname{tr} \sum_{j=1}^{J} E[(Z_j' Z_j)^{-1} Z_j' \varepsilon_j \varepsilon_j' Z_j]$ does not depend on Z_j. This in general rules out heteroskedasticity over time with a common pattern across companies, autocorrelation of the ε_{jt} or lagged dependent variables among the X_j.

of the gradient evaluated at $\psi=\iota$. Thus we can base the test on the fact that under $H^0: \psi^0=\iota$,

$$J(\partial S^*/\partial\psi|_{\psi=\iota})' \hat{V}^{-1}(\partial S^*/\partial\psi|_{\psi=\iota}) \overset{A}{\sim} \chi^2(T-1).$$

Alternatively, but in the same vein presuming that $\gamma_j=0$, we can test using the asymptotic distribution of $\hat{\psi}$ and the consistent estimate of its covariance matrix, $\hat{Q}$, since then

$$J(\hat{\psi}-\iota)'\hat{Q}^{-1}(\hat{\psi}-\iota) \overset{A}{\sim} \chi^2(T-1).$$

Asymptotically the two tests are equivalent, but this will not be the case in finite samples.

The second approach does not produce an equivalent test and uses instead an aspect of the specification which is available whatever the values of the parameters. Let

$$\nu_j=(X_j'X_j)^{-1}\sum_{t=1}^{T}\phi_t x_{jt}x_{jt}'\beta_j.$$

Note from (1) and (2) that $E(x_{jt}y_{jt}|X_j)=\phi_t x_{jt}x_{jt}'\beta_j+x_{jt}x_{jt}'\gamma_j$ while

$$E(x_{jt}x_{jt}'(X_j'X_j)^{-1}X_j'y_j|X_j)=x_{jt}x_{jt}'\nu_j+x_{jt}x_{jt}'\gamma_j.$$

Define the K-vector

$$h_{jt}=x_{jt}y_{jt}-x_{jt}x_{jt}'(X_j'X_j)^{-1}X_j'y_j, \tag{14}$$

which is the cross product of the t-th residual from the regression of y_j on X_j with x_{jt} and includes that residual as an element if x_{jt} contains a constant. Then under the null hypothesis that ϕ is a constant vector, $E(h_{jt}|X_j)=0$. Under the alternative that ϕ varies,

$$E(h_{jt}|X_j)=\phi_t x_{jt}x_{jt}'\beta_j-x_{jt}x_{jt}'\nu_j. \tag{15}$$

Note that $\sum_{t=1}^{T} h_{jt}=0$, reflecting the orthogonality produced by least squares so that the joint distribution of only $(T-1)$ of the h_{jt} is non-degenerate. Define the $(KT-K)$-vector $h_j'=\{h_{j1}' \ldots h_{j,T-1}'\}$ and consider $\bar{h}=\sum_{j=1}^{J} h_j/J$ and $H=\sum_{j=1}^{J} h_j h_j'/J$. Then under $H^0: E(h_{jt})=0$. With the supplementary moment assumptions indicated in the appendix, $J\bar{h}'H^{-1}\bar{h} \overset{A}{\sim} \chi^2(KT-K)$. For this test to have power requires that

$$\lim \sum_{j=1}^{J} E\{x_{tj}x_{tj}'[\phi_t I-(X_j'X_j)^{-1}\sum_{s=1}^{T}\phi_s x_{js}x_{js}']\beta_j\}/J \neq 0, \qquad t=1,\ldots,T-1.$$

Given that the $\beta_j \neq 0$, it is hard to think of a substantive set of circumstances that would make this a problem other than that both $\sum_{j=1}^{J} \beta_j = 0$ (for which there would usually be no reason) and also that any differences in the covariance matrices of the x_{jt} were unrelated to the β_j.

2.3. Inference about δ

Inference about δ_j is at first sight more problematical because its estimates depend on $\hat{\phi}$. However, this does not create serious problems asymptotically. Let $\hat{\delta}_j^0$ be $\hat{\delta}_j$ calculated using ϕ^0 and δ_j^0 be the population value of δ_j. Since $p \lim \hat{\phi} = \phi^0$, $p \lim (\hat{\delta}_j - \hat{\delta}_j^0) = 0$ so that $(\hat{\delta}_j - \delta_j^0) \xrightarrow{d} (\hat{\delta}_j^0 - \delta_j^0)$. For large J therefore one is justified in testing hypotheses about δ_j (for particular values of j) by treating the estimates as ordinary regression estimates. For example, consider the standard $H^0 : A_j \delta_j = a_j$, where A_j and a_j are a known $N \times M$ matrix of rank N and a known N-vector respectively. If the ε_{jt} are assumed to be normally distributed, one can use the usual F-test. The test statistic is

$$(2.16) \qquad \begin{aligned} f_j &= [(A \hat{\delta}_j - a_j)' (A_j (\hat{Z}_j' \hat{Z}_j)^{-1} A_j')^{-1} (A_j \hat{\delta}_j - a_j)/N] / \\ &\quad [y_j' [I - \hat{Z}_j (\hat{Z}_j' \hat{Z}_j)^{-1} \hat{Z}_j'] y_j / (T - M)] \\ &= [(s_j^R - s_j^U)/s_j^U] [(T - M)/N], \end{aligned}$$

where s_j^R and s_j^U are the residual sums of squares using the restricted and unrestricted models respectively and M and N are the number of coefficients in the unrestricted version and the number of restrictions respectively. Then under H^0, $\underset{J \to \infty}{p \lim} (f_j - f_j^0) = 0$ where f_j^0 is the value of f_j in formula (16) using Z_j^0 and $\hat{\delta}_j^0$ in place of $\hat{Z}_j$ and $\hat{\delta}_j$. In consequence, $f_j \xrightarrow{d} F(N, T - M)$. Under the alternative hypothesis, the limiting distribution of f_j as $J \to \infty$ is the non-central F, with non-centrality parameter

$$(17) \qquad (A_j \delta_j^0 - a_j)' (A_j (Z_j^{0\prime} Z_j^0)^{-1} A_j')^{-1} (A_j \delta_j^0 = a_j)/\sigma_j^2.$$

Hypotheses about the δ_j are more likely to involve values for all the different j rather than for a particular one. The same approach, ignoring the dependence on $\hat{\phi}$ and the resulting correlation across j, is justified asymptotically since the dependence on $\hat{\phi}$ is of lower order in probability than the variability of the basic test statistic. Specifically, suppose we want to test

$$(18) \qquad H^0 : A\delta_j = a, \qquad \forall j = 1, \ldots, J.$$

against

$$H^a : A \delta_j \neq a, \text{ for any number of } j, j = 1, \ldots, J.$$

Consider using as a test statistic $\bar{f} = \sum_{j=1}^{J} f_j/J$ with the f_j calculated with the common values of A and a for A_j and a_j. Let $\bar{f}^0 = \sum_{j=1}^{J} f_j^0/J$. Then, as shown in the appendix, with appropriate assumptions under H^0

$$p \lim \bar{f} = p \lim \bar{f}^0 = q,$$

$$\sqrt{J}(\bar{f} - q) \xrightarrow{d} N(0, Q),$$

where $q = E(f_j) = (T-M)/(T-M-2)$, the mean of $F(N, T-M)$ and $Q = 2(T-M)^2(T-M+N-2)/(N(T-M-2)^2(T-M-4))$, its variance. Thus one may test the hypothesis of a common restriction on the δ_j parameters with a one-sided normal-test of the hypothesis that the f_j have mean q.

This test involves the assumption of normally distributed residuals leading to an F distribution. Since $T \not\to \infty$ and may well be small, this small-sample assumption is central to the test. It is not needed, however, in order to test the hypothesis (18). Specifically, consider

$$\text{(19)} \qquad \begin{aligned} b_j &= (\hat{A}\,\hat{\delta}_j - a)'(A(\hat{Z}_j'\hat{Z}_j)^{-1}A')^{-1}(A\,\hat{\delta}_j - a) - N s_j^U/(T-M) \\ &= s_j^R - s_j^U - N s_j^U/(T-M). \end{aligned}$$

Let b_j^0 be the values of b_j calculated using ϕ^0 and $\hat{\delta}_j^0$ rather than $\hat{\phi}$ and $\hat{\delta}_j$ in (19). Under H^0, $E b_j^0 = 0$ while under the alternative hypothesis,

$$E(b_j^0 \mid X_j) = (A\,\delta_j^0 - a)'(A(Z_j^{0\prime}Z_j^0)^{-1}A')^{-1}(A\,\delta_j^0 - a) > 0.$$

Define $\bar{b} = \sum_{j=1}^{J} b_j/J$ and $\bar{b}^0 = \sum_{j=1}^{J} b_j^0/J$. As shown in the appendix, using suitable assumptions about x_{jt} and ε_{jt},

$$\text{(20)} \qquad \begin{aligned} & p \lim \bar{b} = p \lim \bar{b}^0 = 0; \\ & \sqrt{J}\,\bar{b} \xrightarrow{d} N(0, r); \\ & p \lim \sum_{j=1}^{J} b_j^2/J = r. \end{aligned}$$

The three properties of b_j in (20) provide asymptotic justification for testing H^0 by a one-sided normal-test that the average is zero which does not require normality of the ε_{jt}.

The test using $\bar{b}$ will have power asymptotically only if under the alternative

$$\text{(21)} \qquad \sum_{j=1}^{J} (A\,\delta_j^0 - a)'(A(Z^{0\prime}Z^0)^{-1}A')^{-1}(A\,\delta_j^0 - a)/J > o_p(J^{-1/2})$$

since $\bar{b}$ converges in probability to the L.H.S. of (21) under H^a. A corresponding condition involving the average of the non-centrality parameters in (17) is needed for the test using $\bar{f}$.

3 An Illustration Using Corporate Earnings

We illustrate the model and its usefulness by fitting it to U.S. Corporate earnings over the period 1958-1983, using data from the COMPUSTAT tape for S & P-400 companies with fiscal year ending December 31. There was a total of 271 such companies for which enough data were available so that in this application T=26 and J=271. The model is illustrative, and details of specification or its adequacy or completeness are not here of primary concern.

The dependent variable is corporate earnings in real terms (obtained by division by the GNP implicit price deflator). It is hypothesized to depend on the capital of the corporation represented by its real assets, and on "business conditions". This is the underlying variable that will be estimated as the ϕ_t's. The dependence of earnings on real assets, A_{jt}, is specified to be quadratic. Thus letting P_{jt} be the earnings figure for the j-th company in year t, the model becomes

(22) $$P_{jt} = \phi_t(\beta_{j0} + A_{jt}\beta_{j1} + A_{jt}^2\beta_{j2}) + (\gamma_{j0} + A_{jt}\gamma_{j1} + A_{jt}^2\gamma_{j2}) + \varepsilon_{jt}.$$

The least-squares estimates of ϕ_t and their standard errors are found in Table 1. The model apparently performs very strongly in accounting for the earnings data. In particular, the hypothesis that ϕ_t is zero (or is a constant with all γ_j being zero) was rejected beyond the .005 level using any of the tests suggested in section 2.2.

Having found that the account of earnings using the overall, unobserved variables ϕ_t has considerable merit, it is of interest to enquire whether or not the ϕ_t simply represent standard aggregate variables. To this end, we considered as explanatory variables for the ϕ_t the unemployment rate, short-term nominal interest rates as measured by the 90-day treasury bill rate, and the rate of inflation as measured by the annual rate of change of the GNP implicit price index. The extent to which these variables account

TABLE 1

Estimates of ϕ_t

Year	est.	s.e.	Year	est.	s.e.
1958.	−.701	.103	1971.	−.426	.055
1959.	−.692	.098	1972.	−.424	.051
1960.	−.678	.095	1973.	−.178	.097
1961.	−.662	.090	1974.	.098	.355
1962.	−.640	.084	1975.	−.170	.182
1963.	−.618	.079	1976.	−.274	.127
1964.	−.597	.065	1977.	−.363	.120
1965.	−.579	.058	1978.	−.283	.231
1966.	−.548	.059	1979.	.394	.259
1967.	−.507	.053	1980.	1.890	.345
1968.	−.482	.051	1981.	2.669	.341
1969.	−.474	.051	1982.	2.980	.219
1970.	−.434	.057	1983.	1.700	1.296

for ϕ is substantial with R^2 being 0.95. The regression is summarized in the first part of Table 2. The coefficients appear to be highly significant. [7]

One feature of the results in the first part of Table 2 is misleading. Despite the value of the Durbin-Watson statistic, it is the case that introducing a second-order autoregressive process adds significantly to the fit. The resulting equation is summarized in the second part of Table 2. The autoregressive process has complex roots well outside the unit circle. This feature accounts both for the apparent lack of first-order autocorrelation indicated by the Durbin-Watson statistic and for a first-order process adding comparatively little to the account of ϕ. Furthermore, the strength of the second-order process does not arise simply from the lagged values of the dependent variable representing the lagged values of the explanatory variables. Instead, these lagged values were insignificant when they were also included in the regression while the lagged values of ϕ remained significant. [8]

Several findings concerning the specification of equation (22) are worth noting. The major overall questions concern whether such an elaborate

7. Also considered as possible explanatory variables were real and money GNP, the rate of growth of GNP, and the level of prices. These variables added little to those already included, but extensive experimentation with variables and forms was not pursued. Though these results are based on the regression over the period 1960-1983 (used because the identifying constraints mean that only 24 separate ϕ need to be estimated), the results would be qualitatively identical if the regression were run for the whole period or for a different subperiod. While allowance is made for possible heteroskedasticity, no account is taken of the fact that the sampling errors in the estimates of ϕ_t are not independent of each other.
8. These findings hold also for the variations in the specification of explanatory variables mentioned in the previous footnote.

TABLE 2

Regressions for ϕ 1960-1983

Variable	Estimate	O.S.E.	H.C.S.E.
Constant	−2.804	.196	.218
Int. rate	.389	.031	.030
Unemp	.235	.035	.040
Inflation	−.189	.033	.034

R^2 .951; D.W. 2.09

Variable	Estimate	O.S.E.	H.C.S.E.
Constant	−1.870	.387	.403
Int. rate	.246	.050	.050
Unemp	.146	.049	.044
Inflation	−.111	.038	.032
ϕ_{t-1}	.862	.238	.257
ϕ_{t-2}	−.669	.214	.273

R^2 .972
O.S.E.: Ordinary Standard Errors; H.C.S.E.: Standard Errors based on heteroskedasticity-consistent covariance matrix.

quadratic specification is needed. According to the test procedures developed in section 2.3 they are. The hypotheses that all $\gamma=0$, that all $\beta_{j0}=\gamma_{j0}=0$ and that all $\beta_{j2}=\gamma_{j2}=0$ were each rejected well beyond the 0.05 level using either of the tests [9] suggested in section 2.3.

One may be a bit skeptical of these test results, since the procedures have only asymptotic justification. It is therefore of some interest to note that restricting the model does produce substantial changes. For example, r^2 between ϕ calculated when all $\gamma_j=0$ and the unrestricted ϕ is 0.45. By contrast, dropping the constants produced a ϕ vector which had an r^2 of 0.94 with the unconstrained one. Dropping the square terms while retaining the constants also led to estimates with a correlation of about the same magnitude with the unrestricted estimates. However, dropping both the constants and the square terms produced ϕ values with r^2 of only 0.64 with the unrestricted ones.

This is possibly the most interesting of the substantive findings. It might seem reasonable to suppose that the rate of return on a company's assets will depend on the company's basic riskiness and on business conditions. This corresponds to the model restricted to having no constants and no A_{jt}^2 terms. The rate of return should not depend on the size of the company's

9. Using $\bar{J}$ the standardized normal test statistic had values of 11.73, 6.52 and 5.87 respectively for the three hypotheses. Using $\bar{F}$ the corresponding values were 12.65, 5.71 and 5.67. For the tests that $\gamma_j=0$, the value of ϕ calculated under H^0 was used since if H^0 is true ϕ is not identified under H^a.

assets, which would be the inference to be drawn from concluding that either the constant or the square term has significance.

Investigation of the β and γ coefficients reveals some interesting features. The estimates of these coefficients vary substantially across the companies, and a major issue is the extent to which this variation is related to the observed independent variables. This association is summarized in Table 3 where we report the regressions of the coefficients on the average (over the period) values of the explanatory variables A_{jt} and A_{jt}^2, $\bar{A}_j$ and $\bar{A}_j^2$. Adding the averages of higher powers of A_{jt} did not alter the results qualitatively. The regressions are reported in the form of deviations from the mean so that the estimates of the constants are the averages of the β_j and γ_j coefficients.

TABLE 3

Regression of Estimates of δ_j on Average Values of A_{jt} and A_{jt}^2. Estimated Coefficients. (Standard errors in parentheses) [*"t"-Ratios in brackets*]

	β_0	β_1	β_2	γ_0	γ_1	γ_2
Constant	.449	−.290	.191	.335	−.150	.177
	(.467)	(.102)	(.091)	(.398)	(.076)	(.082)
	[0.96]	[−2.84]	[2.09]	[0.84]	[−1.99]	[2.16]
$\bar{A}_j$	−.787	.065	−.035	−.634	.050	−.037
	(.216)	(.047)	(.042)	(.184)	(.035)	(.038)
	[−3.64]	[1.37]	[−0.83]	[0.67]	[1.42]	[−0.98]
$\bar{A}_j^2$	.031	−.002	.001	.024	−.001	.001
	(.005)	(.001)	(.001)	(.004)	(.001)	(.001)
	[6.38]	[−1.48]	[0.63]	[5.99]	[−1.63]	[0.77]
R^2	.243	.009	.004	.209	.011	.006

Two rather different types of results appear in these regressions. For β_{j0} and γ_{j0} there is a fairly strong association both with $\bar{A}_j$ and also with $\bar{A}_j^2$. By contrast, the other coefficients show significant averages, but virtually no association with the independent variables.

4 Conclusion

This paper has developed a bilinear model that can be used in panel data to capture the idea that some overall variable has varying multiplicative effects on individual behavior. Least-squares estimates of the over-all parameters can be calculated for which large-sample theory of the usual sort was developed.

A simple model fitted to corporation data illustrated the suggested procedures. It is encouraging that the procedures led to estimates of the underlying variable that appeared to be very closely related to some aggregate economic variables. This finding suggests also that it would be worthwhile to develop explicit procedures to investigate whether such linear combinations really capture fully the unobserved influences.

Over-all the fact that the model could be fitted successfully with apparently sensible results to a body of data on diverse corporations is highly encouraging. Nevertheless, many questions remain about the sampling properties of the techniques that require investigation. Hopefully the model may also stand as a prototype pointing towards evolution of more adventurous specification and estimation in panel data.

APPENDIX

Asymptotic Properties of $\hat{\phi}$, $\bar{f}$ and $\bar{b}$

Properties of $\hat{\phi}$

For convenience, we shall presume that $z'_{jt} = \{\phi_t x'_{jt}\ x'_{jt}\}$, the needed changes when $\gamma_j = 0$, or when there are particular omissions among the variables being obvious. Assume that $\psi^0 \in \Psi$ and $\delta_j^0 \in \Delta, j = 1, \ldots, J$, where Ψ and Δ are bounded compact sets in $\mathscr{R}^{T-2}$ and $\mathscr{R}^{2K}$ respectively, and that the x_{jt} are random vectors (or constants) independent across J and of all ε_{jt} such that for all $\psi \in \Psi$, $Z'_j Z_j$ is non-singular almost surely, $j = 1, \ldots, J$. Given independence, the main requirement is that S* in (13) converge in probability to a non-stochastic function having the critical properties of the expectation of the individual terms in S*. Define $c_{jts} = z'_{jt}(Z'_j Z_j)^{-1} z_{js}$ and assume, letting x_{kjt} be the k-th element of x_{jt}, that for some Λ, $0 < \Lambda < \infty$, and some $\delta > 0$,

$$
(23)\quad
\begin{gathered}
E(|x_{kjt} x_{mjs} c_{jts}|^{(1+\delta)}) \leqslant \Lambda;\\
E(|x_{kjt} x_{mjt}|^{(1+\delta)}) \leqslant \Lambda;\\
E(|x_{kjt} c_{jts}|^{(1+\delta)}) \leqslant \Lambda;\\
E(|\varepsilon_{jt}|^{(2+\delta)}) \leqslant \Lambda;\\
\forall \psi \in \Psi;\ j = 1, \ldots, J;\ t, s = 1, \ldots, T;\ k, m = 1, \ldots, K.
\end{gathered}
$$

Assume $\lim_{J \to \infty} \sum_{j=1}^{J} \sigma_j^2/J$ and $\lim_{J \to \infty} \left(\sum_{j=1}^{J} E(Z_j^{0\prime}[I - Z_j(Z'_j Z_j)^{-1} Z'_j] Z_j^0)/J \right)$ are well defined (including any needed limits when the x_{jt} contain fixed variates). (Assumptions (23) guarantee the existence of these moments.) Then $p\lim S^*$ exists as a non-stochastic function of ψ which reaches its minimum at $\psi = \psi^0$. Specifically,

$$
\begin{aligned}
(24)\quad p\lim S^* &= p\lim \sum_{j=1}^{J} y'_j [I - Z_j(Z'_j Z_j)^{-1} Z'_j] y_j/J \\
&= p\lim \sum_{j=1}^{J} (\varepsilon_j + Z_j^0 \delta_j^0)' [I - Z_j(Z'_j Z_j)^{-1} Z'_j](\varepsilon_j + Z_j^0 \delta_j^0)/J \\
&= \lim \sum_{j=1}^{J} E(\varepsilon_j + Z_j^0 \delta_j^0)' [I - Z_j(Z'_j Z_j)^{-1} Z'_j](\varepsilon_j + Z_j^0 \delta_j^0)/J \\
&= (T-K) \lim \sum_{j=1}^{J} \sigma_j^2/J \\
&\quad + \lim E\left\{ \sum_{j=1}^{J} (\delta_j^{0\prime} Z_j^{0\prime} [I - Z_j(Z'_j Z_j)^{-1} Z'_j] Z_j^0 \delta_j^0)/J \right\} \\
&\geqslant (T-K) \lim \sum_{j=1}^{J} \sigma_j^2/J.
\end{aligned}
$$

The last line of (24) gives $p\lim S^*$ evaluated at ψ^0. The inequality is strict for $\psi \neq \psi^0$ since the expectation of the second, positive semi-definite form in the penultimate line is positive definite except when $\psi = \psi^0$. The third line of (24) arises from the assumptions on moments by Markov's theorem (cf. LOÈVE [1977], p. 287) and because $E(Z_j^0 \delta_j^0 \varepsilon_j') = 0$. The evaluation

$$E(\varepsilon_j' [I - Z_j (Z_j' Z_j)^{-1} Z_j'] \varepsilon_j) = (T - K) \sigma_j^2$$

used in obtaining the penultimate line of (24) relies on the assumption (3) giving

$$E(\varepsilon_j \varepsilon_j' \mid Z_j) = \sigma_j^2 I. \tag{25}$$

Result (24) together with the assumptions made yields (cf. AMEMIYA, 1985, p. 106) that $p\lim \hat{\psi} = \psi^0$.

The asymptotic normality of $\hat{\psi}$ is equally straightforward if we make assumptions to allow use of Markov's theorem and of the Liapounov CLT. To this end we need assumptions such that in a neighbourhood of ψ^0, $p\lim \partial^2 S^* / \partial \psi \, \partial \psi'$ exists as a positive definite matrix and that $\sqrt{J}\, \partial S^* / \partial \psi^0 \overset{A}{\sim} N(0, V)$ where the notation $\partial x / \partial a^0$ indicates $\partial x / \partial a_{|a=a^0}$. Let

$$s_j^* = y_j' [I - Z_j (Z_j' Z_j)^{-1} Z_j'] y_j, \tag{26}$$

the terms of which S^* is the average. Then

$$\partial s_j^* / \partial \phi_t = -2 y_{jt} (x_{jt}' \;\; 0') (Z_j' Z_j)^{-1} Z_j' y_j + 2 y_j' Z_j (Z_j' Z_j)^{-1} (x_{jt}' \;\; 0')' z_{jt}' (Z_j' Z_j)^{-1} Z_j' y_j. \tag{27}$$

This expression involves the sum of cross products of the ε_{jt} and of the x_{kjt} with themselves and each other multiplied by the K^2 items of which the c_{jts} are composed and with the cross products of these latter elements. We shall assume that finite $(2+\delta)$-th moments exist for these various elements. Similarly we shall assume the existence of the $(1+\delta)$-th moments of the products of the cross products of the ε_{jt} and of the z_{kjt} with the third cross products of the items entering the c_{jts}, which are also involved in $\partial^2 s_j^* / \partial \phi \, \partial \phi'$. These assumptions ensure the existence of $E[(\partial s_j^* / \partial \phi)(\partial s_j^* / \partial \phi)']$ and of $E[\partial^2 s_j^* / \partial \phi \, \partial \phi']$ and also by Markov's theorem that

$$p\lim \sum_{j=1}^{J} \{ (\partial s_j^* / \partial \phi)(\partial s_j^* / \partial \phi)' - E[(\partial s_j^* / \partial \phi)(\partial s_j^* / \partial \phi)'] \} / J = 0,$$

and that

$$p\lim \sum_{j=1}^{J} \{ \partial^2 s_j^* / \partial \phi \, \partial \phi' - E[\partial^2 s_j^* / \partial \phi \, \partial \phi'] \} / J = 0.$$

Assume for notational convenience that the average expectations in these expressions converge to constants as $J \rightarrow \infty$. Note that

$$(28) \qquad \begin{aligned} E(\partial s_j^*/\partial \phi_t^0) &= E[\hat{\varepsilon}_{jt}^0 (x_{jt}' \;\; 0')(Z_j^{0\prime} Z_j^0)^{-1} Z_j^{0\prime} y_j] \\ &= E[\hat{\varepsilon}_{jt}^0 (x_{jt}' \;\; 0')(\hat{\delta}_j^0 - \delta_j^0)] = 0 \end{aligned}$$

where $\hat{\varepsilon}_j^0 = [I - Z_j^0 (Z_j^{0\prime} Z_j^0)^{-1} Z_j^{0\prime}] y_j$, $\hat{\delta}_j^0 = (Z_j^{0\prime} Z_j^0)^{-1} Z_j^{0\prime} y_j$ (the estimate of δ_j that would be obtained if Z_j^0 were used) and we rely on (25) to produce the usual lack of correlation between residuals and estimates in a correctly specified regression.

Let $D = \{d_{ij}\} = \{\partial \phi_i / \partial \psi_j\}$, with D^0 and $\hat{D}$ indicating the matrix evaluated at ψ^0 and $\hat{\psi}$ respectively. Then our assumptions suffice for the Liapounov CLT to apply to $\partial s_j^*/\partial \psi^0$; that is, $\sqrt{J}(\partial s_j^*/\partial \psi^0) \overset{A}{\sim} N(0, V)$ where

$$V = D^{0\prime} \lim_{J \rightarrow \infty} \left\{ \sum_{j=1}^{J} E[(\partial s_j^*/\partial \phi^0)(\partial s_j^*/\partial \phi^0)']/J \right\} D^0,$$

which is consistently estimated by $\hat{V} = \hat{D}' \left\{ \sum_{j=1}^{J} (\partial s_j^*/\partial \hat{\phi})(\partial s_j^*/\partial \hat{\phi})'/J \right\} \hat{D}$.

This is the basis of the first test proposed in section 2.2.

Let

$$U = \lim_{J \rightarrow \infty} E[\partial^2 S^*/\partial \psi^0 \partial \psi^{0\prime}] = \lim_{J \rightarrow \infty} D^{0\prime} \left(\sum_{j=1}^{J} E[\partial^2 s_j^*/\partial \phi^0 \partial \phi^{0\prime}]/J \right) D^0.$$

Then, using the standard Taylor series expansion,

$$\sqrt{J}(\hat{\psi} - \psi^0) \overset{A}{\sim} N(0, U^{-1} V U^{-1}).$$

U is consistently estimated by $\hat{U} = \hat{D}' \left(\sum_{j=1}^{J} (\partial^2 s_j^*/\partial \hat{\phi} \partial \hat{\phi}')/J \right) \hat{D}$.

The distribution of $\sqrt{J}(\hat{\phi} - \phi^0)$ is degenerate, but the asymptotic distribution of any $(T-2)$ or fewer elements is normal with covariance matrix obtained from the corresponding submatrix of $DU^{-1}VU^{-1}D'$, consistently estimated by $\hat{D}\hat{U}^{-1}\hat{V}\hat{U}^{-1}\hat{D}'$.

The assumptions made guarantee finite $(2+\delta)-$th moments for h_{jt} of (14), so $\bar{h}$ is asymptotically normally distributed with covariance matrix given by $\sum_{j=1}^{J} E(h_j h_j')/J$ which is consistently estimated by $\sum_{j=1}^{J} h_j h_j'/J$.

Properties of $\bar{b}$ and $\bar{f}$

Similar analyses apply to the test in section 2.3 using $\bar{b}$, the average of

$$b_j = s_j^R - (T - M + N) s_j^U/(T - M)$$

where $M=2K$ and s_j^R and s_j^U are the residual sums of squares imposing the null hypothesis (18) and imposing only the maintained hypothesis respectively.

Let $s_j^{U_0}=y_j'[I-Z_j^0(Z_j^{0\prime}Z_j^0)^{-1}Z_j^{0\prime}]y_j$ for which $E\,s_j^{U_0}=(T-M)\sigma_j^2$. Correspondingly, define $s_j^{R_0}=(y_j-Z_j^0\tilde{\delta}_j)'(y_j-Z_j^0\tilde{\delta}_j)$ where

$$\tilde{\delta}_j=\hat{\delta}_j^0-(Z_j^{0\prime}Z_j^0)^{-1}A'[A(Z_j^{0\prime}Z_j^0)^{-1}A']^{-1}(A\hat{\delta}_j^0-a)$$

is the restricted least-squares estimates subject to restrictions (18) using ϕ^0. Then

$$E(s_j^{R_0})=(T-M+N)\sigma_j^2+E(A\delta_j^0-a)'(A(Z_j^{0\prime}Z_j^0)^{-1}A')^{-1}(A\delta_j^0-a).$$

Hence under H^0, $E(b_j^0)=0$ while under H^a, for any j for which $(A\delta_j^0-a)\neq 0$, $E(b_j^0)>0$. By assumption, the b_j^0 are independent across J. Assume for convenience that under H^a $\lim\sum_{j=1}^{J}E(b_j^0)/J$ is well defined. Again making appropriate assumptions on moments of the x_{jt} and ε_{jt} and recalling that $p\lim\phi=\phi^0$,

$$(29)\quad p\lim\sum_{j=1}^{J}b_j/J=p\lim\sum_{j=1}^{J}b_j^0/J$$
$$=\lim\sum_{j=1}^{J}E[(A\delta_j^0-a)'(A(Z_j^{0\prime}Z_j^0)^{-1}A')^{-1}(A\delta_j^0-a)]/J.$$

Relying on the assumed non-singularity of $Z_j'Z_j$ in a neighbourhood of ϕ^0, expand $\sqrt{J}(\bar{b})$ in Taylor series about ϕ^0:

$$(30)\qquad \sqrt{J}(\bar{b})=\sqrt{J}(\bar{b}^0)+\partial\bar{b}/\partial\phi_{|\phi=\phi^*}\sqrt{J}(\hat{\phi}-\phi^0)$$

where ϕ^* lies between $\hat{\phi}$ and ϕ^0. Under H^0

$$(31)\qquad p\lim\partial\bar{b}/\partial\phi_{|\phi=\phi^*}=p\lim\partial\bar{b}/\partial\phi^0=p\lim\sum_{j=1}^{J}(\partial b_j^0/\partial\phi^0)/J$$

since $p\lim\hat{\phi}=\phi^0$.

Now $\partial b_j/\partial\phi^0=\partial s_j^R/\partial\phi^0-[(T-M+N)/(T-M)]\,\partial s_j^U/\partial\phi^0$. From (28), $E(\partial s_j^U/\partial\phi^0)=0$. A similar analysis applies to $\partial s_j^R/\partial\phi^0$. Possibly following rearrangement of δ_j and the columns of A and Z_j, which we shall ignore for notational simplicity, we can partition δ' into $\{\delta_{j1}'\ \delta_{j2}'\}$ of $(M-N)$ and N elements respectively, partition A and Z_j conformably, and write (18) as

$$\delta_{j2}=A_2^{-1}A_1\delta_{j1}+A_2^{-1}a$$

Then, defining $G=A_2^{-1}A_1$ and $g=A_2^{-1}a$ and letting $q_j^0=y_j-Z_{j2}^0 g$ and $Q_j^0=Z_{j1}^0+Z_{j2}^0 G$, the restricted estimate at ϕ^0 is $\tilde{\delta}_{j1}^0=(Q_j^{0\prime}Q_j^0)^{-1}Q_j^{0\prime}q_j^0$ and

$s_j^{Ro} = q_j^{0\prime}[I - Q_j^0 (Q_j^{0\prime} Q_j^0)^{-1} Q_j^{0\prime}] q_j^0$. In consequence under H^0 as in (28),

$$\text{(32)} \qquad E(\partial s_j^{Ro}/\partial \phi_t^0) = E[\hat{\varepsilon}_{jt}^{Ro}\{x_{jt}^{\prime}\ 0^{\prime}\} G (Q_j^{0\prime} Q_j^0)^{-1} Q_j^{0\prime} q_j^0]$$
$$= E[\hat{\varepsilon}_{jt}^{Ro}\{x_{jt}^{\prime}\ 0^{\prime}\} G (\tilde{\delta}_{j1}^0 - \delta_{j1}^0)] = 0$$

where $\hat{\varepsilon}_j^{Ro} = [I - Q_j^0 (Q_i^{0\prime} Q_j^0)^{-1} Q_j^{0\prime}] \varepsilon_j$ under H^0 and

$$(\tilde{\delta}_{j1}^0 - \delta_{j1}^0) = (Q_j^{0\prime} Q_j^0)^{-1} Q_j^{0\prime} \varepsilon_j.$$

The assumptions about moments already made in connection with (27) then ensure that

$$p\lim \sum_{j=1}^{J} (\partial b_j/\partial \phi^0)/J = \lim \sum_{j=1}^{J} E(\partial b_j/\partial \phi^0)/J = 0.$$

In consequence from (30), under H^0, $p\lim \sqrt{J}(\bar{b}) = \sqrt{J}(\bar{b}^0)$, where $\bar{b}^0$ is the sum of independent random variables having bounded $(2+\delta)$-th moments. Therefore,

$$\text{(33)} \qquad \sqrt{J}(\bar{b}^0) \overset{A}{\sim} N\left(0, \lim \sum_{j=1}^{J} E(b_j^{02})/J\right).$$

Under H^0 the variance in (33) is consistently estimated by $\sum_{j=1}^{J} b_j^2/J$ so

$$\sqrt{J}\,\bar{b} \Big/ \left(\sum_{j=1}^{J} b_j^2/J\right)^{1/2} \overset{A}{\sim} N(0, 1).$$

Under H^a, assuming that all $E|b_j^{02}|^{1+\delta} \leqslant \Lambda$, so that $p\lim \sum_{j=1}^{J} b_j^2/J$ exists and provided that

$$\sum_{j=1}^{J} E(A\delta_j^0 - a)^{\prime}(A(Z_j^{0\prime} Z_j^0)^{-1} A^{\prime})^{-1}(A\delta_j^0 - a)/J > O(J^{-1/2})$$

it follows from (29) that $\sqrt{J}(\bar{b}) \Big/ \left(\sum_{j=1}^{J} b_j^2/J\right)^{1/2} > O_p(1)$ so the test is consistent.

Dealing with $\bar{f}$ proceeds in similar fashion, aided by the assumption that the ε_{jt} are normally distributed. Assumptions about the existence of moments of functions of the x_{jt} and ε_{jt} analogous to those in (23) are needed to guarantee that $E|f_j^0|^{2+\delta} \leqslant \Lambda$ and that $E|\partial f_j/\partial \phi^0|^{1+\delta} \leqslant \Lambda$. Then by Markov's theorem since under $H^0 f_j^0 \sim F(N, T-M)$, and presuming that $(T-M) > 4$, $p\lim \bar{f} = E(f_j) = q = (T-M)/(T-M-2)$ and

$$\sqrt{J}(\bar{f} - q) = \sqrt{J}(\bar{f}_j^0 - q) + \left(p\lim \sum_{j=1}^{J} (\partial f_j/\partial \phi^0)/J\right)\sqrt{J}(\hat{\phi} - \phi^0) + o_p(1).$$

Now

$$(34)\quad \partial f_j/\partial\phi^0 = [(T-M)/N]\,[(\partial s_j^R/\partial\phi^0))/s_j^{U_0} - (\partial s_j^U/\partial\phi^0)/s_j^{U_0} - (s_j^{R_0} - s_j^{U_0})(\partial s_j^{U_0}/\partial\phi^0)/(s_j^{U_0})^2].$$

The first term in (34) contains the derivative exhibited in (32) and so involves the product of $(\tilde{\delta}_{j1}^0 - \delta_{j1}^0)$ with an expression involving the x_{jt}, $\hat{\varepsilon}_j^{R_0}$ and $\hat{\varepsilon}_j^0$. Since

$$(Q_j^{0\prime} Q_j^0)^{-1} Q_j^{0\prime} [I - Z_j^0 (Z_j^{0\prime} Z_j^0)^{-1} Z_j^{0\prime}] = 0,$$

$\tilde{\delta}_{j1}^0$ is independent of $\hat{\varepsilon}_j^0$ as well as of $\hat{\varepsilon}_j^{R_0}$ and the x_{jt}, the first term in (34) has expectation 0 under H^0. Similarly, from (28) the second term involves the product of $(\hat{\delta}_j^0 - \delta_j^0)$ with items from which it is independent and so has expectation 0. Finally, the third term of (34) is under H^0 $(A(\hat{\delta}_j^0 - \delta_j^0))'\ (A(Z_j^{0\prime} Z_j^0)^{-1} A')^{-1} (A(\hat{\delta}_j^0 - \delta_j^0))(\partial s_j^U/\partial\phi^0)/(s_j^{U_0})^2$ and so involves the products of the third cross-products of $(\hat{\delta}_j^0 - \delta_j^0)$ with terms from which they are independent and so also has expectation 0. Therefore under H^0,

$$\sqrt{J}(\bar{f} - q) \overset{A}{\sim} N(0,\ E(f_j^0 - q)^2).$$

Because $f_j^0 \sim F(N,\ T-M)$,

$$E(f_j^0 - q)^2 = 2(T-M)^2 (T-M+N-2)/(N(T-M-2)^2 (T-M-4)).$$

• References

AMEMIYA, T., *Advanced Econometrics*, Cambridge, Mass: Harvard University Press, 1985.

CHAMBERLAIN, Gary (1984). – "Panel Data", *Handbook of Econometrics*, ed. Zvi Griliches and M. Intrilligator, Amsterdam, North Holland, 1984.

HSIAO, Cheng (1985). – "Benefits and Limitations of Panel Data", *Econometric Reviews*, 4, pp. 121-74.

JUDGE, George G., GRIFFITHS, William E., CARTER HILL, R. and TSEUNG-CHAO, Lee (1980). – *The Theory and Practice of Econometrics*, New York: Wiley.

LOÈVE, M. (1977). – *Probability Theory I*, fourth ed., New York: Springer-Verlag.

NEYMAN, J. and SCOTT, E. L. (1948). – "Consistent Estimates Based on Partially Consistent Observations", *Econometrica*, 16, pp. 1-32.

WALD, A. (1949). – "Note on the Consistency of the Maximum Likelihood Estimate". *Annals of Mathematical Statistics*, 60, pp. 595-601.

WHITE, H. (1980). – "A Heteroskedasticity-Consistent Covariance Estimator and a Direct Test for Heteroskedasticity", *Econometrica*, 48, pp. 817-838.