Estimating Multi-Way Error Components Models with Unbalanced Data Structures[*]

Peter Davis[†]
MIT Sloan

This version: December 2000.

Abstract

I develop simple matrix algebra techniques that simplify and unify much of the previous literature on estimating error components models (ECMs). In fact, the simple analytic results provided here are useful for analyzing a very broad set of models with complex error structures. To illustrate the techniques, I develop the algebra for three- and four-way ECMs explicitly. In addition I provide Monte Carlo simulation evidence on the performance of several estimators for the three-way ECM and estimate the model using data from a retail market where the three dimensions of data variation are products selling in many locations over time.

Key Words: Unbalanced Panel data, Instrumental variables, Error Component Models.
JEL Classification: C13, C23.

[*]Thanks are due to Steve Berry, Ariel Pakes, Nadia Soboleva, and Tom Stoker for helpful comments and suggestions. This paper is a substantially revised version of Chapter 3 of my Ph.D. dissertation which has benefited greatly from the suggestions of two anonymous referees and an associate editor of this journal. Financial Support from Yale University and the Robert M. Leylan Doctoral Fellowship Fund is gratefully acknowledged.

[†]Mailing address: Assistant Professor of Applied Economics, E52-449, MIT Sloan School of Management, 50 Memorial Drive, Cambridge, MA 02142. Tel: (617) 253-1277, fax: (617) 258-6855, e-mail: pjdavis@MIT.EDU, Web: http://web.mit.edu/pjdavis/www

1 Introduction

The error components model (ECM) is one of the most frequently used econometric models for panel data. Work by Balestra and Nerlove (1973), Wallace and Hussain (1969), Amemiya (1971),Nerlove (1971a), Nerlove (1971b), Mazodier (1972) and Fuller and Batese (1974) among others, provide a comprehensive analysis of the one- and two-way ECM for balanced panels. In practice, researchers often have to deal with missing observations. Consequently, Baltagi (1985) and Wansbeek and Kapteyn (1989) (henceforth WK) respectively developed the one- and two-way ECM for unbalanced data.

In a related literature, the nested ECM has also been developed. Most recently, Baltagi, Song, and Jung (1999b) and Antweiler (1998) respectively develop the two- and three-way nested ECM for unbalanced panels, building on earlier work by Fuller and Battesse (1973). The nested ECM has proven particularly useful for datasets with a natural grouping structure. For example, data on firms may be grouped by industry.

The aim of this paper is to extend the literature in a number of directions. First, I provide a series of mathematical results which allow an elegant analysis of the ECM for unbalanced panel data with an *arbitrary* number of error components.[1] I illustrate the results by developing estimators for the three-way ECM in the text while results for the four-way model are presented in an appendix.[2] By developing the paper's results as recurrence relations, the extensions to the five- or more-way ECMs are easily developed, while the one- and two-way ECMs are observed as special cases.

Second, I show that the standard and nested ECMs have the same mathematical structure and therefore the results I present can be applied to both. In fact, the results are

[1]Datasets with three or more dimensions of data variation are now commonly available to applied researchers in fields such as marketing and industrial organization. For example, when estimating differentiated product demand systems, perhaps using the techniques developed by Berry (1994) and Berry, Levinsohn, and Pakes (1995), observations on a cross-section of products, sold in many locations over time are frequently available. Disaggregated data sets are usually unbalanced. Fortunately, the ideas and models developed for panel data can be extended to allow for these richer data structures and using them makes it possible to empirically distinguish between models which could not be differentiated using only panel data (see Pakes (1983) for example.)

[2]Ghosh (1976) develops the three-way ECM for balanced panels. See also Baltagi (1987).

useful for estimating a broad array of models with complex error structures. In each case, modeling the correlation structure among the errors is important because it allows efficient estimation of the parameters of interest and consistent estimates of the standard errors of the parameters. Naturally, the latter is crucially important for performing inference correctly.

Third, whenever possible, I relax WK's assumption that all regressors are exogenous, thereby allowing estimation of the general ECM using instrumental variables. Applied demand analysis motivated these extensions of the literature and this aspect of the results is particularly important in that and other contexts.[3]

The paper proceeds as follows: Sections 2 and 3 develop the multi-way fixed- and random-effects models respectively, for given estimates of the variance of each error component. In section 4 quadratic unbiased estimators (QUE's), minimum norm quadratic unbiased estimators (MINQUE's) and maximum likelihood estimators (MLE's) of the variance of each error component are developed. Section 5 provides Monte Carlo evidence on estimators of the three-way ECM. In section 6, I apply the results by estimating a differentiated product demand equation using a new dataset from a retail market. Section 7 concludes.

2 The Fixed Effects Model

Consider the three-way error components linear regression model:

$$y_{ijk} = X_{ijk}\beta + u_{ijk}, \quad \text{where} \quad u_{ijk} = \mu_i + \lambda_j + \gamma_k + \nu_{ijk} \tag{1}$$

[3]The previous literature on instrumental variable estimation of ECMs has emphasized two other important contexts: dynamic panel data and simultaneous equation ECMs. Instrumental variable estimation of the one-way ECM with lagged endogenous variable is considered by Hausman and Taylor (1981), Arellano and Bond (1991) and Arellano and Bover (1995) among others. Baltagi (1981) and Amemiya and MaCurdy (1986) consider instrumental variable estimation of the simultaneous equation ECM. The reader is directed to Baltagi (1995) and the references therein for further details.

and $i = 1, .., N_1$ might index products selling in location $j = 1, .., N_2$ at date $k = 1, .., N_3$. In that case, N_1 would denote the total number of different products ever observed being sold at any location in any time period. Similarly, N_2 and N_3 would denote the number of locations and time periods for which data are available. Let n denote the total number of observations in the dataset.

Using matrix notation, the three-way ECM may be written

$$y \;=\; X\beta + u, \quad \text{where} \;\; u = \Delta_1\mu + \Delta_2\lambda + \Delta_3\gamma + \nu, \tag{2}$$

$\mu' = (\mu_1, ..., \mu_{N_1})$, $\lambda' = (\lambda_1, ..., \lambda_{N_2})$ and $\gamma' = (\gamma_1, ..., \gamma_{N_3})$ are parameters in the fixed effects model, Δ_s for $s = 1, 2, 3$ are $n\text{x}N_s$ matrices of indicator variables denoting observations on products, locations and time periods respectively, and ν is the vector of econometric unobservables.

The three-way nested ECM is

$$y_{ijkl} = X_{ijkl}\beta + u_{ijkl}, \quad \text{where} \;\; u_{ijkl} = \mu_i + \lambda_{ij} + \gamma_{ijk} + \nu_{ijkl} \tag{3}$$

where $i = 1, .., N_1$, $j = 1, .., N_2$, $k = 1, .., N_3$ and $l = 1, .., N_4$. Writing the model in matrix form also yields equation (2), but with $\mu = (\mu_1, .., \mu_{N_1})$, $\lambda = (\lambda_{11}, .., \lambda_{1N_2}, .., \lambda_{N_1N_2})$, $\gamma = (\gamma_{111}, .., \gamma_{N_1,N_2,N_3})$, and Δ_s $s = 1, .., 3$ are matrices of zeros and ones denoting inclusion of the observation in the relevant group.[4]

When μ, λ, and γ are parameters and the data set is large, direct estimation of either model is often infeasible. For example, OLS requires the inverse of a matrix which grows with the number of columns in the entire data and indicator variable matrix, $(X, \Delta_1, \Delta_2, \Delta_3)$.

[4]Clearly, the appropriate definition of Δ_s will change between the conventional ECM and the nested ECM. However, the mathematical results provided below do not rely on the detailed structure of these inclusion matrices so I shall consider the two classes of models together. This fact also means that many other error component structures, such as combinations of nested and non-nested ECMs, can be considered using the same analytic framework. In addition, the same feature of the results also implies that the ordering of the data is immaterial provided that the inclusion matrices are appropriately defined.

To circumvent this familiar difficulty, we may use the "within" transformation to obtain estimates of the parameters of interest.[5]

Developing an expression for the within transformation requires various projection arguments. Throughout, I follow the convention that $P_{[A]} = A(A'A)^{-}A'$ is a matrix (or corresponding linear transformation) which projects onto the range of A, denoted $R(A)$. Similarly, denote the orthogonal projection onto the null space of the matrix A, $N(A)$, by $Q_{[A]} = I - P_{[A]}$. First, I establish that the orthogonal projection onto the columns of an arbitrary matrix X may be performed in stages. This result allows the within transformation to be performed in stages, reducing the computational burden substantially.

Lemma 1 *Let $X = (X_1, X_2)$, and $P_{[\cdot]}$ denote an orthogonal projection onto $R(\cdot)$, then $P_{[X]} = P_{[X_1]} + P_{[Q_{[X_1]}X_2]}$.*

Proof: (See Appendix.)

Clearly, with $X = (\Delta_2, \Delta_1)$ this lemma corresponds to the lemma in WK, p.344. Applying the lemma to cover the three- or four-way models is simple. Here I present the results for the three-way ECM.

Corollary 1 *Let $Y = (Y_1, Y_2, Y_3)$, then $P_{[Y]} = P_{[A]} + P_{[B]} + P_{[C]}$ where*

$$A = Y_3, \quad B = Q_{[A]}Y_2 \ \text{ and } \ C = Q_{[B]}Q_{[A]}Y_1. \tag{4}$$

Proof: Let $X_1 = (Y_1, Y_2)$ and $X_2 = Y_3 = A$. By lemma 1, $P_{[X]} = P_{[X_2]} + P_{[Q_{[X_2]}X_1]} = P_{[A]} + P_{[Q_{[X_2]}X_1]}$. Next define $\tilde{Y}_1 = Q_{[Y_3]}Y_1$ and $B = Q_{[Y_3]}Y_2$. Applying lemma 1 a second time yields $P_{[\tilde{Y}_1, B]} = P_{[B]} + P_{[Q_{[B]}\tilde{Y}_1]}$. Hence, $P_{[Y]} = P_{[A]} + P_{[B]} + P_{[C]}$. **Q.E.D.**

Defining the matrix $\Delta = (\Delta_1, \Delta_2, \Delta_3)$, the within transformation of the three-way error

[5] Recall that OLS estimates of β from the full model are equivalently provided by transforming y and X using the within transformation and performing OLS on the transformed regression.

components model is therefore given by $Q_{[\Delta]} = Q_{[A]} - P_{[B]} - P_{[C]}$, where[6,7]

$$\begin{aligned} Q_{[A]} &= I - \Delta_3(\Delta_3'\Delta_3)^{-}\Delta_3' \\ P_{[B]} &= Q_{[\Delta_3]}\Delta_2(\Delta_2' Q_{[\Delta_3]}\Delta_2)^{-}\Delta_2' Q_{[\Delta_3]} \\ P_{[C]} &= Q_{[A]}Q_{[B]}\Delta_1(\Delta_1'(Q_{[A]}Q_{[B]})\Delta_1)^{-}\Delta_1' Q_{[A]}Q_{[B]}. \end{aligned}$$

The parameters of the transformed model $Q_{[\Delta]}y = Q_{[\Delta]}X\beta + Q_{[\Delta]}\nu$ can be consistently estimated as $n \to \infty$ using any standard instrumental variable estimator, such as two stage least squares, provided the researcher has a sufficient number of within transformed instrumental variables $Q_{[\Delta]}Z$ to identify the parameters on the transformed regressors $Q_{[\Delta]}X$. The within two stage least squares estimator is

$$\hat{\beta}^{W2SLS} = (X' P_{QZ} X)^{-1} X' P_{QZ} y$$

where $P_{QZ} = Q_{[\Delta]}Z(Z'Q_{[\Delta]}Z)^{-1}Z'Q_{[\Delta]}$.[8] Special cases of this estimator include the single equation variant of the error component two stage least squares estimator derived in Baltagi (1981) and WK's within estimator which takes $Z = X$ in the two-way model.

3 The Random Effects Model

Equation (2) describes the error structure of the three-way random effects model, when μ, λ, γ and ν are random vectors. I assume μ, λ, γ and ν are independent of each other and among themselves with zero means and variances σ_1^2, σ_2^2,σ_3^2 and σ_0^2, respectively. Thus, $\Omega \equiv E[uu'] = \sigma_0^2 I_n + \sigma_1^2\Delta_1\Delta_1' + \sigma_2^2\Delta_2\Delta_2' + \sigma_3^2\Delta_3\Delta_3'$.

---

[6]Note that $Q_{[A]}Q_{[B]} = Q_{[A]} - P_{[B]} = Q_{[B]}Q_{[A]}$ and $P_{[B]}Q_{[A]} = P_{[B]}$ so $Q_{[A]}Q_{[B]}Q_{[A]} = Q_{[A]}Q_{[B]}$.

[7]Since the only requirement is that each matrix represent an orthogonal projection, any generalized inverse may be used.

[8]Naturally, we may not wish to explicitly calculate the (nxn) matrix P_{QZ}. Fortunately, for arbitrary matrices A and B with dimensions (axn) and (nxb) respectively, the matrix $AP_{QZ}B = (AQ_{[\Delta]}Z)(Z'Q_{[\Delta]}Z)^{-1}(Z'Q_{[\Delta]}B)$. Similarly $AQ_{[\Delta]}B = AB - AP_{[\Delta]}B = AB - (A\Delta)(\Delta'\Delta)^{-1}(\Delta' B)$. By grouping the matrices before multiplication, in the way indicated by the parentheses, the evaluation of all (nxn) matrices is avoided.

Given a consistent estimate of Ω^{-1}, a set of instruments Z and the identification condition that $E[u(X,y;\beta_0)|Z,\Delta]=0$ at the true parameter β_0, we may use the GMM framework developed by Hansen (1982) to provide an estimator for β. Specifically, defining $\hat{\beta}$ as the solution to the minimum distance problem

$$\min_{\beta\in\Theta}\|G_n(\beta)\|_A \quad \text{where} \quad G_n(\beta)=\frac{1}{n}Z'\Omega^{-1}u \quad \text{and} \quad A=Var[\lim_{n\to\infty}\sqrt{n}G_n(\beta)],$$

it is easy to show that

$$\begin{aligned}\hat{\beta} &= (X'\Omega^{-1/2}P_{[\Omega^{-1/2}Z]}\Omega^{-1/2}X)^{-1}X'\Omega^{-1/2}P_{[\Omega^{-1/2}Z]}\Omega^{-1/2}y, \quad \text{and}\\ Var(\hat{\beta}) &= (X'\Omega^{-1/2}P_{[\Omega^{-1/2}Z]}\Omega^{-1/2}X)^{-1}.\end{aligned}$$

Conceptually therefore, efficient estimation of an ECM with unbalanced data using instrumental variables poses no new problems. In practice, two problems remain. First, to provide a practical technique for calculating Ω^{-1} for given values of the variances associated with each error component, $(\sigma_0^2,\sigma_1^2,\sigma_2^2,\sigma_3^2)$. This is a non-trivial task since Ω is nxn and n is typically too large for a direct numerical calculation of the inverse to be feasible. The second remaining problem is to provide consistent estimates of the variance components. I postpone the latter to Section 4.

Notice that, however many error components there are, the variance-covariance matrix of the composite error term always has the form $\Omega=\sigma_0^2(I+XX')$. Thus, $\Omega^{-1}=\sigma_0^{-2}(I+XX')^{-1}$. In fact, *any* variance-covariance matrix may always be written in this form, provided only that there is at least one idiosyncratic component which is uncorrelated across observations. Lemma 2 provides an expression for the inverse of a matrix of the general form $I+XX'$.

Lemma 2 *If* $|I+XX'|\neq 0$, *then* $(I+XX')^{-1}=\tilde{Q}_{[X_2]}-\tilde{Q}_{[X_2]}X_1S_1^{-1}X_1'\tilde{Q}_{[X_2]}$ *where*

$X = (X_1, X_2)$ *and*

$$\tilde{Q}_{[X_2]} \;=\; I - X_2(I + X_2'X_2)^{-1}X_2' = (I + X_2X_2')^{-1} \quad \text{and} \quad S_1 = I + X_1'\tilde{Q}_{[X_2]}X_1.$$

Proof (See Appendix A.)

By implicitly applying this lemma once, with $X = (\frac{\sigma_1}{\sigma_0}\Delta_1, \frac{\sigma_2}{\sigma_0}\Delta_2)$, WK provide an analytic expression for the two-way error components model. Now suppose $X_1 = Y_1$ and $X_2 = (Y_2, Y_3)$. Since $\tilde{Q}_{[X_2]} = I - X_2(I + X_2'X_2)^{-1}X_2' = (I + X_2X_2')^{-1}$, lemma 2 may be applied a second time to provide an analytic expression for $\tilde{Q}_{[X_2]} = \tilde{Q}_{[Y_2,Y_3]}$.

Corollary 2 *If* $Y = (Y_1, Y_2, Y_3)$ *and* $\Omega = \sigma_0^2(I + YY')$, *then* $\Omega^{-1} = \sigma_0^{-2}(I + YY')^{-1}$ *where* $(I + YY')^{-1} = \tilde{Q}_{[Y_2,Y_3]} - \tilde{Q}_{[Y_2,Y_3]}Y_1S_1^{-1}Y_1'\tilde{Q}_{[Y_2,Y_3]}$ *and*

$$\begin{aligned}
\tilde{Q}_{[Y_2,Y_3]} &= \tilde{Q}_{[Y_3]} - \tilde{Q}_{[Y_3]}Y_2S_2^{-1}Y_2'\tilde{Q}_{[Y_3]}, \qquad \tilde{Q}_{[Y_3]} = I - Y_3(I + Y_3'Y_3)^{-1}Y_3' \\
S_1 &= I + Y_1'\tilde{Q}_{[Y_2,Y_3]}Y_1, \qquad S_2 = I + Y_2'\tilde{Q}_{[Y_3]}Y_2.
\end{aligned}$$

Proof (See Appendix A.)

Notice that the inverses of three matrices remain to be calculated: S_1, S_2 and $(I + Y_3'Y_3)$. These matrices have dimensions that are the number of columns of Y_1, Y_2 and Y_3 respectively. Moreover, if the columns of Y_3 are a collection of indicator variables representing mutually exclusive groupings of the data, $Y_3'Y_3$ will be diagonal and consequently, this third inversion can be performed analytically.[9]

Applying lemma 2 multiple times, expressions for Ω^{-1} are easily obtained for the four- and five-way ECM's. In each case the variance-covariance matrix is of the form $(I + XX')$.

[9] Judicious choice of Y_3 will minimize the dimension of the numerical inversions required. For example, in the application presented in Section 6, N_2 and N_3 are small relative to N_1, thus the required computational resources are minimized by setting $Y = (\frac{\sigma_2}{\sigma_0}\Delta_2, \frac{\sigma_3}{\sigma_0}\Delta_3, \frac{\sigma_1}{\sigma_0}\Delta_1)$ before applying the lemma.

Indeed, since the lemma may be applied any number of times, models with an arbitrary number of error components can easily be developed. In each case, inversions of nxn matrices are avoided entirely and numerical inversion is avoided entirely in one dimension of the data.

4 Estimation of the Variance Components

The expression for Ω^{-1} reduces computational time considerably relative to inverting Ω numerically given any specified values of the variances of the error components. Various methods are available to estimate the variance components themselves. In this section I develop analysis of variance type quadratic unbiased estimators (QUEs) analogous to WK's QUE's for the two-way model. Then I present the Minimum Norm Quadratic Unbiased Estimator (MINQUE) developed by Rao (1971) and finally maximum likelihood estimators.

4.1 Quadratic Unbiased Estimators

Define $e = y - X\hat{\beta}^{W2SLS}$, $q_n = e'Q_{[\Delta]}e$ and $q_i = e'P_{[\Delta_i]}e$ for $i = 1, 2, 3$. In Lemma 3, I provide expressions for the expected values of q_n, q_1, q_2, and q_3. QUE's are obtained by solving the four equations that result from equating the observed quadratic forms q_n, q_1, q_2 and q_3 to their expectations for the parameters σ_0^2, σ_1^2, σ_2^2, and σ_3^2.

Lemma 3

$$
\begin{aligned}
E[q_n] &= (n - t_1 - t_2 - t_3 + k_n)\sigma_0^2 \\
E[q_1] &= (N_1 + k_1)\sigma_0^2 + n\sigma_1^2 + k_{12}\sigma_2^2 + k_{13}\sigma_3^2 \\
E[q_2] &= (N_2 + k_2)\sigma_0^2 + k_{21}\sigma_1^2 + n\sigma_2^2 + k_{23}\sigma_3^2 \\
E[q_3] &= (N_3 + k_3)\sigma_0^2 + k_{31}\sigma_1^2 + k_{32}\sigma_2^2 + n\sigma_3^2
\end{aligned}
$$

where $k_n = tr\{(X'P_{[QZ]}X)^{-1}X'Q_{[\Delta]}X\}$, $k_i = tr\{(X'P_{[QZ]}X)^{-1}X'P_{[\Delta_i]}X\}$, $k_{ij} = tr\{\Delta_j'P_{[\Delta_i]}\Delta_j\}$

for $i,j = 1,2,3,\ t_1 = tr\{(A'A)^{-}A'A)\} = rank(A),\ t_2 = tr\{(B'B)^{-}B'B\} = rank(B)$ *and* $t_3 = tr\{(C'C)^{-}C'C\} = rank(C)$,[10] *where* A, B, C *are defined as in corollary 1.*[11]

Proof

Let $M \equiv I_n - X(X'P_{QZ}X)^{-1}X'P_{QZ}$, then by definition $e = My = Mu$. Since $Q_{[\Delta]}\Delta_i = 0$ for $i = 1,2,3$, $Q_{[\Delta]}\Omega = \sigma_0^2 Q_{[\Delta]}$, $M\Omega = \Omega - \sigma_0^2 X(X'P_{QZ}X)^{-1}X'P_{QZ}$ and $M\Omega M' = \Omega - \sigma_0^2 X(X'P_{QZ}X)^{-1}X'P_{QZ} - \sigma_0^2 P_{QZ}X(X'P_{QZ}X)^{-1}X' + \sigma_0^2 X(X'P_{QZ}X)^{-1}X'$.

Since $P_{[\Delta_i]}P_{[QZ]} = 0$, $tr\{P_{[\Delta_i]}M\Omega M'\} = tr\{P_{[\Delta_i]}\Omega\} + \sigma_0^2 tr\{P_{[\Delta_i]}X(X'P_{QZ}X)^{-1}X'\} = tr\{P_{[\Delta_i]}\Omega\} + \sigma_0^2 k_i$. Hence,

$$\begin{aligned} E[q_i] &= E[e'P_{[\Delta_i]}e] = tr\{P_{[\Delta_i]}M\Omega M'\} = tr\{P_{[\Delta_i]}\Omega\} + \sigma_0^2 k_i \\ E[q_n] &= E[e'Q_{[\Delta]}e] = tr\{Q_{[\Delta]}M\Omega M'\} = \sigma_0^2 tr\{Q_{[\Delta]}\} + \sigma_0^2 k_n \end{aligned}$$

The results follow immediately from these expressions since $tr\{P_{[\Delta_i]}\Delta_i\Delta_i'\} = n$, $tr\{P_{[\Delta_i]}\} = N_i$, and $tr\{Q_{[\Delta]}\} = tr\{I_n - P_{[A]} - P_{[B]} - P_{[C]}\} = n - t_1 - t_2 - t_3$. **Q.E.D.**

4.2 MINQUE Estimation of the variance components

The Minimum Norm Quadratic Unbiased Estimator of a general error component model is developed in Rao (1971). Let

$$\begin{aligned} R &= \Omega^{-1}[I_n - X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}] \\ S &= \{s_{i+1,j+1}\} = \{tr\{V_i R V_j R\}\},\quad i,j = 0,..,3 \text{ where} \\ u &= \{u_{i+1}\} = \{y'RV_iRy\}\quad i = 0,..,3 \text{ with} \\ V_i &= \Delta_i\Delta_i'\quad i = 0,..,3 \text{ and } \Delta_0 \equiv I_n. \end{aligned}$$

[10] *For any matrix A, $trace((A'A)^{-}A'A) = rank(A'A) = rank(A)$. The first equality follows since $(A'A)^{-}A'A$ is a projection matrix and the latter since the $range(A) = range(A'A)$. See, for example, Lancaster and Tismenetsky (1985), p.430 and Ostaszewski (1990), p.41. respectively.*

[11] *Notice that the constants k_{ij} just represent average sums of squares. For example, in the conventional error components model if Δ_1 and Δ_2 contain the indicator variable structure for products and locations respectively, n_i denotes the total number of observations available on product i, and n_{ij} denotes the total number of observations available on product i at location j, then $k_{12} = tr\{\Delta_2'P_{[\Delta_1]}\Delta_2\} = \sum_{i=1}^{N_1}\frac{1}{n_i}\sum_{j=1}^{N_2}n_{ij}^2$.*

Rao (1971) shows that the vector of MINQUE's is given by $\hat{\theta} = S^{-1}u$ where $\hat{\theta}' = (\hat{\sigma}_0^2, \hat{\sigma}_1^2, \hat{\sigma}_2^2, \hat{\sigma}_3^2)$. Clearly, an initial estimate of Ω is required to compute the matrix R, which is required to compute the MINQUE. The WK estimators provided in the previous sections provide one way to construct the required initial estimate of Ω. Appendix D provides the computational details.

4.3 MLE and the Information Matrix

Amemiya (1971) studied maximum likelihood estimation of complete data ECMs. Applying the general results obtained by Magnus (1978), the first order conditions for ML estimation of β and the parameters in Ω are

$$\hat{\beta} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y \tag{5}$$

$$tr\{\hat{\Omega}_\theta^{-1}\Omega\} = e'\hat{\Omega}_\theta^{-1}e, \text{ for } \theta = \sigma_0^2, \sigma_1^2, \sigma_2^2 \text{ and } \sigma_3^2 \tag{6}$$

with $e = y - X\hat{\beta}$ and Ω_θ^{-1} the partial derivative of Ω^{-1} with respect to θ. A well known result in matrix differentiation is that $\Omega_\theta^{-1} = -\Omega^{-1}\Omega_\theta\Omega^{-1}$ (see Magnus and Neudecker (1999) for example.) Defining $\Delta_0 \equiv I_n$, it follows immediately that

$$e'\Omega_\theta^{-1}e = -e'\Omega^{-1}\Omega_\theta\Omega^{-1}e = -e'\Omega^{-1}\Delta_i\Delta_i'\Omega^{-1}e \quad \text{if } \theta = \sigma_i^2,\ i = 0,..,3 \tag{7}$$

while

$$tr\{\Omega_\theta^{-1}\Omega^{-1}\} = -tr\{\Omega^{-1}\Omega_\theta\} = -tr\{\Delta_i'\Omega^{-1}\Delta_i\} \quad \text{if } \theta = \sigma_i^2,\ i = 0,..,3. \tag{8}$$

These expressions allow both sides of the first order conditions for the ML estimators of the variance components to be evaluated for each proposed value of the four parameters $(\sigma_0^2, \sigma_1^2, \sigma_2^2, \sigma_3^2)$. In Appendix C, I demonstrate that each of these expressions may easily be evaluated without explicitly calculating any nxn matrices. Thus, finding the MLEs amounts to finding a solution of these four equations in the four unknowns.[12] As usual, the recursive structure of the matrix algebra ensures that this is will also be true for the fourth or higher variance component models. In each case, good initial values are available from the QUE estimators derived previously.

Following Magnus (1978), the information matrix of the log-likelihood function is given by the $(k+4)$x$(k+4)$ matrix

$$\Phi \;=\; \begin{bmatrix} X'\Omega^{-1}X & 0 \\ 0 & \frac{1}{2}\Phi_\theta \end{bmatrix},$$

where $(\Phi_\theta)_{i+1,j+1} = tr\{\Omega_{\theta_i}^{-1}\Omega\Omega_{\theta_j}^{-1}\Omega\} = tr\{\Delta_i'\Omega^{-1}\Delta_j\Delta_j'\Omega^{-1}\Delta_i\}$ for $i,j = 0,..,3$ $\theta_i = \sigma_i^2$ and $\theta_j = \sigma_j^2$.

In Appendix C, I demonstrate that each element in the information matrix may be calculated without explicitly evaluating any $(n$x$n)$ matrices. In Appendix D, I provide an alternative method for directly evaluating the likelihood function.[13]

[12] In fact, by transforming the parameters of the model, the non-analytic dimensions of the first order conditions can be reduced to only three. Specifically, by writing the likelihood function in terms of β, σ_0^2 and the transformed parameters $\rho_i = \frac{\sigma_i^2}{\sigma_0^2}$, the MLE of σ_0^2 is available analytically for each value of $\beta, \rho_1, \rho_2, \rho_3$: $\hat{\sigma}_0^2 = e'(I + \sum_{i=1}^3 \rho_i\Delta_i\Delta_i')^{-1}e/n$.

[13] Having provided an expression for the likelihood function and information matrix, Lagrange multiplier and Likelihood ratio test statistics can easily be computed using the results developed in this section and Appendix C. The availability of likelihood ratio tests obviously follow directly from the ability to evaluate the likelihood function for any number of error components, while Lagrange Multiplier tests require the derivative of the likelihood and the relevant component of the information matrix evaluated at the restricted parameter point:

$$\frac{\partial L}{\partial \theta} = -\frac{1}{2}tr\{\Omega^{-1}\Omega_\theta\} + \frac{1}{2}u'\Omega^{-1}\Omega_\theta\Omega^{-1}u \tag{9}$$

while

$$E[-\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}] = \frac{1}{2}tr\{\Omega^{-1}\Omega_{\theta_i}\Omega^{-1}\Omega_{\theta_j}\}. \tag{10}$$

5 Monte Carlo Simulations

5.1 Design of the Monte Carlo Study

While previous Monte Carlo studies of ECMs with unbalanced data have considered designs which emphasize the traditional panel data context with observations missing in only one dimension (individuals are only observed in some time periods), I consider the simplest possible regression model with a three-way error component structure and data sets which are symmetrically unbalanced in multiple dimensions. Specifically, I use the data generating process

$$y_{ijk} = x_{ijk}\beta + \mu_i + \lambda_j + \gamma_k + \nu_{ijk},$$

where the exogenous variable $x_{ijt} \sim IIN(0,1)$, β is fixed at 1, $\mu_i \sim IIN(0,\sigma_1^2)$, $\lambda_j \sim IIN(0,\sigma_2^2)$, $\gamma_k \sim IIN(0,\sigma_3^2)$, and $\nu_{ijk} \sim IIN(0,\sigma_0^2)$. I fix the total variance of the error term to be $\sigma_0^2 + \sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 1$ vary each σ_i^2 over the set $\{0, 0.2, 0.4, 0.6, 0.8\}$ such that $\sigma_0^2 = 1 - \sum_{i=1}^{3} \sigma_i^2 > 0$. Following Ahrens and Pincus (1981), I define the three, one-way, measures of unbalancedness

$$r_i = \frac{N_i^2}{tr\{(\Delta_i'\Delta_i)^{-1}\}tr\{\Delta_i'\Delta_i\}}, \quad i = 1, 2, 3.$$

Note that r_i takes the value 1 when the data are balanced in *both* of the other two dimensions, but takes on smaller values as the data pattern becomes more unbalanced. In contrast to the Monte Carlo studies for the two-way model reported in Baltagi and Chang (1994) and Baltagi, Song, and Jung (1999a),who use an asymmetric design with unbalancedness in one dimension, I consider a symmetrically unbalanced design with unbalancedness in multiple dimensions and with $r_1 = r_2 = r_3$.[14] This difference reflects the differences in typical data sets; three dimensional data structures are often multi-dimensionally unbalanced while panel

[14]In terms of the AP unbalancedness statistics, Baltagi and Chang (1994) and Baltagi, Song, and Jung (1999a) consider designs which allow r_1 to vary, but constrain $r_2 = 1$. The symmetric design in the two-way case would consider the case where both r_1 and r_2 decrease, but do so with $r_1 = r_2$.

data unbalancedness arises from not observing all individuals in every time period (while some data is usually observed each time period.)

A symmetrically unbalanced data pattern with three dimensions of data variation can be described by the matrix $\{(a,b,c),(b,c,a),(c,a,b)\}$ where the first vector (a,b,c) denotes the data pattern: observe a products at b stores for c time periods. The second vector denotes observe b products at c stores for a time periods and so forth. Thus, the matrix generates a data set with 3xaxbxc observations. I report this symmetric pattern using the shorthand notation, (a,b,c). The advantage of this symmetric approach is that it provides a single index of unbalancedness while, at the same time, allowing multiple dimensions of unbalancedness.

I consider three unbalancedness patterns $(4,4,4)$, $(4,8,2)$ and $(2,16,2)$. The first is clearly a balanced data pattern, while the others are increasingly unbalanced. The AP statistics for these three patterns are 1, 0.777 and 0.505 respectively. For each case, I compute estimates of the variance components using the Wansabeek and Kapteyn (WK) method, the Maximum Likelihood Estimator (MLE), and three types of MINQUE estimators corresponding to three initial values. MV_1 reports MINQUE taking the initial value of the variance components to each be equal to zero and the initial value of $\sigma_0^2 = 1$. MV_2 takes the initial value of $\sigma_i^2 = 1$, $i = 0,1,2,3$ and MV_3 takes the computationally more intensive approach of using the WK estimates as initial values. The results for the linear parameters also report OLS and the within (WTN) estimators for completeness.

5.2 Comparison of the Estimators

Table 1 reports the mean square error (MSE) of the various estimators of β relative to that of true GLS. The results are relatively similar to those presented by Baltagi, Song, and Jung (1999a) for the linear paratmers in the two-way model. Specifically, OLS is inferior to GLS,MLE,MV_1,MV_2, and MV_3 unless $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0$, while it will be superior to the WTN estimator if the σ_i^2's are all sufficiently close to zero since OLS is equivalent to MLE in that case. The WK and MV estimators perform well relative to true GLS for a correctly specified three-way model and are much simpler to compute. Even when the model is misspecified, the simpler ANOVA methods demonstrate relatively good performance. The only exception is MV_1 which performs poorly relative to the other estimators when the data is highly unbalanced. In contrast, if the data pattern is sufficiently unbalanced, the WTN estimator can suffer from non-trivial increases in MSE over GLS. For example, with the unbalanced data pattern 3, $\sigma_1^2 = \sigma_2^2 = 0$, and $\sigma_3^2 = 0.8$ the WTN increase in MSE reaches 35%. Clearly, over-transforming the data in this way can lead to poor small sample performance of the within estimator.

Table 2 reports the MSE of the variance components relative to that of MLE for the three unbalanced patterns considered. Not much difference between the WK, MV_2 and MV_3 estimates of σ_0^2 is evident in the table, however all three show MSE performance which is significantly better than MLE when the true value of σ_0^2 generates twenty percent or less of the total variance in the error term and all the variance components are non-zero. Since MV_1 and MV_2 both use biased initial estimators of the variance components, one might expect both of these estimators to show poorer performance when the true values of the variance components differ substantially from the initial estimates. Somewhat surprisingly however, MV_1 performs poorly relative to MV_2 generally and demonstrates consistently worse performance and as σ_0^2 decreases toward zero.

The remaining columns report the MSE of the various estimators of σ_i^2 for $i = 1, .., 3$ relative to the MSE of MLE. The most striking pattern in the table is that the WK, MV_1, MV_2, and MV_3 estimators are evidently not robust to the model misspecification when some, but not all, variance components in the data generating process are truthfully zero.

Pattern	σ_1^2	σ_2^2	σ_3^2	GLS	OLS	WTN	WK	MLE	MV1	MV2	MV3
(4,4,4)	0.00	0.00	0.00	1.00	1.00	1.19	1.01	1.01	1.01	1.01	1.01
(4,4,4)	0.00	0.00	0.20	1.00	1.14	1.14	1.02	1.01	1.02	1.02	1.01
(4,4,4)	0.00	0.00	0.40	1.00	1.86	1.13	0.99	0.98	0.99	0.99	0.99
(4,4,4)	0.00	0.00	0.60	1.00	2.39	1.17	1.04	0.99	1.05	1.02	1.01
(4,4,4)	0.00	0.00	0.80	1.00	3.85	1.06	1.02	0.99	1.02	1.00	0.98
(4,4,4)	0.00	0.20	0.20	1.00	1.58	1.03	1.01	1.00	1.01	1.00	1.00
(4,4,4)	0.00	0.20	0.40	1.00	2.57	1.08	0.99	0.99	1.00	0.99	0.99
(4,4,4)	0.00	0.20	0.60	1.00	4.02	1.07	1.08	1.03	1.12	1.03	1.02
(4,4,4)	0.00	0.40	0.40	1.00	4.80	1.00	1.00	1.00	1.01	0.99	1.00
(4,4,4)	0.20	0.20	0.20	1.00	2.11	1.02	1.04	1.04	1.05	1.00	1.00
(4,4,4)	0.20	0.40	0.20	1.00	5.36	1.01	1.04	1.05	1.04	1.00	1.00
(4,8,2)	0.00	0.00	0.00	1.00	1.00	1.26	1.01	1.02	1.02	1.01	1.02
(4,8,2)	0.00	0.00	0.20	1.00	1.21	1.18	1.01	1.01	1.01	1.02	1.02
(4,8,2)	0.00	0.00	0.40	1.00	1.27	1.17	1.01	1.01	1.02	1.00	1.00
(4,8,2)	0.00	0.00	0.60	1.00	2.48	1.07	1.02	1.01	0.98	1.01	1.00
(4,8,2)	0.00	0.00	0.80	1.00	4.94	1.27	1.08	1.01	1.11	1.04	1.04
(4,8,2)	0.00	0.20	0.20	1.00	1.28	1.13	1.00	0.99	1.03	0.98	0.98
(4,8,2)	0.00	0.20	0.40	1.00	1.94	1.05	1.03	1.02	1.07	1.01	1.01
(4,8,2)	0.00	0.20	0.60	1.00	5.12	1.14	1.06	1.03	1.14	1.03	1.02
(4,8,2)	0.00	0.40	0.40	1.00	3.97	1.03	1.04	1.01	1.12	1.01	1.00
(4,8,2)	0.20	0.20	0.20	1.00	1.72	1.05	1.01	1.02	0.99	1.01	1.01
(4,8,2)	0.20	0.40	0.20	1.00	4.16	1.01	1.07	1.04	1.17	1.00	1.00
(2,16,2)	0.00	0.00	0.00	1.00	1.00	1.36	1.01	1.00	1.00	1.03	1.01
(2,16,2)	0.00	0.00	0.20	1.00	1.18	1.14	0.99	0.99	0.98	1.01	0.98
(2,16,2)	0.00	0.00	0.40	1.00	1.48	1.18	1.02	1.01	1.05	1.01	1.00
(2,16,2)	0.00	0.00	0.60	1.00	2.36	1.20	1.02	1.01	1.07	1.03	1.02
(2,16,2)	0.00	0.00	0.80	1.00	4.24	1.35	1.04	1.00	1.11	1.00	1.01
(2,16,2)	0.00	0.20	0.20	1.00	1.65	1.20	1.06	1.03	1.11	1.04	1.05
(2,16,2)	0.00	0.20	0.40	1.00	2.11	1.00	0.99	0.98	1.07	0.99	0.97
(2,16,2)	0.00	0.20	0.60	1.00	4.12	1.15	1.01	1.02	1.20	0.99	0.99
(2,16,2)	0.00	0.40	0.40	1.00	5.27	1.15	1.06	1.02	1.28	1.01	1.00
(2,16,2)	0.20	0.20	0.20	1.00	2.02	1.13	0.98	0.95	1.10	0.99	0.98
(2,16,2)	0.20	0.40	0.20	1.00	4.30	1.03	1.07	1.04	1.38	1.01	1.02

Table 1: Regression Coefficient Estimates. The Table reports the mean squared error of the linear parameters, β, achieved by each estimation technique relative to the mean squared error achieved by GLS evaluated at the true parameter values. The Monte-Carlo experiment uses 200 repetitions of each estimation procedure for each value of the parameters and each data unbalancedness pattern.

				$MSE(\hat{\sigma}_0^2)/MSE^{MLE}$				$MSE(\hat{\sigma}_1^2)/MSE^{MLE}$				$MSE(\hat{\sigma}_2^2)/MSE^{MLE}$				$MSE(\hat{\sigma}_3^2)/MSE^{MLE}$			
Pattern	σ_1^2	σ_2^2	σ_3^2	WK	MV_1	MV_2	MV_3	WK	MV_1	MV_2	MV_3	WK	MV_1	MV_2	MV_3	WK	MV_1	MV_2	MV_3
(4,4,4)	0	0	0	1.1	1.1	1.1	1.1	1.3	1.3	1.6	1.2	1.4	1.4	1.5	1.3	1.2	1.2	1.4	1.1
(4,4,4)	0	0	0.2	1.1	1.9	1.1	1.1	3.7	4.2	1.1	1.3	3.1	3.5	1.1	1.2	1.0	1.0	1.2	0.9
(4,4,4)	0	0	0.4	1.2	3.8	1.2	1.1	22.5	26.0	2.0	2.0	15.4	17.7	1.5	1.5	1.1	1.0	1.3	1.0
(4,4,4)	0	0	0.6	1.1	16.3	1.1	1.1	108.1	121.7	6.4	3.0	154.5	175.2	8.9	4.4	1.1	1.1	1.3	1.0
(4,4,4)	0	0	0.8	1.1	62.7	1.1	1.1	1617.3	1866.3	73.5	10.5	1418.4	1634.4	64.4	9.4	1.0	1.0	1.1	0.9
(4,4,4)	0	0.2	0.2	1.1	4.7	1.0	1.0	25.2	28.6	2.3	1.6	1.1	1.1	1.0	0.9	1.2	1.2	0.9	0.9
(4,4,4)	0	0.2	0.4	1.1	12.1	1.1	1.1	93.1	107.2	5.2	2.6	1.4	1.5	0.9	0.9	1.0	1.0	1.0	0.9
(4,4,4)	0	0.2	0.6	0.4	19.5	0.3	0.3	483.4	554.0	22.8	4.3	1.4	1.5	0.8	0.8	1.0	1.0	1.3	1.0
(4,4,4)	0	0.4	0.4	1.0	54.4	1.0	1.0	708.7	818.7	35.2	7.2	1.1	1.1	1.0	0.9	1.1	1.1	1.0	0.9
(4,4,4)	0.2	0.2	0.2	0.9	11.4	0.9	0.9	1.3	1.3	0.9	0.8	1.3	1.3	1.0	0.8	1.3	1.4	0.9	0.8
(4,4,4)	0.2	0.4	0.2	0.4	14.3	0.4	0.4	1.7	1.8	0.9	0.9	1.1	1.1	1.1	0.9	1.7	1.8	0.9	0.9
(4,8,2)	0	0	0	1.1	1.0	1.1	1.0	1.4	0.8	2.9	1.3	1.2	0.8	2.1	1.2	1.3	0.9	2.6	1.2
(4,8,2)	0	0	0.2	1.1	2.1	1.1	1.1	8.5	7.2	2.3	1.6	3.8	4.0	1.4	1.1	1.1	1.9	1.2	1.0
(4,8,2)	0	0	0.4	1.1	6.2	1.1	1.1	20.2	16.6	2.2	1.4	15.6	30.5	1.5	1.1	1.1	2.1	1.2	1.0
(4,8,2)	0	0	0.6	1.2	20.2	1.1	1.1	163.0	135.9	5.9	2.9	319.5	438.9	8.6	3.6	1.3	2.8	1.1	1.0
(4,8,2)	0	0	0.8	1.1	97.6	1.1	1.1	764.7	546.0	16.9	3.8	732.2	815.2	15.8	4.4	1.0	1.8	1.2	1.0
(4,8,2)	0	0.2	0.2	1.1	5.4	1.0	1.1	15.9	16.9	2.5	1.6	1.2	2.4	1.0	0.9	1.3	2.5	1.0	0.9
(4,8,2)	0	0.2	0.4	1.1	14.3	1.1	1.1	61.2	85.3	3.6	1.8	1.8	3.7	0.9	1.0	1.2	2.3	1.1	1.0
(4,8,2)	0	0.2	0.6	0.6	34.3	0.6	0.5	1220.6	1141.0	23.7	5.0	2.1	4.4	0.8	0.8	1.3	3.1	1.1	0.9
(4,8,2)	0	0.4	0.4	1.1	70.3	1.0	1.0	1066.7	2124.2	36.3	9.8	1.3	2.2	1.0	0.9	1.6	3.3	1.0	0.9
(4,8,2)	0.2	0.2	0.2	0.9	11.9	0.9	0.9	1.5	3.0	0.9	0.9	1.4	2.8	1.0	0.9	1.5	2.9	1.0	0.9
(4,8,2)	0.2	0.4	0.2	0.5	18.2	0.4	0.5	2.2	3.2	0.8	0.8	1.3	3.1	1.0	0.9	2.2	4.1	0.9	0.9
(2,16,2)	0	0	0	1.3	0.9	1.3	1.0	2.4	1.1	7.5	1.9	1.5	0.5	5.3	1.5	1.9	0.5	7.3	1.5
(2,16,2)	0	0	0.2	1.1	1.9	1.1	1.0	2.6	6.7	3.9	1.3	4.0	6.1	5.0	1.2	1.4	2.4	1.3	1.0
(2,16,2)	0	0	0.4	1.1	5.1	1.1	1.0	11.1	33.4	3.5	1.1	10.9	20.5	3.0	1.4	1.4	2.9	1.2	1.0
(2,16,2)	0	0	0.6	1.2	23.0	1.2	1.2	53.5	102.1	3.5	1.9	43.5	115.4	3.8	1.5	1.9	3.8	1.2	1.0
(2,16,2)	0	0	0.8	1.2	76.4	1.2	1.1	975.5	1804.3	15.4	3.6	650.3	1693.0	16.9	4.4	2.3	4.8	1.1	1.0
(2,16,2)	0	0.2	0.2	1.0	4.1	1.0	1.0	7.5	18.2	2.3	1.1	1.4	2.7	1.0	0.9	1.6	3.1	1.1	1.0
(2,16,2)	0	0.2	0.4	1.1	13.0	1.0	1.0	59.2	128.4	4.6	2.4	1.9	4.5	0.9	0.9	1.7	4.1	1.0	1.0
(2,16,2)	0	0.2	0.6	0.4	23.1	0.4	0.5	1367.9	3252.6	29.4	6.9	3.7	7.2	0.7	0.8	2.1	5.0	1.1	1.0
(2,16,2)	0	0.4	0.4	0.6	31.7	0.6	0.6	789.2	899.2	9.1	3.0	2.6	6.5	1.0	1.0	2.8	7.6	1.0	0.9
(2,16,2)	0.2	0.2	0.2	0.8	8.3	0.7	0.9	1.7	3.0	0.9	0.9	1.6	3.1	0.9	0.8	2.3	4.2	0.9	0.9
(2,16,2)	0.2	0.4	0.2	0.3	13.1	0.3	0.6	2.7	5.2	0.7	0.8	2.2	5.0	0.9	1.0	2.1	4.2	0.8	0.8

Table 2: The table reports the ratio of the mean squared error of $\hat{\sigma}_0^2$,$\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\sigma}_3^2$ from each estimation method relative to the mean squared error provided by MLE. All methods except MV1 perform relatively well, but WK, MV2, and MV3 can perform better than MLE when most of the variance in the residual is coming from the error components rather than the idiosyncratic component of the error term. The Monte-Carlo experiment uses 200 repetitions of each estimation procedure for each value of the parameters and each data unbalancedness pattern.

In contrast, the estimators each generally perform well whenever the i^{th} variance component is truly non-zero. This is most clearly seen in the final four columns of the table since $\sigma_3^2 \neq 0$ in most of the table entries. Evidently, there are some efficiency losses due to WK and MV_1 since Table 2 shows that MV_2 and MV_3 estimators can benefit from as much as a twenty to thirty percent reduction in MSE relative to MLE, even when some of the random components are zero, *provided that the true value of σ_i^2 is non-zero.* However, these potential gains are very small relative to the potential for huge MSE losses incurred by not using MLE when the true σ_i^2 is in fact zero.

Qualitatively, these results are similar to those provided by Baltagi, Song, and Jung (1999a) and indicate substantial gains from the computationally more intensive MLE estimator. Quantitatively however, the results are much *less* favorable for the WK estimator when the true value of the variance component is in fact zero; WK can have a MSE of more than a *thousand* times that of the MLE. The potential losses in MSE from using WK instead of MLE dwarf the magnitudes of the effects reported in Baltagi, Song, and Jung (1999a).

6 Application: The Demand for Retail Products

Berry (1994) demonstrates that discrete choice random utility models of differentiated product demand may be estimated from the linear equation[15]

$$y_{jht} = x'_{jht}\beta + \psi \ln(s_{fht|g}) + u_{jht} \tag{11}$$

where $y_{jht} = ln(s_{jht}/s_{ot})$, with s_{jht} the observed market share of product j selling at location h on date t, s_{ot} is the fraction of people choosing not to consume any of the products in the market on day t and x_{jht} represents the observed characteristics of the product, location and day. Berry (1994) demonstrates that if $\psi = 0$, the model corresponds

[15]The reader is directed to Davis (2000) and the references provided therein for a full discussion of this class of differentiated product demand models.

to the multinomial logit model, while if $\psi \neq 0$, this corresponds to the nested multinomial logit model where $s_{jht|g}$ is the observed market share of jht given that one of the products in nest $g = 1, .., G$ is purchased. In either case, the error term u_{jht} represents the unobserved (to the econometrician, but observed by the consumer) quality of the product and location on that day.

I estimate the demand model using data from a retail market, the motion picture exhibition market, in order to determine the between theater cross-price elasticity of demand.[16] The dataset consists of sales, price, and theater characteristics for six movie theaters, located near New Haven CT, from the six week period (17/4/1998 to 5/28/1998.) This period coincides with a major marketing experiment undertaken by the Hoyts theater chain: for three weeks, during the middle of the sample period, Hoyts substantially reduced adult evening admission prices (from \$7.75 to \$5), and subsequently raised the adult admission price to \$8. Several other local movie theaters immediately responded, reducing their prices for the same period. Daily sales data for each film playing in each theater during the period were obtained from *Entertainment Data Inc.* The number of weeks the film has played at a particular theater is also observed, while price and theater characteristics data were obtained separately by surveying individual theaters repeatedly during the six week experiment.[17] Table 3 reports the theaters in the sample, the theater chain which owned the theater, and their pricing strategy in response to Hoyts marketing experiment.[18]

[16]One particularly high profile issue in this market is the appropriate treatment of mergers between theater chains in highly concentrated local markets. For example, in the recent Sony Theaters and Cineplex-Odeon merger, antitrust concern resulted in forced divestiture of individual theaters in New York and Chicago on the presumption that these divestitures would sustain competition in the post-merger market. Obviously, determining appropriate regulatory treatment of proposed mergers requires an understanding of the substitutability of theaters in a market and for that reason determining own- and cross-price elasticities of demand in this market is a useful activity. Since admission prices rarely vary substantially either across films within a theater, or across theaters, either within or across theater chains, this marketing experiment provides a rare opportunity to observe demand responses to variation in prices in this market.

[17]On average, a film playing in a single location achieved revenues of \$647 each day over the sample, but the data are highly skewed with one film achieving a daily box office of \$13,608 in a single location. The number of weeks the film has played at the theater is also available. The average film plays for only 3-4 weeks, however some films play for an extremely long time with one film in the sample achieving a run of 22 weeks at the same theater.

[18]The pricing response of theaters provides important additional identifying information about the extent of substitution between the movie theaters. No price response by theater A to a price decrease at theater B

	Theater	Chain	Pricing strategy during sample
1	BRANFORD 12	HOYTS	\$5 for three weeks
2	MILFORD QUAD	National Amusements	No price change
3	SHOWCASE ORANGE	National Amusements	\$4.50 for three weeks
4	SHOWCASE 5 MILFORD	National Amusements	No price change
5	SHOWCASE 8 NORTH HAVEN	National Amusements	\$4.50 for three weeks
6	NEW HAVEN	York Square Cinema	No price change

Table 3: Theaters in the sample.

Table 4 reports a selection of the 47 films in the sample and illustrates the major sources of unbalancedness in the dataset. First, not all films are shown in all locations. Second, films start and stop being shown at theaters during the sample period. This pattern is described in the table entries which are the number of days a film played at the theater during the sample period. In addition some revenue data are missing because of data reporting failures.[19] The Ahrens-Pincus unbalancedness statistics are $(r_{time}, r_{theater}, r_{films}) = (0.97, 0.72, 0.44)$, indicating that the data is approximately balanced in terms of number of observations for each day of data. In contrast, the data are substantially more unbalanced across theater locations and extremely unbalanced across films. Evidently the Branford 12 theater is much larger than the Milford 4, hence we tend to see many more observations from the Branford 12 than from the smaller theaters. Similarly, some films show in more locations or for a longer time period than others.

The estimation results are reported in Table 5. The first five columns report various estimators of the multinomial logit demand equation, while the last three columns report nested

indicates that the owners of theater A expect very little substitution will occur. This is useful information that can be incorporated into the demand model by using it to form groups or nests of products. Specifically, the two National Amusement theaters in Milford did not respond to Hoyts price experiment, therefore they are placed in one nest. In contrast, the National Amusement theaters in Orange and North Haven did change their prices in response to Hoyts's experiment. Thus, the Hoyts Branford 12 and these two National Amusement theaters are placed in a second nest. The York Square art house in downtown New Haven did not respond, so it is given its own nest. Finally, the fourth nest is the outside good.

[19]Each evening, *Entertainment Data* attempt to contact each theater either electronically or via telephone. Despite multiple attempts to contact the theater, inevitably some data are missing. Reasons that the sales data are missing include: the figures weren't ready, only a recording could be reached, no one at the theater answered the phone and so forth.

FILM	THEATER NUMBER						
	1	2	3	4	5	6	Total
AS GOOD AS IT GETS	14	0	7	0	0	21	42
BARNEYS GREAT ADVENTURE	35	0	35	0	0	0	70
BLACK DOG	21	14	0	0	14	0	49
BULWORTH	7	0	7	0	0	0	14
CHINESE BOX	0	0	0	0	0	7	7
CITY OF ANGELS	42	0	42	0	0	0	84
LOST IN SPACE	28	28	0	0	28	0	84
PAULIE	42	42	0	0	28	0	112
THE BIG HIT	28	7	0	21	21	0	77
TITANIC	42	0	21	21	21	0	105
TWO GIRLS AND A GUY	0	0	0	0	0	14	14
WOO	0	14	0	0	7	0	21
Total	588	196	350	119	273	147	1673

Table 4: Data unbalancedness due to films only playing in some locations. This table reports the number of days a selection of films played at each location during the sample. The final row of the table reports total numbers of observations in the sample broken down by theater. Thus, the dataset provides a total of 1673 observations on 47 films that play in at least one theater during the six week data period. The regression results use 1478 observations because sales figures are sometimes not observed as a result of electronic reporting failures.

multinomial logit results. The first column reports the naive, inefficient, OLS estimator of the random effects model.[20] This naive estimator ignores the correlation structure in the errors, does not use all of the information in the sample optimally[21] and nicely illustrates the problems that result: the parameters of particular interest for the calculation of price elasticities are not well determined. The second column reports the estimates that would be obtained using the within regression.[22] The point estimates of the linear parameters obtained from the WK,MV_1,MV_2, and[23] MV_3 estimators are similar to each other, but all perform poorly for correct inference about the price coefficients relative to MLE or the fixed

[20]Irrespective of the correlation structure, provided $E[u(\beta_0)|Z,\Delta]=0$, a standard instrumental variables estimator, $\beta^{inefficient}=(X'P_{[Z]}X)^{-}X'P_{[Z]}y$, provides a consistent estimator of the linear parameters. In this specification, all variables are assumed exogenous.

[21]Consistent estimates of the standard errors are reported, using the MLE estimates of the error components.

[22]These results are actually calculated using OLS on the full fixed effects model. To ensure comparability across the specifications, the R-squared reported in this column is the proportion of variance explained by the full model relative to a model with just film, time, and theater fixed effects.

[23]The estimates for MV_1 and MV_2 are not reported due to space limitations. The estimates of the linear parameters are quantitatively virtually identical to those obtained by MV_3. The vector of variance component estimates for MV_1 and MV_2 are $(0.092, 0.041, 0.407, 0.673)$ and $(0.499, 0.101, 1.665, 0.324)$ respectively.

effects estimator. In contrast, the estimates from OLS(2) and MLE are very similar except that the coefficient on $p * th3$ is larger and statistically significant in the MLE results. In this dataset, modeling the error components results in substantial increases in the precision with which the linear parameters are estimated relative to the naive estimator.

Param.	OLS(1)	OLS(2)	WK	MV3	MLE	WTN	2SLS	WK
const	−8.91** (0.92)	−6.99** (0.32)	−7.91** (0.45)	−8.00** (0.45)	−7.98** (0.40)	. .	−5.18** (0.37)	−5.81** (0.55)
week	−0.21 (0.20)	0.01 (0.03)	−0.04 (0.03)	−0.01 (0.03)	−0.01 (0.03)	0.00 (0.03)	0.00 (0.03)	−0.01 (0.01)
week2	0.01 (0.01)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)	−0.00 (0.00)
sat	0.41 (0.43)	. .	0.43 (0.51)	0.43 (0.44)	0.43 (0.42)	. .	. .	0.40 (0.49)
sun	0.09 (0.43)	. .	0.12 (0.51)	0.12 (0.44)	0.12 (0.42)	. .	. .	0.05 (0.49)
mon	−1.06* (0.43)	. .	−1.11* (0.51)	−1.11* (0.44)	−1.11** (0.42)	. .	. .	−1.14* (0.49)
tue	−1.39** (0.43)	. .	−1.50** (0.51)	−1.50** (0.44)	−1.51** (0.42)	. .	. .	−1.52** (0.49)
wed	−1.32** (0.43)	. .	−1.51** (0.51)	−1.51** (0.44)	−1.52** (0.42)	. .	. .	−1.38** (0.49)
thu	−1.33** (0.43)	. .	−1.49** (0.51)	−1.50** (0.44)	−1.50** (0.42)	. .	. .	−1.40** (0.49)
prices	−0.03 (0.11)	−0.23** (0.03)	−0.23** (0.03)	−0.23** (0.03)	−0.26** (0.02)	−0.16** (0.03)	−0.16** (0.03)	−0.16** (0.01)
p*th2	−0.04 (0.05)	0.04** (0.01)	0.01 (0.05)	0.01 (0.05)	0.04** (0.01)	−0.10 (0.40)	−0.11** (0.02)	−0.11 (0.10)
p*th3	0.08 (0.05)	0.01 (0.03)	0.02 (0.03)	0.02 (0.03)	0.10** (0.01)	0.01 (0.02)	0.01 (0.02)	0.01 (0.01)
p*th4	0.12 (0.07)	0.08** (0.01)	0.06 (0.05)	0.05 (0.05)	0.08** (0.01)	−0.07 (0.45)	−0.09** (0.03)	−0.09 (0.10)
p*th5	0.10* (0.04)	0.07* (0.03)	0.07** (0.03)	0.07** (0.03)	0.08** (0.01)	0.04 (0.03)	0.04 (0.03)	0.04** (0.01)
p*th6	−0.01 (0.08)	0.13** (0.03)	0.03 (0.06)	0.06 (0.07)	0.09** (0.03)	−0.27 (2.04)	−0.23** (0.06)	−0.24* (0.12)
$\ln(sh_{fht\|g})$	. .	. .	. .	. .	. .	0.66** (0.09)	0.66** (0.09)	0.64** (0.05)
R^2	0.39	0.27	0.25	0.22	0.22	0.10	0.27	0.39
Num.	1478.00	1478.00	1478.00	1478.00	1478.00	1478.00	1478.00	1478.00
	Estimated Variance Components							
σ^2_{time}			0.777	0.568	0.521** (0.29)			0.731
$\sigma^2_{theater}$			0.093	0.102	0.000 (0.04)			0.426
σ^2_{film}			0.601	1.853	1.952** (0.41)			0.211
σ^2_{ν}			0.313	0.319	0.317** (0.01)			0.054

Table 5: ** and * denote statistically different from zero at the 1% and 5% level of significance respectively. All standard errors are computed assuming that the true data generating process corresponds to the three-way error components model. The variable "Price" is the adult evening admission price to the theater. "Weeks at theater" corresponds to the number of weeks the film had played at that particular theater. $p * th2$, say, denotes an interaction between price and a dummy variable indicating that the observation is from theater two. Day of the week fixed effects are included in each of the specifications but are collinear with time fixed effects included in the full fixed effects specification reported in OLS(2), in the 2SLS specification and are annhialated by the within transformation so do not appear in those columns. The variance components modelled are: (i) A time random effect. This effect allows correlation in the error term across theaters and films within the time period that the data were observed. (ii) A "theater" random effect. This effect allows correlation in the errors across time and films for a given theater. (iii) A "film" random effect. This final effect allows errors to be correlated across theaters and time for a given film. The sixth column reports regression estimates from two stage least squares estimation of the within regression. The seventh column reports two stage least squares estimation of the full regression equation allowing for film, theater, and time fixed effects. The final column reports instrumental variable WK estimates of the random efffects model. In the last three sets of estimates, $\ln(sh_{fht\|g})$ is instrumented using the set of films playing in *other* theaters in the same group as theater h. All R^2 reported are the appropriate generalized R^2 for instrumental variable regressions, while for conformity, the 2SLS column reports the proportion of variance explained relative to a model which includes only fixed film, theater, and time effects.

The estimates from the full model are also useful because they determine the correlation structure in the residuals and therefore provide relative contributions that each source of variance makes toward the total variance of the error term. While MV_2, MV_3 and MLE each provide similar estimates of the variance components, the WK and MV_1 estimates are quite different. This is precisely as the Monte Carlo simulations predicted; the true value of the theater random effect is evidently close to zero in the non-nested multinomial logit specifications.

Instrumental variable estimates of the Nested multinomial logit are reported in the final three columns of the table.[24] First the within equation, then two stage least squares performed with film theater and time fixed effects and finally WK estimates are provided.[25] The general pattern of the estimates is similar to those found for the multinomial logit model. However, the additional parameter, ψ, representing the level of correlation between utilities within a group is also statistically highly significant in both specifications.[26]

Table 6 reports the estimated elasticities of demand between theaters in the sample for the nested multinomial model with three-way error components. Notice that the estimated cross-price elasticity between theaters not in the same group is essentially zero. Within group elasticities are always of the right sign and several are large. The predicted theater cross elasticities are asymmetric. For example, an increase in price at Hoyts 12 screen theater in Branford is found to have a significantly bigger impact on demand at National

[24]Berry (1994) argues that $\ln(s_{jht|g})$ will likely be high when the unobserved quality of product jht is high, hence an instrument for $\ln(s_{jht|g})$ must be constructed. The films that are playing at other theaters within the same nest will be correlated with $s_{jht|g}$, but are assumed uncorrelated with the unobserved quality of the film playing at that theater, u_{jht}. Consequently, they provide valid instruments to identify ψ. This identification assumption is standard in the differentiated product demand literature, but is certainly mildly troubling since there is a market which determines the way films are allocated to theaters. In the absence of obvious alternatives I follow the literature.

[25]Neither MV nor MLE estimators are provided for the nested model since, unfortunately, these estimators are not easily generalized to allow for instrumental variable estimation, although the former could probably be with some considerable effort.

[26]A Hausman misspecification test of the WK estimates relative to the within estimator is easily accepted indicating the random effects specification is not misspecified. The within estimator is consistent whether the effects are random or fixed, but inefficient if they are random. In contrast, the WK estimator is efficient if the effects are random but inconsistent if they are fixed. The Hausman statistic yields a value of 0.4, but is distributed as a $\chi^2(9)$ under the null hypothesis. Thus the null of correct specification cannot be rejected.

	1	2	3	4	5	6
1	-2.78	0.002	0.683	0.003	0.545	0.004
2	0.005	-4.11	0.008	1.597	0.002	0.004
3	1.094	0.002	-2.26	0.003	0.823	0.004
4	0.005	2.033	0.008	-3.53	0.002	0.004
5	0.451	0.002	0.425	0.003	-2.30	0.004
6	0.004	0.002	0.006	0.002	0.002	-2.61

Table 6: Estimates of theater own and cross price elasticities. Diagonal elements correspond to estimated own theater price elasticities of demand, while non-diagonal elements are theater cross-price elasticities of demand. The element $(h.k)$ in the table reports the elasticity of demand at theater h in response to a change in price at theater k. The table is not symmetric since these are theater cross price elasticities of demand and the size of theaters and movies playing at each theater differs across theaters.

Amusement's showcase Orange than vice-versa. This is intuitive since the latter is only an 8 screen theater. The estimates show that the art house theater is essentially a monopoly but one whose own-price elasticity is substantial. The theater facing the highest price elasticity of demand is the National Amusement Milford Quad, which would see a 4% decline in demand in response to a one percent increase in the admission price. The sister theater in Milford, also owned by National Amusements, would see a 2% rise in demand. Consistent with the observed price responses to Hoyts experiment, the estimates reveal little evidence of substitution between the groups of theaters, but considerable evidence of substitution within each of the groups.

7 Conclusions

In this paper I show that the econometric techniques required to construct efficient estimators allowing for instrumental variable estimation of multi-way error components models in unbalanced data structures involves simple and elegant generalizations of the panel data methods developed by Wansbeek and Kapteyn (1989). By presenting the results in terms of recurrence relations, feasible estimators for models with an arbitrary number of error components can be easily developed. In addition, the approach allows the analysis of the standard- and nested error component models to be unified, in contrast to their treatment to date. In fact, the inverse provided for the variance-covariance matrix is useful quite generally for estimating regression models with complex error structures. Where possible, I also generalize the results available from the literature by allowing for instrumental variable estimation of the linear parameters.

Monte Carlo simulation results for the three-way model yield three main conclusions. First, the mean-squared error of the linear parameters does not differ dramatically across the different error component estimation methods. Second, MLE performs considerably better than all the other suggested methods whenever the model is misspecified in the sense that at least one variance component is truthfully zero. Finally, when all variance components are truly non-zero, the various ANOVA methods appear to perform at least as well as the more computationally burdensome MLE.

I estimate a three-way error components version of the nested- and non-nested multinomial logit model of demand for differentiated products with unobserved product characteristic using the formulation developed in Berry (1994). I use a new dataset constructed during a major marketing experiment in the motion picture exhibition market to estimate the parameters of the model and to provide estimates of the cross price elasticity of demand between a group of geographically differentiated movie theaters. The estimates reveal a complex set of asymmetric cross-theater price elasticities of demand. These estimates are particularly useful for strategic decision making by the theater chains, for analysis of the impact of mergers on prices and for determining the appropriate extent of geographic market definition in these markets. The extent of appropriate geographic market definition is strikingly evident from Table 6. Further evaluation of these issues is beyond the scope of this paper, but several are explored further using a related dataset in Davis (1998).

Appendix A: Proofs

Proof of lemma 1

Partition the matrix X, so that $X = (X_1, X_2)$. Searle (1982) (page 263) shows that a generalized inverse of the matrix $X'X$ is given by

$$(X'X)^- = \begin{bmatrix} (X_1'X_1)^- & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -(X_1'X_1)^- X_1'X_2 \\ I \end{bmatrix} (X_2'Q_{[X_1]}X_2)^- \begin{bmatrix} -X_2'X_1(X_1'X_1)^- & I \end{bmatrix}$$

The lemma follows immediately by pre-multiplying this expression by X and post-multiplying by X'. **Q.E.D.**

Proof of lemma 2

If A is nxm and B is mxn, then $(I_n + AB)^{-1} = I_n - A(I_m + BA)^{-1}B$ provided $|I_m + AB| \neq 0$. Thus, $(I_n + XX')^{-1} = I_n - X(I + X'X)^{-1}X'$ (see for example Berk and Sydsæ ter (1993), p.100.) Applying the partitioned inverse formula for matrices yields

$$(I+X'X)^{-1} = \begin{pmatrix} S^{-1}, & -S^{-1}X_1'X_2(I + X_2'X_2)^{-1} \\ -(I + X_2'X_2)^{-1}X_2'X_1S^{-1}, & (I + X_2'X_2)^{-1}[I + X_2'X_1S^{-1}X_1'X_2(I + X_2'X_2)^{-1}] \end{pmatrix}$$

where $S = I + X_1'X_1 - X_1'X_2(I + X_2'X_2)^{-1}X_2'X_1 = I + X_1'\tilde{Q}_{[X_2]}X_1$. Consequently,

$$\begin{aligned} X[I + X'X]^{-1}X' &= X_1S^{-1}X_1'\tilde{Q}_{[X_2]} - \tilde{P}_{[X_2]}X_1S^{-1}X_1' + \tilde{P}_{[X_2]} + \tilde{P}_{[X_2]}X_1S^{-1}X_1'\tilde{P}_{[X_2]} \\ &= \tilde{P}_{[X_2]} + \tilde{Q}_{[X_2]}X_1S^{-1}X_1'\tilde{Q}_{[X_2]} \end{aligned}$$

and hence $(I + XX')^{-1} = \tilde{Q}_{[X_2]} - \tilde{Q}_{[X_2]}X_1S^{-1}X_1'\tilde{Q}_{[X_2]}$. **Q.E.D.**

Appendix B: Results for the four-way model

In the four-way model, $u = \Delta_1\mu + \Delta_2\lambda + \Delta_3\gamma + \Delta_4\delta + \nu$ and $\Omega = E[uu'] = \sigma_0^2 I_n + \sum_{s=1}^4 \sigma_s^2 \Delta_s \Delta_s'$. In this appendix I present three sets of results. Proofs are simple applications of the results provided in the paper and are consequently left to the reader.

Corollary 3 (Within Transformation) *Let* $Y = (Y_1, Y_2, Y_3, Y_4)$, *then* $P_{[Y]} = P_{[A]} + P_{[B]} + P_{[C]} + P_{[D]}$ *where* $A = Y_4$, $B = Q_{[A]}Y_3$, $C = Q_{[B]}Q_{[A]}Y_2$ *and* $D = Q_{[C]}Q_{[B]}Q_{[A]}Y_1$.

Corollary 4 (Expression for Ω^{-1}) *If* $Y = (Y_1, Y_2, Y_3, Y_4)$ *and* $\Omega = \sigma_0^2(I + YY')$, *then* $\Omega^{-1} = \sigma_0^{-2}(I + YY')^{-1}$ *where* $(I + YY')^{-1} = \tilde{Q}_{[Y_2,Y_3,Y_4]} - \tilde{Q}_{[Y_2,Y_3,Y_4]}Y_1 S_1^{-1} Y_1' \tilde{Q}_{[Y_2,Y_3,Y_4]}$, $\tilde{Q}_{[Y_2,Y_3,Y_4]} = \tilde{Q}_{[Y_3,Y_4]} - \tilde{Q}_{[Y_3,Y_4]}Y_2 S_2^{-1} Y_2' \tilde{Q}_{[Y_3,Y_4]}$, $\tilde{Q}_{[Y_3,Y_4]} = \tilde{Q}_{[Y_4]} - \tilde{Q}_{[Y_4]}Y_3 S_3^{-1} Y_3' \tilde{Q}_{[Y_4]}$, $\tilde{Q}_{[Y_4]} = I - Y_4(I + Y_4'Y_4)^{-1}Y_4'$, $S_1 = I + Y_1'\tilde{Q}_{[Y_2,Y_3,Y_4]}Y_1$, $S_2 = I + Y_2'\tilde{Q}_{[Y_3,Y_4]}Y_2$ *and* $S_3 = I + Y_3'\tilde{Q}_{[Y_4]}Y_3$.

Lemma 4 (QUE's of the Variance Components) *If* $e = y - X\hat{\beta}^{W2SLS}$, $q_n = e'Q_{[\Delta]}e$ *and* $q_i = e'P_{[\Delta_i]}e$, *for* $i = 1, 2, 3, 4$. *Then,*

$$
\begin{aligned}
E[q_n] &= (n - t_1 - t_2 - t_3 - t_4 + k_n)\sigma_0^2 \\
E[q_1] &= (N_1 + k_1)\sigma_0^2 + n\sigma_1^2 + k_{12}\sigma_2^2 + k_{13}\sigma_3^2 + k_{14}\sigma_4^2 \\
E[q_2] &= (N_2 + k_2)\sigma_0^2 + k_{21}\sigma_1^2 + n\sigma_2^2 + k_{23}\sigma_3^2 + k_{24}\sigma_4^2 \\
E[q_3] &= (N_3 + k_3)\sigma_0^2 + k_{31}\sigma_1^2 + k_{32}\sigma_2^2 + n\sigma_3^2 + k_{34}\sigma_4^2 \\
E[q_4] &= (N_4 + k_4)\sigma_0^2 + k_{41}\sigma_1^2 + k_{42}\sigma_2^2 + k_{43}\sigma_3^2 + n\sigma_4^2
\end{aligned}
$$

where $k_n = tr\{(X'P_{[QZ]}X)^{-1}X'Q_{[\Delta]}X\}$, $k_i = tr\{(X'P_{[QZ]}X)^{-1}X'P_{[\Delta_i]}X\}$, $k_{ij} = tr\{P_{[\Delta_i]}\Delta_j\Delta_j',\}$ *for* $i, j = 1, 2, 3, 4$, $t_1 = rank(A), t_2 = rank(B), t_3 = rank(C)$ *and* $t_4 = rank(D)$ *where* A, B, C, D *are defined in corollary 3.*

Appendix C: Some Useful Matrix Algebra

To calculate the MLE and MINQUE estimators, we would like to evaluate the matrix expressions developed above without ever explicitly calculating any $(n\mathrm{x}n)$ matrices. First, consider the matrix calculation $A\Omega^{-1}B$ for an arbitrary pair of matrices A and B, where A is $(a\mathrm{x}n)$ and B is $(n\mathrm{x}b)$. Applying the results and notation from the previous sections, in particular $X_1 = Y_1$, $X_2 = (Y_2, Y_3)$, we may recursively write

$$\begin{aligned}
A\tilde{Q}_{[Y_3]}B &= AB - (AY_3)(I + Y_3'Y_3)^{-1}(Y_3'B), \text{ while} \\
A\tilde{Q}_{[X_2]}B &= A\tilde{Q}_{[Y_3]}B - (A\tilde{Q}_{[Y_3]}Y_2)S_2^{-1}(Y_2'\tilde{Q}_{[Y_3]}B), \text{ and therefore} \\
\sigma_0^2 A\Omega^{-1}B &= A\tilde{Q}_{[X_2]}B - (A\tilde{Q}_{[X_2]}X_1)S_1^{-1}(X_1'\tilde{Q}_{[X_2]}B).
\end{aligned}$$

Since $A\tilde{Q}_{[Y_3]}B$ and $A\tilde{Q}_{[X_2]}B$ are $(a\mathrm{x}b)$ matrices that may be constructed without calculating any $(n\mathrm{x}n)$ matrices directly, so can any matrix of the form $A\Omega^{-1}B$. This approach can also be used for the other expressions we require for the maximum likelihood or MINQUE estimators, such as $tr\{\Omega^{-1}\}$, $e'\Omega^{-2}e$, $tr\{\Omega^{-2}\}$, $tr\{\Delta_j'\Omega^{-2}\Delta_j\}$, $\sigma_0^6 tr\{\Omega^{-3}\}$ and $A\Omega^{-3}B$. Let $\tilde{Q}_3 \equiv \tilde{Q}_{[Y_3]}$, $\tilde{Q}_2 \equiv \tilde{Q}_{[X_2]}$ and $\phi_1 = X_1 S_1^{-1} X_1'$. Then

$$\begin{aligned}
\sigma_0^4 A\Omega^{-2}B &= A(I + XX')^{-2}B = A(\tilde{Q}_2 - \tilde{Q}_2\phi_1\tilde{Q}_2)^2 B \\
&= A\tilde{Q}_2^2 B - A\tilde{Q}_2^2\phi_1\tilde{Q}_2 B - A\tilde{Q}_2\phi_1\tilde{Q}_2^2 B + A\tilde{Q}_2\phi_1\tilde{Q}_2^2\phi_1\tilde{Q}_2 B, \quad \text{while} \\
\sigma_0^6 A\Omega^{-3}B &= A(I + XX')^{-3}B = A(\tilde{Q}_2 - \tilde{Q}_2 X_1 S_1^{-1} X_1'\tilde{Q}_2)^3 B \\
&= A\tilde{Q}_2^3 B - A\tilde{Q}_2^2\phi_1\tilde{Q}_2^2 B - A\tilde{Q}_2\phi_1\tilde{Q}_2^3 B + A\tilde{Q}_2\phi_1\tilde{Q}_2^2\phi_1\tilde{Q}_2^2 B - A\tilde{Q}_2^3\phi_1\tilde{Q}_2 B \\
&\quad + A\tilde{Q}_2^2\phi_1\tilde{Q}_2^2\phi_1\tilde{Q}_2 B + A\tilde{Q}_2\phi_1\tilde{Q}_2^3\phi_1\tilde{Q}_2 B - A\tilde{Q}_2\phi_1\tilde{Q}_2^2\phi_1\tilde{Q}_2^2\phi_1\tilde{Q}_2 B
\end{aligned}$$

First note that any matrix with one of A, B, or ϕ_1 on either side is not a large $n\mathrm{x}n$ matrix and will be amenable to the bracketing technique. For instance, $A\tilde{Q}_2\phi_1\tilde{Q}_2^3\phi_1\tilde{Q}_2 B = (A\tilde{Q}_2 X_1)S_1^{-1}(X_1'\tilde{Q}_2^3 X_1)S_1^{-1}(X_1'\tilde{Q}_2 B)$.

Second, note that $A\tilde{Q}_2^m B = A\tilde{Q}_2^{m-1}B - (A\tilde{Q}_2^{m-1}X_2)(I + X_2'X_2)^{-1}(X_2'B)$ for all integers,

$m \geq 0$ when $\tilde{Q}_2^0 \equiv I_n$. Then, since $A\tilde{Q}_2^1 B$ can be evaluated without explicitly evaluating any nxn matrices, by induction so can $A\tilde{Q}_2^m B$ for all integers $m \geq 1$ and hence each term in the expressions for $A\Omega^{-2}B$ and $A\Omega^{-3}B$ can be calculated numerically without ever evaluating any nxn matrices explicitly.

The trace expressions required are computable using the identical technique, viz., $\sigma_0^2 tr\{\Omega^{-1}\} = tr\{\tilde{Q}_2\} - tr\{S_1^{-1}X_1'\tilde{Q}_2^2 X_1\}$ where $tr\{\tilde{Q}_2\} = n - tr\{(I + X_2'X_2)^{-1}X_2'X_2\}$.

These calculations do require matrices of the form $(I_{dim\{X_2\}} + X_2'X_2)^{-1}$ to be numerically invertible, or else diagonal. However, if the number of columns of X_2 remains too large for this inverse to be computable, the inversion formula may easily be applied a second time to each expression involving $\tilde{Q}_2$. Specifically, $\tilde{Q}_2^s = \left(\tilde{Q}_3 - \tilde{Q}_3 Y_2 S_2^{-1} Y_2' \tilde{Q}_3\right)^s$ can be expanded by direct multiplication for any integer required $s = 1, 2, 3, 4$ (or indeed beyond.) Each term in the expansion has the form $\tilde{Q}_3^m Y_2 S_2^{-1} Y_2' \tilde{Q}_3^2 Y_2 S_2^{-1} Y_2' \tilde{Q}_3^2 Y_2 S_2^{-1} Y_2' .... Y_2' \tilde{Q}_3^2 Y_2 S_2^{-1} Y_2' \tilde{Q}_3^r$ for some integers $r, m \geq 1$. Then any matrix of the form $A\tilde{Q}_2^s B$ can be written as the sum of products of matrices of the form $A\tilde{Q}_3^m Y_2$, $Y_2'\tilde{Q}_3^2 Y_2$ and $Y_2'\tilde{Q}_3^r B$ which may each be calculated without explicitly evaluating any nxn matrices and requiring only numerical or analytic inversion of a matrix of the form $(I_{dim\{Y_3\}} + Y_3'Y_3)^{-1}$ which has dimension of the columns of Y_3.

Appendix D: MINQUE

Construction of the matrix $S = \{s_{i+1,j+1}\} = tr\{V_i R V_j R\}$ involves

$$s_{i+1,j+1} = \begin{cases} tr\{RR\} & \text{if } i,j = 0 \\ tr\{\Delta_j' RR \Delta_j\} & \text{if } i = 0 \text{ and } j = 1,..,3 \\ tr\{\Delta_i' R \Delta_j \Delta_j' R \Delta_i\} & \text{if } i,j = 1,..,3 \end{cases} \tag{12}$$

while the vector u requires components $y'RRy$ and $y'R\Delta_i\Delta_i'Ry$ for $i = 1,..,3$.

Given the matrix expressions provided in Appendix C, these matrix calculations are entirely straight forward. Let A and B denote arbitrary $(a\text{x}n)$ and $(n\text{x}b)$ matrices, then $ARB = A\Omega^{-1}B - A\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}B$ and

$$\begin{aligned}
ARRB &= A\Omega^{-2}B - A\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-2}B - A\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-2}B \\
&\quad + A\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-2}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}B, \quad \text{hence} \\
tr\{RR\} &= tr\{\Omega^{-2}\} - 2tr\{(X'\Omega^{-1}X)^{-1}X'\Omega^{-3}X\} \\
&\quad + tr\{(X'\Omega^{-1}X)^{-1}X'\Omega^{-2}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-2}X\}, \quad \text{while} \\
tr\{\Delta_i'RR\Delta_i\} &= tr\{\Delta_i'\Omega^{-2}\Delta_i - \Delta_i'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-2}\Delta_i - \Delta_i'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-2}\Delta_i \\
&\quad + \Delta_i'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-2}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\Delta_i\}
\end{aligned}$$

while $tr\{\Delta_i'R\Delta_j\Delta_j'R\Delta_i\}$ may be directly computed using ARB with $A = \Delta_i'$ and $B = \Delta_j$.

Appendix E: Maximum Likelihood Estimation

An alternative method for obtaining the ML estimators involves directly numerically maximizing the log likelihood function, $L = -\frac{n}{2}ln2\pi + \frac{1}{2}ln|\Omega^{-1}| - \frac{1}{2}u'\Omega^{-1}u$. I have already provided a computationally efficient expression for the term $u'\Omega^{-1}u$. It remains to provide a convenient expression for $|\Omega^{-1}|$ using the well known matrix result that for any $(n\text{x}m)$ matrix A and $(m\text{x}n)$ matrix B, $|I_n + AB| = |I_m + BA|$ (see Berk and Sydsæ ter (1993), p. 95.) Specifically, $|\Omega^{-1}| = \sigma_0^{-2n}|(I + XX')^{-1}|$, so

$$\begin{aligned}
|(I + XX')^{-1}| &= |\tilde{Q}_{[X_2]}|.|I_n - X_1S_1^{-1}X_1'\tilde{Q}_{[X_2]}| \\
&= |\tilde{Q}_{[X_2]}|.|I_{N_1} - S_1^{-1}X_1'\tilde{Q}_{[X_2]}X_1|, \text{ where} \\
|Q_{[X_2]}| &= |(I_n + X_2X_2')^{-1}| = |I_n - X_2(I + X_2'X_2)^{-1}X_2'| \\
&= |I_{dim\{X_2\}} - (I + X_2'X_2)^{-1}X_2'X_2|
\end{aligned}$$

As usual, if $dim\{X_2\}$ remains too large for computational convenience, the same process may once again be repeated, viz.,

$$\begin{aligned}
|\tilde{Q}_{[X_2]}| &= |\tilde{Q}_{[Y_3]}|.|I_n - Y_2 S_2^{-1} Y_2' \tilde{Q}_{[Y_3]}| = |\tilde{Q}_{[Y_3]}|.|I_{dim\{Y_2\}} - S_2^{-1} Y_2' \tilde{Q}_{[Y_3]} Y_2| \text{ where} \\
|\tilde{Q}_{[Y_3]}| &= |I_{dim\{Y_3\}} - (I + Y_3' Y_3)^{-1} Y_3' Y_3|.
\end{aligned}$$

Appendix F: Modified Wallace and Hussain (WH) Estimators

Wallace and Hussain (1969) suggest using the OLS residuals instead of the true disturbances to estimate the variance components in a balanced two-way model. Baltagi and Chang (1994) extend the approach to consider the one-way unbalanced model. More generally, allowing for a two-stage least squares estimator, the approach leads to the quadratic forms

Lemma 5 *If $q_n = \hat{u}_{2SLS}' Q_{[\Delta]} \hat{u}_{2SLS}$ and $q_i = \hat{u}_{2SLS}' P_{[\Delta_i]} \hat{u}_{2SLS}$ for $i = 1, 2, 3$. Then,*

$$\begin{aligned}
E[q_n] &= (n - \sum_{s=1}^{3} t_s - 2tr\{Q_{[\Delta]}\psi\} + tr\{\psi' Q_{[\Delta]}\psi\})\sigma_0^2 + \sum_{j=1}^{3} \sigma_j^2 tr\{\Delta_j' \psi' Q_{[\Delta]} \psi \Delta_j\} \\
E[q_i] &= (N_i - 2tr\{P_{[\Delta_i]}\psi\} + tr\{\psi' P_{[\Delta_i]}\psi\})\sigma_0^2 + \sum_{j=1}^{3} \sigma_j^2 (k_{ij} - 2a_{ij} + b_{ij})
\end{aligned}$$

where $k_{ij} = tr\{P_{[\Delta_i]}\Delta_j \Delta_j'\}$, $a_{ij} = tr\{\Delta_j' P_{[\Delta_i]} \psi \Delta_j\}$, $b_{ij} = tr\{\Delta_j' \psi' P_{[\Delta_i]} \psi \Delta_j\}$, $\psi = X(X' P_{[Z]} X)^{-1} X' P_{[Z]}$, $t_1 = rank(A)$, $t_2 = rank(B)$ and $t_3 = rank(C)$ where A, B, C are defined in corollary 1.

Proof Let $M = I - X(X' P_{[Z]} X)^{-1} X' P_{[Z]}$ so that $\hat{u}_{2SLS} = My = Mu$. Notice that $tr\{Q_{[\Delta]}\} = n - t_1 - t_2 - t_3$, $tr\{P_{[\Delta_i]}\} = N_i$ and

$$\begin{aligned}
E[q_n] &= E[u' M' Q_{[\Delta]} M u] = tr\{M' Q_{[\Delta]} M \Omega\} \\
&= \sigma_0^2 tr\{M' Q_{[\Delta]} M\} + \sum_{j=1}^{3} \sigma_i^2 tr\{\Delta_j' M' Q_{[\Delta]} M \Delta_j\}, \text{ while} \\
E[q_i] &= E[u' M' P_{[\Delta_i]} M u] = tr\{M' P_{[\Delta_i]} M \Omega\} \\
&= \sigma_0^2 tr\{M' P_{[\Delta_i]} M\} + \sum_{j=1}^{3} \sigma_j^2 tr\{\Delta_j' M' P_{[\Delta_i]} M \Delta_j\}, \quad \text{where}
\end{aligned}$$

$$
\begin{aligned}
tr\{M'P_{[\Delta_i]}M\} &= tr\{P_{[\Delta_i]}\} - 2tr\{P_{[\Delta_i]}\psi\} + tr\{\psi' P_{[\Delta_i]}\psi\} \\
tr\{M'Q_{[\Delta]}M\} &= tr\{Q_{[\Delta]}\} - 2tr\{Q_{[\Delta]}\psi\} + tr\{\psi' Q_{[\Delta]}\psi\} \\
tr\{\Delta_j' M' P_{[\Delta_i]} M\Delta_j\} &= k_{ij} - 2a_{ij} + b_{ij} \quad \text{and} \quad tr\{\Delta_j' M' Q_{[\Delta]} M \Delta_j\} = tr\{\Delta_j' \psi' Q_{[\Delta]} \psi \Delta_j\}.
\end{aligned}
$$

References

AHRENS, H., AND R. PINCUS (1981): "On two measures of unbalancedness in a one-way model and their relation to efficiency," *Biometric Journal*, 23, 227–235.

AMEMIYA, T. (1971): "The Estimation of the Variances in a Variance-Covariance Model," *International Economic Review*, 12, 1–13.

AMEMIYA, T., AND T. E. MACURDY (1986): "Instrumental-Variable Estimation of An Error Components Model," *Econometrica*, 54(4), 869–880.

ANTWEILER, W. (1998): "Nested Random Effects Estimation in Unbalanced Panel Data," *Mimeo, University of British Columbia.*

ARELLANO, M., AND S. BOND (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

ARELLANO, M., AND O. BOVER (1995): "Another look at the instrumental variable estimation of error-components models," *Journal of Econometrics*, 68(1), 29–51.

BALESTRA, P., AND M. NERLOVE (1973): "Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas," *Journal of Econometrics*, 1, 17–28.

BALTAGI, B. H. (1981): "Simultaneous Equations with Error Components," *Journal of Econometrics*, 17, 189–200.

BALTAGI, B. H. (1985): "Pooling Cross-Sections with Unequal Time Series Lengths," *Economics Letters*, 18, 133–136.

——— (1987): "On Estimating from a more general time series cum cross-section data structure," *The American Economist.*

——— (1995): *Econometric Analysis of Panel Data.* Wiley, Chichester, UK.

BALTAGI, B. H., AND Y. J. CHANG (1994): "Incomplete Panels: A Comparative Study of Alternative Estimators for the Unbalanced One-Way Error Component Regression Model," *Journal of Econometrics*, 62, 67–89.

BALTAGI, B. H., S. H. SONG, AND B. C. JUNG (1999a): "Further Evidence on the Unbalanced Nested Error Component Regression Model," Discussion paper, Texas A& M University.

——— (1999b): "The Unbalanced Nested Error Component Regression Model," Discussion paper, Texas A& M University.

BERK, P., AND K. SYDSÆ TER (1993): *Economists' Mathematical Manual.* Springer-Verlag, Germany.

Berry, S. (1994): "Estimating Discrete-Choice Models of Product Differentiation," *Rand Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Davis, P. (1998): "Spatial Competition in Retail Markets: Movie Theaters," Mimeo, MIT.

——— (2000): "Empirical Models of Demand for Differentiated Products," *European Economic Review (papers and proceedings)*, 44(4-6), 993–1005.

Fuller, W. A., and G. E. Batese (1974): "Estimation of linear functions with crossed-error structure," *Journal of Econometrics*, 2, 67–78.

Fuller, W. A., and G. E. Battesse (1973): "Transformations for Estimation of Linear Models with Nested Error Structures," *Journal of the American Statistical Association*, 68, 626–632.

Ghosh (1976): "Estimating from a more general time series cum cross-section data structure," *The American Economist*, 20, 15–21.

Hansen, L. (1982): "Large Sample Properties of Generlized Method of Moment Estimators," *Econometrica*, 50, 1029–1954.

Hausman, J., and W. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49(6).

Lancaster, P., and M. Tismenetsky (1985): *The Theory of Matrices*, Computer Science and Applied Mathematics. Academic Press, 2 edn.

Magnus, J. R. (1978): "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Distrurbance Covariance Matrix," *Journal of Econometrics*, pp. 281–312.

Magnus, J. R., and H. Neudecker (1999): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley.

Mazodier, P. (1972): "l'Estimation des modèles à erreurs composées," *Annales de l'INSEE*, 7, 43–72.

Nerlove, A. (1971a): "Further evidence on the estimation of dynamic economic relations from a time series of cros sections," *Econometrica*, 39, 359–382.

Nerlove, M. (1971b): "A note on error components models," *Econometrica*, 39, 383–396.

Ostaszewski, A. (1990): *Advanced Mathematical Methods*. Cambridge University Press, Cambridge, UK.

Pakes, A. (1983): "On Group Effects and Errors in Variables in Aggregation," *Review of Economics and Statistics*, 65(1), 168–173.

RAO, R. C. (1971): "Estimation of Variance and covariance Components-MINQUE Theory," *Journal of Multivariate Analysis*, 1, 257–275.

SEARLE, S. R. (1982): *Matrix Algebra Useful for Statistics.* John Wiley & Sons, New York.

WALLACE, T. D., AND A. HUSSAIN (1969): "The Use of Error Components Models in Combining Cross-Section and Time-Series Data," *Econometrica*, 37, 55–72.

WANSBEEK, T., AND A. KAPTEYN (1989): "Estimation of the Error-Components Model with Incomplete Panels," *Journal of Econometrics*, 41(3), 341–361.