On the maximum likelihood estimation of multivariate regression models containing serially correlated error components

Magnus, J.R.; Woodland, A.D.

ON THE MAXIMUM LIKELIHOOD ESTIMATION OF MULTIVARIATE REGRESSION MODELS CONTAINING SERIALLY CORRELATED ERROR COMPONENTS*

By Jan R. Magnus and Alan D. Woodland[1]

1. INTRODUCTION

Ever since the important paper by Balestra and Nerlove (1966), econometricians and statisticians have made extensive use of error components models, that is, regression models whose error terms are the sum of two or more independent components. Error components models are typically (although not exclusively) used in analyzing panel data, and the sharp increase in availability of panel data in recent years has led to a renewed interest in such models.

In the *three-error components* model, the error structure is

$$(1.1) \qquad u_{it} = c_t + \mu_i + \varepsilon_{it}, \qquad i = 1, \ldots, q; t = 1, \ldots, T,$$

where c_t and μ_i are the time-specific and the cross-section-specific components, and $\{c_t\}$, $\{\mu_i\}$ and $\{\varepsilon_{it}\}$ are mutually independent white noise sequences. More popular (and easier to estimate) is the *two-error components* model where either $\mu_i = 0$ or $c_t = 0$. In this paper, we shall concentrate on the two-error components model with error structure consisting of ε_{it} and time-specific component c_t: the time-specific, two-error components model. Most of the subsequent analysis, however, applies equally well to the cross-section-specific, two-error components model.

Although linear and nonlinear regression models with two- or three-error components have performed well in empirical applications, there are two major areas for generalization within the error components framework. First, there is the need to allow for possible serial correlation in the disturbances, and secondly we wish to be able to handle multivariate error components models where the components are vectors rather than scalars. Let us briefly discuss these two generalizations.

The discussion of serial correlation within the error components framework originates with Lillard and Willis (1978), who used the two-error components model in which $\{\varepsilon_{it}\}$ is assumed to follow a first-order autoregressive AR(1) process with a single autocorrelation coefficient applicable for all i. They estimated the parameters of their model by a two-step procedure. Revankar (1979)

studied the three-error components model (1.1) in which $\{e_t\}$ follows an AR(1) process. He too proposes a two-step estimator. The properties of the maximum likelihood estimator for the Lillard-Willis model were studied by Anderson and Hsiao (1982). Similar models were presented by Lillard and Weiss (1979) and Hause (1980). MaCurdy (1982) generalized the error structure of Lillard and Willis to a more general time-series process for $\{\varepsilon_{it}\}$. Pagano (1974) and Revankar (1980) studied error components models of a slightly different sort where the error structure can be written as

$$(1.2) \qquad\qquad u_t = e_t + \varepsilon_t, \qquad t = 1, \ldots, T,$$

where the first component $\{e_t\}$ follows an AR(p) process and the second (independent) component $\{\varepsilon_t\}$ is white noise. See also King (1986).

A multivariate, two-error, components model was first estimated by Chamberlain and Griliches (1975) using maximum likelihood (ML) techniques. A full treatment of multivariate two error components analysis by maximum likelihood (but without serial correlation) can be found in Magnus (1982).

The aim of this paper is to study the multivariate version of the time-specific, two-error components structure, where the components are vectors rather than scalars, allowing serial correlation in both components. In fact, even more generalization is possible. It suffices to assume that $\{e_t\}$ and $\{\varepsilon_{it}\}$ are independently and normally distributed vectors with zero means such that

$$(1.3) \qquad \mathsf{E}\, e_t e_s' = g_{st}\Gamma \quad \text{and} \quad \mathsf{E}\, \varepsilon_{it}\varepsilon_{js}' = \lambda_{ij}\mu_{st}\Delta,$$

$$s, t = 1, \ldots, T; \; i, j = 1, \ldots, q,$$

where Γ is positive semidefinite and Δ positive definite. The $T \times T$ matrices $G = (g_{st})$ and $M = (\mu_{ij})$ can be taken as the covariance matrices of arbitrary autoregressive moving average (ARMA) processes. We shall show that full ML estimation of regression models (linear or nonlinear) with a time-specific, two-error components structure defined by (1.3) is feasible, even though the dimension of the covariance matrix and the number of parameters can be very large.

The model dealt with in this paper should be of particular interest to analysts of panel data sets in which the number of individuals is relatively small. Such data sets include those consisting of a time series of observations on different industries (as in our empirical example), states, or countries. In this type of panel data set, it is often desirable to include a pure time component in the disturbance to account for shocks common to all industries, states, or countries. If the time component is included as a vector of fixed effects, the number of parameters increases with the number of time series observations. Accordingly, it is more appropriate to include the time component as a random vector.

The plan of the paper is as follows. In Sections 2 and 3, we discuss the multivariate *nonlinear* regression model where the disturbances have a time-specific, two-error components structure and both components follow AR(1) processes. Theorem 1 shows that the likelihood function can be written as a function of matrices of small dimension, allowing full ML estimation of the whole system.

In Section 4 we specialize to the *linear* regression model in which the matrix of

explanatory variables does not depend on i. This situation arises quite often, for example, when the regressors are function of prices only, and each of the q industries faces the same prices. (This is, in fact, the case in our empirical example of Section 5.) Theorem 2 then shows that we can concentrate the likelihood function with respect to the regression parameters, which dramatically facilitates the optimization.

An empirical example is provided in Section 5, where we study the interfuel substitution possibilities in Dutch manufacturing. In this illustrative example, annual data on the cost shares and prices of various fuels in each of six manufacturing industries are used to estimate the model by maximum likelihood. The example shows that it is practical to take account, simultaneously, of contemporaneous correlation between cost shares of fuels in all industries and serial correlation of each component of the disturbances.

Section 6 summarizes our conclusions. Finally, the appendix gives the mathematical background of Theorems 1 and 2, and provides the tools for extending the theory to the more general structure.

2. A MULTIVARIATE ERROR COMPONENTS MODEL WITH SERIAL CORRELATION

Let us consider a set of nonlinear regression equations

$$(2.1) \qquad y_{it} = f_i(X_{it}, \beta_i) + u_{it}, \qquad i = 1, \ldots, q, \qquad t = 1, \ldots, T,$$

where y_{it}, f_i and u_{it} are $p \times 1$ vectors, the nonrandom inputs X_{it} are $p \times l_i$ matrices, and the unknown parameter vectors β_i are $k_i \times 1$ vectors. We shall make the following assumption regarding the disturbances.

ASSUMPTION 2.1. *The* $p \times 1$ *disturbance vectors* $\{u_{it}\}$ *decompose as.*

$$u_{it} = e_t + \varepsilon_{it}, \qquad\qquad i = 1, \ldots, q, \qquad i = 1, \ldots, T$$

where

$$e_t = \rho e_{t-1} + v_t, \qquad\qquad t = 2, \ldots, T,$$

and

$$\varepsilon_{it} = \varkappa\varepsilon_{i,t-1} + \lambda_i^{1/2}\eta_{it}, \qquad i = 1, \ldots, q, \qquad t = 2, \ldots, T,$$

with initial conditions

$$e_1 = (1 - \rho^2)^{-1/2}v_1 \quad and \quad \varepsilon_{i1} = (1 - \varkappa^2)^{-1/2}\lambda_i^{1/2}\eta_{i1}.$$

The vectors $\{v_t\}$ *are i.i.d.* $N(0, \Gamma)$, Γ *positive semidefinite; the vectors* $\{\eta_{it}\}$ *are i.i.d.* $N(0, \Delta)$, Δ *positive definite; the sequences* $\{v_t\}$ *and* $\{\eta_{it}\}$ *are independent; and* $|\varkappa| < 1, |\rho| < 1, \lambda_i > 0$ $(i = 1, \ldots, q)$.

If we think of our data as arising from a combined time-series cross-section of q industries over T years, then the disturbance vectors u_{it} are composed of a vector which is different for each industry, and a common vector, which reflects

macroeconomic disturbances affecting all industries in the same manner. Both vectors can be subject to autocorrelation.

In order to assess the appropriateness and flexibility of Assumption 2.1, let us consider some special cases.

Case (i): $\alpha = \rho = 0$, $\lambda_i = 1$.   The disturbance vector can be written as

$$u_{it} = v_t + \eta_{it}.$$

This is the multivariate version of the well-known, two-error components model, as introduced by Balestra and Nerlove (1966), which suggests itself naturally when combining time-series data with cross-section data.[2] In the present case, the disturbances u_{it} are *vectors* so that the error components are *matrices*, vis à vis Γ and Δ. This multivariate extension of the two-error components model was extensively studied in Magnus (1982).

Case (ii): $\Gamma = 0$, $\lambda_i = 1$.   Here we have

$$\mu_{it} = \alpha u_{i,t-1} + \eta_{it},$$

i.e., first-order vector autocorrelation. Notice that the autocorrelation parameter α is assumed to be the same for every industry; this has an important practical reason, which we shall discuss later in this section. Also notice that we have a single parameter α rather than a $p \times p$ matrix A.[3]

Case (iii): $\Gamma = 0$, $\alpha = 0$.   Now

$$u_{it} = \lambda_i^{1/2}\eta_{it} \simeq N(0, \lambda_i\Delta).$$

This specification is intermediate between a Zellner-type seemingly unrelated regression model for all industries combined [$u_{it} \simeq N(0, \Delta)$] and a Zellner-type seemingly unrelated regression model for each industry separately [$u_{it} \simeq N(0, \Delta_i)$].

Case (iv): $\alpha = 0$.   Then

$$u_{it} = e_t + \lambda_i^{1/2}\eta_{it}, \qquad e_t = \rho e_{t-1} + v_t,$$

i.e., two error components with the common component subject to autocorrelation.

Case (v): $\rho = 0$.   Then

$$u_{it} = v_t + \varepsilon_{it}, \qquad \varepsilon_{it} = \alpha\varepsilon_{i,t-1} + \lambda_i^{1/2}\eta_{it},$$

[2] A linear multiple regression model with covariance matrix as in case (i) was first estimated by Chamberlain and Griliches (1975) using LM techniques.

[3] The reason for this is provided by Berndt and Savin (1975), who show that all diagonal elements of a diagonal matrix of autocorrelation parameters must be equal to be consistent with the adding-up property of share equations.

i.e., two error components with the residual component subject to auto-correlation.

Case (vi): $\alpha = \rho$. Then

$$u_{it} = \alpha u_{i,t-1} + w_{it}, \qquad w_{it} = v_t + \lambda_i^{1/2} \eta_{it}.$$

This is the mirror image of cases (iv) and (v). Cases (iv) and (v) have two error components with one component subject to autocorrelation; case (vi) has auto-correlation with a residual consisting of two components.

From these six cases we conclude that Assumption 2.1 is a flexible assumption to make, and that it allows interesting hypotheses to be tested.

Let us define

$$(2.2) \qquad u_i = (u_{i1}', u_{i2}', \dots, u_{iT}')'; \qquad u = (u_1', u_2', \dots, u_q')'.$$

One may then verify that the covariance matrix takes the form

$$(2.3) \qquad \Omega = E\, uu' = \Lambda \otimes M_\alpha \otimes \Delta + \iota\iota' \otimes M_\rho \otimes \Gamma$$

where

$$M_\alpha = \frac{1}{1-\alpha^2} \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{T-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{T-2} \\ \alpha^2 & \alpha & 1 & \cdots & \alpha^{T-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \alpha^{T-1} & \alpha^{T-2} & \alpha^{T-3} & \cdots & 1 \end{bmatrix},$$

M_ρ is defined similarly (with ρ instead of α), ι is a $q \times 1$ vector of ones, and Λ is the diagonal $q \times q$ matrix with $\lambda_1, \dots, \lambda_q$ on the diagonal. Note that, without a further constraint, only ratios of the λ's can be identified, not the λ's themselves. We therefore normalize the λ's in the following convenient way.

ASSUMPTION 2.2. *The $q \times q$ matrix Λ is normalized by* tr $\Lambda^{-1} = q$.

Since the complete covariance matrix Ω is of the order $pqT \times pqT$, ML esti-mation of model (2.1) under Assumption 2.1 is only feasible if we can express the determinant and inverse of Ω in terms of matrices of lower dimension. The next section shows that this is possible, due to the Kronecker structure of Ω. This Kronecker structure is destroyed if α (or ρ) is allowed to differ over industries, and the expressions for Ω, $|\Omega|$ and Ω^{-1} become cumbersome and impractical in that case.

3. THE LIKELIHOOD FUNCTION

With M_α as defined in (2.4) and M_ρ defined similarly (with ρ replacing α), we can now prove the following results, showing that the likelihood can be written as a function of $p \times p$ matrices only, even though the complete covariance matrix is $pqT \times pqT$.

THEOREM 1. *Consider the nonlinear regression model (2.1) under Assumptions 2.1 and 2.2. Let Q be the $T \times T$ matrix*

(3.1)

$$
Q = \begin{bmatrix}
(1 - \alpha^2)^{1/2} & -\alpha & 0 & \cdots & 0 & 0 \\
0 & 1 & -\alpha & \cdots & 0 & 0 \\
\cdot & & \cdot & & & \cdot \\
\cdot & & & \cdot & & \cdot \\
\cdot & & & & \cdot & \cdot \\
0 & 0 & 0 & \cdots & 1 & -\alpha \\
0 & 0 & 0 & \cdots & 0 & 1
\end{bmatrix},
$$

so that $QQ' = M_\alpha^{-1}$, and let Z be an orthogonal $T \times T$ matrix such that

(3.2)
$$Z'Q'M_\rho QZ = \Xi,$$

where Ξ is a diagonal matrix with $\xi_1, \ldots, \xi_T$ on the diagonal. Define the $T \times T$ matrix $S = (\sigma_{st})$ by $S = QZ$, and the $p \times p$ matrices W_t by

(3.3)
$$W_t = \Delta + q\xi_t \Gamma, \qquad t = 1, \ldots, T.$$

Then the log-likelihood function is

(3.4)
$$L = -(1/2)qTp \ln 2\pi - (1/2)\ln |\Omega| - (1/2)u'\Omega^{-1}u,$$

with

(3.5)
$$|\Omega| = \left(\prod_{i=1}^{q} \lambda_i\right)^{pT} (1 - \alpha^2)^{-pq} |\Delta|^{(q-1)T} \prod_{t=1}^{T} |W_t|$$

and

(3.6)
$$
u'\Omega^{-1}u = \sum_{t=1}^{T} (1 + \delta_t \alpha^2)\left(\sum_{j=1}^{q} \lambda_j^{-1} u'_{jt} \Delta^{-1} u_{jt} - q\bar{u}'_t \Delta^{-1} \bar{u}_t\right)
$$
$$
- 2\alpha \sum_{t=1}^{T-1} \left(\sum_{j=1}^{q} \lambda_j^{-1} u'_{jt} \Delta^{-1} u_{j,t+1} - q\bar{u}'_t \Delta^{-1} \bar{u}_{t+1}\right)
$$
$$
+ q \sum_{t=1}^{T} \left(\sum_{s=1}^{T} \sigma_{st} \bar{u}_s\right)' W_t^{-1} \left(\sum_{s=1}^{T} \sigma_{st} \bar{u}_s\right),
$$

where $\delta_1 = \delta_T = 0$, $\delta_t = 1 \ (t = 2, \ldots, T - 1)$, and

(3.7)
$$u_{jt} = y_{jt} - f_j(X_{jt}, \beta_j), \qquad \bar{u}_t = (1/q) \sum_{j=1}^{q} \lambda_j^{-1} u_{jt}.$$

PROOF. The proof proceeds by applying the lemmas of the appendix taking $L = \Lambda$, $M = M_\alpha$, $G = M_\rho$, and $a = \iota = (1, 1, \ldots, 1)'$. In particular, we use (A.4), (A.5), and (A.19). For a full proof, see Magnus and Woodland (1987a).   Q.E.D.

Theorem 1 allows substantial generalisation. In fact, it appears from the Appendix that essentially the same result holds for

(3.8)
$$\Omega = L \otimes M \otimes \Delta + aa' \otimes G \otimes \Gamma$$

where L is a positive definite $q \times q$ matrix (not necessarily diagonal), a is a $q \times 1$ vector (whose components are not necessarily equal), and M and G are arbitrary $T \times T$ matrices (M positive definite, G positive semidefinite). In particular, M and G can be covariance matrices of very general ARMA processes. (In Theorem 1 both M and G represent an AR(1) process.)

Theorem 1 shows that although the complete covariance matrix is $pqT \times pqT$, the likelihood function only contains $p \times p$ matrices. Evaluation of L is therefore easy and inexpensive. Even so, the total number of parameters can be large. It seems worthwhile, therefore, to investigate whether concentration of the likelihood with respect to the β parameters is possible. It turns out that concentration is indeed possible, but only in a somewhat restricted class of linear models. The next section deals with this case.

4. THE LINEAR CASE

In Sections 2 and 3, we considered a set of *nonlinear* regression equations given by (2.1). Let us now consider the following set of *linear* regressions:

$$(4.1) \qquad y_{it} = X_t \beta_i + u_{it}, \qquad i = 1, \ldots, q, \qquad t = 1, \ldots, T,$$

where y_{it} and u_{it} are $p \times 1$ vectors, the nonrandom inputs X_t are $p \times k$ matrices, and the unknown parameter vectors β_i are $k \times 1$ vectors. Notice that the inputs X_t are the same for each i. This is essential. Without this feature, much of the subsequent treatment would not lead to tractable results.

In this section, we shall investigate whether in the linear model (4.1) concentration of the likelihood function with respect to the β parameters is feasible. The following theorem shows that this is the case.

THEOREM 2. *Consider the linear regression model (4.1) under Assumptions 2.1 and 2.2. The maximum likelihood equations for $\beta_1, \beta_2, \ldots, \beta_q$ are*

$$(4.2) \qquad \beta_i = Q_1^{-1} d_i + Q_2^{-1} d,$$

and the "asymptotic covariance matrix" of the ML estimators $\tilde{\beta}_1, \ldots, \tilde{\beta}_q$ is given by

$$(4.3) \qquad \text{as.var}\,(\tilde{\beta}_i) = \lambda_i Q_1^{-1} + (1/q)(Q_2^{-1} - Q_1^{-1})$$

and

$$(4.4) \qquad \text{as.cov}\,(\tilde{\beta}_i, \tilde{\beta}_j) = (1/q)(Q_2^{-1} - Q_1^{-1}), \qquad i \neq j,$$

where

$$(4.5) \qquad d_i = \sum_{t=1}^{T} (1 + \delta_t \alpha^2) c_{it}^i - \alpha \sum_{t=1}^{T-1} (c_{t,t+1}^i + c_{t+1,t}^i),$$

$$(4.6) \qquad d = \sum_{s=1}^{T} \left(\sum_{t=1}^{T} \sigma_{ts} X_t \right)' W_s^{-1} \left(\sum_{t=1}^{T} \sigma_{ts} \bar{s}_t \right),$$

$$(4.7) \qquad Q_1 = \sum_{t=1}^{T} (1 + \delta_t \alpha^2) C_{tt} - \alpha \sum_{t=1}^{T-1} (C_{t,t+1} + C_{t+1,t}),$$

$$(4.8) \qquad Q_2 = \sum_{s=1}^{T} \left(\sum_{t=1}^{T} \sigma_{ts} X_t \right)' W_s^{-1} \left(\sum_{t=1}^{T} \sigma_{ts} X_t \right),$$

and

$$(4.9) \qquad C_{st} = X_s' \Delta^{-1} X_t, \quad c_{st}^i = X_s' \Delta^{-1}(y_{it} - \bar{y}_t), \quad \bar{y}_t = (1/q) \sum_{j=1}^{q} \lambda_j^{-1} y_{jt},$$

while δ_t and σ_{ts} are defined in Theorem 1.

PROOF. From Magnus (1978, Theorems 1 and 3), the ML equation (first-order condition) for $\beta = (\beta_1', \ldots, \beta_q')'$ is

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

and the "asymptotic covariance matrix" of the ML estimator β is

$$\text{as.cov}\,(\tilde{\beta}) = (X'\Omega^{-1}X)^{-1}.^{[4]}$$

Again, we apply the results of the Appendix, taking $N = I$, $a = \iota$, $L = \Lambda$, and using (A.7), (A.8), (A.11), (A.13), (A.14), and (A.18). A detailed proof is available in Magnus and Woodland (1987a).   Q.E.D.

Given the ML equations for $\beta_1, \ldots, \beta_q$ from (4.2), we may now define

$$(4.10) \qquad \tilde{u}_{jt} = y_{jt} - X_t \tilde{\beta}_j, \qquad \tilde{\bar{u}}_t = (1/q) \sum_{j=1}^{q} \lambda_j^{-1} \tilde{u}_{jt}$$

and replace u_{jt} and $\bar{u}_t$ with $\tilde{u}_{jt}$ and $\tilde{\bar{u}}_t$ in (3.6) and, hence, in (3.4). Thus, we obtain a concentrated log-likelihood function containing only $m \equiv p(p + 1) + q + 1$ parameters (Γ, Δ, Λ, α, ρ). Maximization of the concentrated log-likelihood function yields ML estimates for these m parameters, and (4.2) can then be used to obtain ML estimates for the remaining qk parameters. That this is a worthwhile procedure is shown by our empirical example (Section 5). There we have $p = 3$, $q = 6$, and $k = 9$, so that, of the 73 parameters to be estimated, 54 are concentrated out of the likelihood function.

The information matrix, which yields the asymptotic variances and covariances of the ML estimators, is block-diagonal (Magnus 1978, Theorem 3). For the "structural" parameters β_i we find from (4.3),

$$\text{as.var}\,(\tilde{\beta}_i) = \lambda_i Q_1^{-1} + (1/q)(Q_2^{-1} - Q_1^{-1}).$$

Denoting the m covariance parameters (Γ, Δ, Λ, α, ρ) as $\theta_1, \theta_2, \ldots, \theta_m$, we can

---

[4] Strictly speaking, we should differentiate between the *function* Ω and the true (unknown) *value* of Ω, say Ω_0. Thus the asymptotic covariance matrix of β is $(X'\Omega_0 X)^{-1}$.

compute the symmetric $m \times m$ matrix Ψ with typical element

$$(4.11) \qquad \Psi_{ij} = \text{tr}\left(\frac{\partial \Omega^{-1}}{\partial \theta_i} \Omega \frac{\partial \Omega^{-1}}{\partial \theta_j} \Omega \right).$$

The matrix $2\Psi^{-1}$ then gives the asymptotic variances and covariances of $\hat{\theta}_1, \ldots,$ $\hat{\theta}_m$.[5]

Two problems remain. The first is that maximization of the concentrated log-likelihood function with respect to Γ, Δ, Λ, α and ρ does not guarantee that the estimate for Γ is positive semi-definite, Δ positive definite, $\lambda_i (i = 1, \ldots, q)$ positive, $|\alpha| < 1$, and $|\rho| < 1$. Hence, we write

$$\Gamma = L_c L_c' \quad \text{and} \quad \Delta = L_d L_d',$$

where L_c and L_d are lower triangular matrices, and maximize the concentrated log-likelihood with respect to the lower triangular elements of L_c and L_d, and the λ_i, under the constraint that the diagonal elements of L_c and L_d, and the λ_i are all nonnegative.[6] This procedure ensures that ML estimates for Γ and Δ are positive semi-definite, but it does not imply nonsingularity of Δ.

The second problem is that most estimation procedures of dynamic error components models produce inconsistent estimates. A unified treatment of this problem underlining the importance of initial values is given in Sevestre and Trognon (1983). While the maximum likelihood estimator will not be consistent as q increases with T fixed (although the inconsistency is likely to be slight, see Sevestre and Trognon 1985), it will be consistent and asymptotically normal as T increases for given q; see Heijmans and Magnus (1986a, 1986b) for regularity conditions in a more general context. The latter case (where T increases and q is fixed) is the one we have in mind in this study and an empirical example is described in the following section.

5. AN EMPIRICAL ILLUSTRATION: INTERFUEL SUBSTITUTION IN DUTCH MANUFACTURING

In this section, we illustrate the model discussed above by estimating it using a combined time-series, cross-section data set relating to the inputs of fuels in various Dutch manufacturing industries. In particular, we estimate a system of fuel cost share equations for six industries taking into account contemporaneous and serial correlations in the disturbances by using our error components model.

In recent years there has been substantial interest in the effects of rapidly changing energy prices upon industrial structure and the demand for energy and non-energy inputs. Not only have energy prices been changing relative to the prices of other inputs, but prices of the various energy inputs have also been changing relative to each other. Our empirical application is concerned with measuring the changes in the mix of four types of energy inputs in response to

[5] Magnus (1978, Theorem 3).
[6] See Magnus (1982, p. 247 and footnote 7).

changes in the relative prices of these inputs using annual time-series data for six manufacturing industries in The Netherlands. The substitution between energy and non-energy inputs in The Netherlands was studied by Magnus (1979). Previous attempts to measure the substitution and complementarity relationships between fuels in The Netherlands are by Pindyck (1979) and Griffin (1977), both of whom use aggregate data. In contrast, the present study represents a disaggregated approach, using combined time-series and cross-section data for six separate industries within the Dutch manufacturing sector.

We distinguished between four types of energy: (i) solid fuels (COAL), (ii) liquid fuels (OIL), (iii) natural and manufactured gas (GAS), and (iv) electricity (ELEC). We also distinguish between six industries of the Dutch manufacturing sector: (i) food, beverages and tobacco products (FOOD), (ii) textiles (TEXT), (iii) paper and paper products (PAPR), (iv) chemical industry (CHEM), (v) building materials, earthenware, glass, and glass products (BLDG), and (vi) fabricated metal products, transport equipment, and mechanical and electrical engineering (METL). The six industries combined account for roughly 75 percent of Dutch manufacturing output.[7] Data requirements prohibit further disaggregation.

Under the assumption that each industry faces the same energy prices, the estimation requires data for price indices of COAL, OIL, GAS, and ELEC, and input costs for each of the four energy types in each industry. The time period is 1958 to 1976, for this is the period during which the data collected by The Netherlands Central Bureau of Statistics are available.[8]

In order to model the energy outlays of different industries, we assume that each industry minimises its cost in two stages, whereby the optimal mix of fuel inputs is chosen in the first stage, and in the second stage the optimal amount of "aggregate energy" is chosen along with other variable inputs and outputs.[9] This assumption implies that we can investigate the substitution possibilities between the various fuel inputs without having to concern ourselves with substitution between fuel inputs and other commodities.

Assuming, in addition, that the functional form for the unit cost function in the first stage is the translog function due to Christensen, Jorgenson, and Lau (1973), we obtain the following cost share equations:

$$(5.1) \qquad y_{ijt} = b^i_j + \sum_{k=1}^{n} b^i_{kj} \ln p_{kt} + u_{ijt} \qquad \begin{aligned} j &= 1, \ldots, n \equiv 4; \\ i &= 1, \ldots, q \equiv 6; \\ t &= 1, \ldots, T \equiv 19, \end{aligned}$$

where y_{ijt} denotes the share of fuel input j in the total energy cost of production

---

[7] The main industry left out is the basic metal industry. From 1958–1968 this industry was a net producer of gas. Since we wished to concentrate upon the demand for the inputs of energy, rather than on their supplies, the basic metal industry did not fit into our framework.

[8] A full documentation and discussion of the data (in Dutch) is provided in Magnus and Vastenou (1978). The data are tabled in the Appendix to Magnus and Woodland (1987b).

[9] The "two-stage optimization" is equivalent to homothetic separability of fuel inputs from all other inputs and outputs in the technology. See Shephard (1970, p. 143–146).

in industry i at time t, p_{kt} is the price of the kth fuel input at time t (which is the same for each industry), the u_{ijt} are disturbances, and the parameters $b^i_{.j}$ and b^i_{kj} satisfy[10]

$$(5.2) \qquad b^i_{kj} = b^i_{jk}, \qquad \sum_{j=1}^{n} b^i_{.j} = 1, \qquad \sum_{j=1}^{n} b^i_{kj} = 0.$$

Since we are dealing with shares, the disturbances are constrained by $\sum_{j=1}^{n} u_{ijt} = 0$. This implies that we may arbitrarily drop one of the n equations, say the last, in each year and for each industry (for example, see Barten 1969). Incorporating the parameter restrictions, we can write (5.1) as

$$(5.3) \qquad y_{it} = X_t \beta_i + u_{it}$$

with

$$y_{it} = (y_{1it}, y_{2it}, y_{3it})', \qquad u_{it} = (u_{1it}, u_{2it}, u_{3it})',$$

$$X_t = \begin{bmatrix} 1 & 0 & 0 & \phi_{1t} & \phi_{2t} & \phi_{3t} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \phi_{1t} & 0 & \phi_{2t} & \phi_{3t} & 0 \\ 0 & 0 & 1 & 0 & 0 & \phi_{1t} & 0 & \phi_{2t} & \phi_{3t} \end{bmatrix},$$

where $\phi_{jt} = \ln(p_{jt}/p_{4t})$, $j = 1, 2, 3$, and

$$\beta_i = (b^i_{.1}, b^i_{.2}, b^i_{.3}, b^i_{11}, b^i_{12}, b^i_{13}, b^i_{22}, b^i_{23}, b^i_{33})'.$$

The model described by (5.3) thus fits into the framework developed in previous sections. In the present case, $T = 19$, $q = 6$, $p = n - 1 = 3$, and $k = 9$. Also notice that the matrix of explanatory variables is independent of i and so the results of Section 4 apply.

We shall assume that the disturbance vectors u_{it} are distributed according to Assumption 2.1. The stochastic specification thus employs a multivariate error components model involving serial correlation. This model specifies that the vector of random disturbances in the system of share equations is composed of a vector which is different for each industry, and a common vector, which reflects macro-economic disturbances affecting all industries in the same manner. As a result, the disturbances for the different industries are correlated with each other, requiring joint estimation of the share equations for all industries on efficiency grounds. Moreover, both components of the disturbance vectors are assumed to follow first-order autoregressive processes. Thus, the model involves both contemporaneous and intertemporal correlations between disturbances. It is estimated by the method of maximum likelihood. Our tests show that both the intertemporal correlations and the correlations between the industry disturbances are not negligible and, hence, that there are efficiency gains arising from joint estimation of the industry share equations.

---

[10] The first set of restrictions are the symmetry restrictions required for identification. The remaining restrictions are necessary and sufficient for the cost function to be homogeneous of degree one in prices or, equivalently, for the share equations to sum to unity and be homogeneous of degree zero in prices.

TABLE 1

PARAMETER ESTIMATES FOR TRANSLOG COST FUNCTIONS[a]

	FOOD	TEXT	PAPR	CHEM	BLDG	METL
b_C	.1030	.0927	.1468	.0578	.2076	.0671
	(.0424)	(.0417)	(.0469)	(.0490)	(.0420)	(.0430)
b_O	.3088	.2376	.2216	.1643	.1914	.1339
	(.0862)	(.0857)	(.0896)	(.0912)	(.0859)	(.0866)
b_G	.2071	.1523	.2533	.2742	.2910	.1498
	(.0856)	(.0853)	(.0875)	(.0885)	(.0855)	(.0859)
b_E	.3811	.5174	.3784	.5037	.3100	.6492
	(.0278)	(.0270)	(.0330)	(.0354)	(.0274)	(.0286)
b_{CC}	−.0157	−.0242	−.0819	.0603	−.0822	.0162
	(.0430)	(.0406)	(.0574)	(.0634)	(.0417)	(.0452)
b_{CO}	−.0683	−.0605	−.0510	−.0519	−.0453	−.0205
	(.0407)	(.0384)	(.0547)	(.0605)	(.0395)	(.0429)
b_{CG}	.0832	.0597	.1398	.0519	.0893	.0009
	(.0237)	(.0227)	(.0299)	(.0325)	(.0232)	(.0246)
b_{CE}	.0007	.0250	−.0069	−.0603	.0382	.0034
	(.0271)	(.0254)	(.0369)	(.0410)	(.0262)	(.0286)
b_{OO}	.0604	.1320	.0695	.2489	.1049	.1142
	(.0614)	(.0589)	(.0771)	(.0839)	(.0600)	(.0637)
b_{OG}	.0470	.0291	.0253	−.0165	.0063	.0044
	(.0428)	(.0421)	(.0480)	(.0504)	(.0424)	(.0436)
b_{OE}	−.0391	−.1006	−.0438	−.1805	−.0660	−.0982
	(.0259)	(.0241)	(.0362)	(.0404)	(.0249)	(.0275)
b_{GG}	−.0386	−.0109	−.0944	.0200	−.0263	.0830
	(.0424)	(.0419)	(.0459)	(.0475)	(.0421)	(.0429)
b_{GE}	−.0915	−.0779	−.0707	−.0555	−.0693	−.0883
	(.0185)	(.0175)	(.0252)	(.0279)	(.0180)	(.0196)
b_{EE}	.1300	.1535	.1214	.2963	.0971	.1831
	(.0303)	(.0285)	(.0415)	(.0462)	(.0293)	(.0321)

[a] Asymptotic standard errors are in parentheses. Fuels are indicated by C = coal, O = oil, G = gas, E = electricity.

Since the model (5.3) describes the type of linear regression equations investigated in Section 4, Theorem 2 applies and we can concentrate the likelihood function with respect to the 54 β_i parameters of the share equations. The concentrated likelihood function then contains only 19 independent parameters, consisting of 6 lower triangular elements of Δ, another 6 for Γ, 5 λ parameters and the autoregression parameters α and ρ. The full covariance matrix in this empirical example is of dimension $Tq(n-1) = 19 \times 6 \times 3 = 342$. The decomposition of the inverse and determinant of the covariance matrix, as provided by Theorem 1, requires calculation of inverses and determinants of matrices of order 3 rather than of the full covariance matrix. Thus, Theorems 1 and 2 allow considerable simplification of the computational burden involved in calculating the likelihood function.

The full model was estimated by maximizing the concentrated log-likelihood function with respect to the 19 covariance parameters.[11] The β_i estimates are presented together with the estimates of their asymptotic standard errors in Table 1.

Approximately one half of the parameters of the share equations are signifi-

[11] The ML estimates were obtained using the Quasi-Newton algorithm by Fletcher (1972).

TABLE 2

ESTIMATES OF COVARIANCE PARAMETERS[a]

		Δ^*			
		COAL	OIL	GAS	ELEC
COAL		2.87 (.44)			
OIL		−1.15 (.42)	5.37 (.82)		
GAS		−.93 (.32)	−2.64 (.51)	3.03 (.47)	
ELEC		−.79 (.25)	−1.58 (.37)	.54 (.25)	1.84 (.28)

		Γ^*			
COAL		2.72 (1.05)			
OIL		.20 (1.54)	13.08 (4.57)		
GAS		−2.32 (1.63)	−11.96 (4.29)	13.23 (4.49)	
ELEC		−.61 (.48)	−1.32 (1.01)	1.05 (.96)	.87 (.39)

λ						Autocorrelation parameters	
FOOD	TEXT	PAPR	CHEM	BLDG	METL	α	ρ
.8175	.6271	2.2214	2.9321	.7139	1.0101	.7398	.9378
(.1698)	(.1267)	(.4686)	(.6187)	(.1468)	(.2116)	(.0385)	(.0372)

[a] Asymptotic standard errors are in parentheses. Estimates of parameters and standard errors of Δ^* and Γ^* are the values in the table multiplied by 10^{-4}. Since the λ's are normalized by $\Sigma_{i=1}^{9} \lambda_i^{-1} = 6$, the λ estimate for METL and its standard error were obtained indirectly.

cantly different from zero, using a 5 percent type I error probability.[12] In addition, about 40 percent of the price coefficients are significantly different from zero, showing the importance of the dependence of fuel cost shares upon the relative prices of the four fuels. For example, in most industries b_{CG} is significantly positive, indicating that an increase in the price of gas will cause an increase in the cost share of coal. A more detailed analysis of the estimates of the substitution-complementarity relationships between fuels is provided in Magnus and Woodland (1987b).

The ML estimates for Δ^*, Γ^*, λ_i, and the autocorrelation parameters α and ρ, together with the estimates of their asymptotic standard errors are presented in Table 2.[13] We see that the Γ^* parameters are roughly twice the size of the Δ^* parameters (as measured by their traces), which underlines the importance of the disturbances e_t (which affect all industries in the same way) relative to the disturbances ε_{it} (which differ from industry to industry). It is remarkable that the sign patterns of Δ^* and Γ^* are the same, but for one element. Ignoring for the moment the autocorrelation parameters α and ρ, the disturbance covariance matrix for the ith industry is $\Gamma^* + \lambda_i \Delta^*$ in every year, and since the λ_i's are in the order of unity, the sign of an element of $\Gamma^* + \lambda_i \Delta^*$ is the same as the sign of the corresponding element of Δ^*. Since all non-price effects are represented by the

[12] Since there are only 19 years of observations, it is doubtful if the asymptotic theory would hold and so the significance tests must be interpreted cautiously.

[13] Δ^* and Γ^* are the full 4×4 covariance matrices of which Δ and Γ are the 3×3 north-west submatrices. The last lines of Δ^* and Γ^* have been reconstructed using the adding-up constraint.

TABLE 3

TEST OF COVARIANCE RESTRICTIONS[a]

Restriction	$\Gamma = 0$	$\Lambda = 1$	$\alpha = 0$	$\rho = 0$	$\alpha = \rho$
L	769.21	831.24	793.43	830.50	847.81
LR Statistic	161.81	37.74	113.36	39.22	4.61

[a] L denotes the loglikelihood of the restricted model. The loglikelihood of the unrestricted model is $L^* = 850.11$. The likelihood ratio statistic is $LR = 2(L^* - L)$.

disturbance, the elements of Δ^* indicate that non-price factors affecting the cost share of a particular input will, in general, have an opposite effect on the cost shares of the other inputs. (The exception is gas − electricity.) The matrix Γ^* represents the covariances between input shares which are common to all industries. Thus, for example, if one industry uses a larger than average amount of gas then the share of coal and oil will tend to be smaller in *every* industry, since γ_{GC} and γ_{GO} are negative. Also, since Γ has significantly positive diagonal elements, if one industry has a larger than average share for an input (say oil) then all other industries will tend to have larger than average shares for that input as well. It is in this manner that the error components formulation models the overall allocation effects of economy-wide disturbances.

The λ estimates in Table 2 show that, while the hypothesis that $\lambda_i = 1$ in all industries is rejected (as we shall see shortly), only three industries (TEXT, PAPR, and CHEM) show a λ value which is significantly different from one. The covariance matrices of the disturbances in the CHEM and PAPR industries are somewhat "larger" than average, while the covariance matrix of the disturbances in the TEXT industry is somewhat "smaller."

Finally, the estimates of the two autocorrelation parameters, $\bar{\alpha} = .74$ and $\bar{\rho} = .94$, are both highly significant, thus firmly rejecting the hypothesis of time independence.

As we indicated in Section 2, several special cases of the general covariance structure are of interest. We therefore estimated the following five restricted versions of the model: $\Gamma = 0$, $\Lambda = I$, $\alpha = 0$, $\rho = 0$, $\alpha = \rho$.[14] The resulting likelihood ratio (LR) statistics are presented in Table 3. Each of these hypotheses is rejected at the 5 percent significance level.[15] These results confirm the significance of our assumption regarding the covariance structure.

The test result, $\Gamma \neq 0$, indicates that the error component, e_t, which is common to all industries, is an important part of the disturbance vector in our empirical application. It therefore justifies the approach taken, whereby the share equations

[14] It is not obvious how the hypothesis $H_0 : \Gamma = 0$ should be tested. One problem is that Γ lies on the boundary of the parameter space, if H_0 is true. Another is that nuisance parameters are present only under the alternative hypothesis, i.e. if we write $\Gamma = \sigma^2 \bar{\Gamma}$ with tr $\bar{\Gamma} = 1$, then $\Gamma = 0$ is equivalent to $\sigma^2 = 0$ irrespective of the values in $\bar{\Gamma}$. See Moran (1971) and Davies (1977). In our case the LR statistic, though not strictly applicable, is sufficiently large (161.81) to reject H_0 with confidence.

[15] At the 5 percent significance level the critical χ^2 values are 3.84 (1 degree of freedom) and 11.1 (5 degrees of freedom); at the 1 percent level we have 6.63 (1 degree of freedom) and 15.1 (5 degrees of freedom).

of the six industries were estimated jointly rather than separately for each industry. The result that $\Lambda \neq I$ indicates that the disturbance covariance matrices for the six industries are not the same. Finally, the tests on parameters α and ρ demonstrate that both of the two-error components are subject to serial correlation, but that the first-order autocorrelation parameters are different.

These test results show, therefore, that the data are not consistent with any of our model's special cases of the covariance structure for the disturbances. This model is fairly general in that it allows for contemporaneous correlation between disturbances within and between industries, as well as serial correlation, in a way that is parsimonious regarding parameters. The model also allows a significant decomposition of the likelihood function. The question of whether this specification would be rejected in favour of some more general model is beyond the scope of this paper.

6. CONCLUSION

In this paper, we extend the two-error component regression model to deal simultaneously with the case where each component is a vector (multivariate error components) and with the case where each component is serially correlated. The model is discussed and the maximum likelihood estimator analyzed. In particular, in the nonlinear regression context it is shown that the determinant and inverse of the covariance matrix can be decomposed so that the likelihood function can be evaluated by calculating determinants and inverses of order at most p (the number of equations).

In the case of a linear regression model where the matrix of explanatory variables does not depend upon the cross-section, it is shown that the likelihood function may be analytically concentrated with respect to the regression parameters, thus permitting the ML estimates to be obtained by maximizing the concentrated log-likelihood function with respect to only the covariance parameters.

In addition, the asymptotic covariance matrix for the ML estimators is obtained. The paper also shows how the results may be extended to deal with more complicated covariance structures for the error components.

The model and ML estimation procedure are illustrated by applying them to a combined cross-section, time-series data set on fuel consumption in six Dutch manufacturing industries over the post-war period.

The use of combined cross-section and time-series data requires special attention to the stochastic specification. The error components structure for the disturbance in a univariate regression model is well established in the theoretical literature. Our generalization of this model allows for the usual correlations between disturbances in the fuel equations, for the correlation between disturbances in the different industries (via the error components structure), and for correlations between disturbances in different time periods. These correlations are achieved by a minimal number of covariance parameters. Our empirical results show that the interindustry and the intertemporal correlations are significant, indicating that joint estimation of the complete set of industry fuel share equations is justified on grounds of efficiency.

Despite the large number of parameters to be estimated, the error components model we specify is relatively easy to estimate by maximum likelihood. It should be useful in a variety of contexts, especially since multivariate data are becoming increasingly available as combined cross-section and time-series.

London School of Economics, England
University of Sydney, Australia

APPENDIX
SOME MATRIX RESULTS USEFUL FOR MULTIVARIATE TWO ERROR COMPONENTS ANALYSIS

The three lemmas in this appendix give the mathematical background of Theorems 1 and 2. They also provide the tools for extending the theory to more general covariance structures. The proofs, which make repeated use of Lemma 2.2 of Magnus (1982), are given in detail in Magnus and Woodland (1987a).

Let L, M and Δ be positive definite matrices of orders $q \times q$ ($q \geq 2$), $T \times T$ and $p \times p$, respectively. Let G and Γ be positive semidefinite matrices of orders $T \times T$ and $p \times p$, respectively, and let $a \neq 0$ be a $q \times 1$ vector. Let S be a nonsingular $T \times T$ matrix with columns $\sigma_1, \ldots, \sigma_T$, such that

(A.1) $$S'MS = I_T \quad \text{and} \quad S'GS = \Xi$$

where Ξ is a diagonal $T \times T$ matrix with nonnegative diagonal elements $\xi_1, \ldots, \xi_T$.[16] Define

(A.2) $$\alpha = a'L^{-1}a, \qquad W_t = \Delta + \alpha \xi_t \Gamma, \qquad t = 1, \ldots, T,$$

and

(A.3) $$\Omega = L \otimes M \otimes \Delta + aa' \otimes G \otimes \Gamma.$$

Lemma A.1 gives the determinant and the inverse of the very complicated $qTp \times qTp$ matrix Ω in a manageable and computable form.

LEMMA A.1. *The $qTp \times qTp$ matrix Ω defined in (A.3) is positive definite with determinant*

(A.4) $$|\Omega| = |L|^{Tp} |M|^{qp} |\Delta|^{(q-1)T} \prod_{t=1}^{T} |W_t|$$

and inverse

(A.5) $$\Omega^{-1} = (L^{-1} - (1/\alpha)L^{-1}aa'L^{-1}) \otimes M^{-1} \otimes \Delta^{-1}$$
$$+ (1/\alpha)L^{-1}aa'L^{-1} \otimes \sum_{t=1}^{T} (\sigma_t \sigma_t' \otimes W_t^{-1}).$$

---

[16] The matrix S can be constructed as follows: let Q be a square matrix such that $M^{-1} = QQ'$, and let Z be an orthogonal matrix such that $Z'Q'GQZ = \Xi$ (diagonal). Then $S = QZ$ has the desired properties. See also Bellman (1970, p. 58).

Now, in addition to the matrices defined above, let $X_1, X_2, \ldots, X_T$ be $p \times k$ matrices such that the matrix

$$\text{(A.6)} \qquad \bar{X} = (X_1', X_2', \ldots, X_T')'$$

has full column-rank k. Let N be a $q \times r$ matrix $(1 \leq r \leq q)$ with full column-rank r and define

$$\text{(A.7)} \qquad X = N \otimes \bar{X}, \qquad L_0 = (N'L^{-1}N)^{-1} \quad \text{and} \quad b = N'L^{-1}a.$$

Denote by μ^{st} and σ_{st} the st-th element of M^{-1} and S, respectively. Defining

$$\text{(A.8)} \qquad Q_1 = \sum_{s=1}^{T} \sum_{t=1}^{T} \mu^{st} X_s' \Delta^{-1} X_t$$

and

$$\text{(A.9)} \qquad Q_2 = \sum_{t=1}^{T} \left(\sum_{s=1}^{T} \sigma_{st} X_s \right)' W_t^{-1} \left(\sum_{s=1}^{T} \sigma_{st} X_s \right),$$

we obtain Lemma A.2.

LEMMA A.2. *The $k \times k$ matrices Q_1 and Q_2 defined in (A.8) and (A.9) are both positive definite. Furthermore, if the $q \times 1$ vector a is a linear combination of the columns of N, then*

$$\text{(A.10)} \qquad X'\Omega^{-1}X = (L_0^{-1} - (1/\alpha)bb') \otimes Q_1 + (1/\alpha)bb' \otimes Q_2$$

and

$$\text{(A.11)} \qquad (X'\Omega^{-1}X)^{-1} = (L_0 - (1/\alpha)L_0 bb'L_0) \otimes Q_1^{-1} + (1/\alpha)L_0 bb'L_0 \otimes Q_2^{-1}.$$

REMARK. The assumption that a lies in the column-space of N is made for simplicity only. It will be satisfied in most applications; in particular, it is satisfied in our application. Without this condition the expressions become somewhat more involved.

Finally, let y_{jt} $(j = 1, \ldots, q; t = 1, \ldots, T)$ be $p \times 1$ vectors and define the $qTp \times 1$ vector

$$\text{(A.12)} \qquad y = (y_{11}', y_{12}', \ldots, y_{1T}', \ldots, y_{qT}')'.$$

Denote by λ^{ij} and l_j the ijth element and jth column of L^{-1}, respectively, and define

$$\text{(A.13)} \qquad c_j = \sum_{s=1}^{T} \sum_{t=1}^{T} \mu^{st} X_s' \Delta^{-1} y_{jt},$$

$$\text{(A.14)} \qquad c = (1/\alpha) \sum_{j=1}^{q} (a'l_j)c_j,$$

$$\text{(A.15)} \qquad d = \sum_{t=1}^{T} \left(\sum_{s=1}^{T} \sigma_{st} X_s \right)' W_t^{-1} \left(\sum_{s=1}^{T} \sigma_{st} \bar{y}_s \right),$$

where

(A.16) $$\bar{y}_s = (1/\alpha) \sum_{j=1}^{q} (a'l_j) y_{js}.$$

LEMMA A.3. *If a is a linear combination of the columns of N, then*

(A.17) $$X'\Omega^{-1} y = b \otimes (d - c) + \sum_{j=1}^{q} (N'l_j \otimes c_j),$$

(A.18) $$(X'\Omega^{-1} X)^{-1} X'\Omega^{-1} y = L_0 b \otimes (Q_2^{-1} d - Q_1^{-1} c) + \sum_{j=1}^{q} (L_0 N'l_j \otimes Q_1^{-1} c_j),$$

and

(A.19) $$y'\Omega^{-1} y = \sum_{s=1}^{T} \sum_{t=1}^{T} \mu^{st} \left(\sum_{i=1}^{q} \sum_{j=1}^{q} \lambda^{ij} y'_{is} \Delta^{-1} y_{jt} - \alpha \bar{y}'_s \Delta^{-1} \bar{y}_t \right)$$
$$+ \alpha \sum_{t=1}^{T} \left(\sum_{s=1}^{T} \sigma_{st} \bar{y}_s \right)' W_t^{-1} \left(\sum_{s=1}^{T} \sigma_{st} \bar{y}_s \right).$$

REFERENCES

ANDERSON, T. W. AND C. HSIAO, "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics* 18 (1982), 47–82.

BALESTRA, P. AND M. NERLOVE, "Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas," *Econometrica* 34 (1966), 585–612.

BARTEN, A. P., "Maximum Likelihood Estimation of a Complete System of Demand Equations," *European Economic Review* 1 (1969), 7–73.

BELLMAN, R., *Introduction to Matrix Analysis*, Second Edition (New York: McGraw-Hill, 1970).

BERNDT, E. R. AND N. E. SAVIN, "Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances," *Econometrica* 43 (1975), 937–957.

CHAMBERLAIN, G. AND Z. GRILICHES, "Unobservables with a Variance-Components Structure: Ability, Schooling and the Economic Success of Brothers," *International Economic Review* 16 (1975), 422–450.

CHRISTENSEN, L. R., D. W. JORGENSON, AND L. J. LAU, "Transcendental Logarithmic Production Frontiers," *Review of Economics and Statistics* 55 (1973), 28–55.

DAVIES, R. B., "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika* 64 (1977), 247–254.

FLETCHER, R., "Fortran Subroutines for Minimization by Quasi-Newton Methods," Theoretical Physics Division, Atomic Energy Research Establishment, Harwell, England (1972).

GRIFFIN, J. M., "Inter-Fuel Substitution Possibilities: A Translog Application to Intercountry Data," *International Economic Review* 18 (1977), 755–770.

HAUSE, J. C., "The Fine Structure of Earnings and the On-the-Job Training Hypothesis," *Econometrica* 48 (1980), 1013–1029.

HEIJMANS, R. D. H. AND J. R. MAGNUS, "Consistent Maximum-Likelihood Estimation with Dependent Observations," *Journal of Econometrics* 32 (1986a), 253–285.

———— AND ————, "Asymptotic Normality of Maximum Likelihood Estimators Obtained from Normally Distributed But Dependent Observations," *Econometric Theory* 2 (1986b), 374–412. Corrigenda in Journal of Econometrics 35 (1988), 395.

KING, M. L., "Efficient Estimation and Testing of Regressions with a Serially Correlated Error

Component," Working Paper 3/86, Department of Econometrics and Operations Research, Monash University (1986).

LILLARD, L. A. AND Y. WEISS, "Components of Variation in Panel Earnings Data: American Scientists 1960-70," *Econometrica* 47 (1979), 437–454.

—— AND R. J. WILLIS, "Dynamic Aspects of Earning Mobility," *Econometrica* 46 (1978), 985–1012.

MACURDY, T. E., "The Use of Time Series Processes to Model The Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics* 18 (1982), 83–114.

MAGNUS, J. R., "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix," *Journal of Econometrics* 7 (1978), 281–312. Corrigenda in *Journal of Econometrics* 10, 261.

——, "Substitution between Energy and Non-Energy Inputs in The Netherlands 1950–1976," *International Economic Review* 20 (1979), 465–484.

——, "Multivariate Error Components Analysis of Linear and Nonlinear Regression Models by Maximum Likelihood," *Journal of Econometrics* 19 (1982), 239–285.

—— AND L. VASTENOU, "Prijzen en Hoeveelheden van Arbeid, Kapitaal, Energie en Produktie in Acht Sektoren van de Nederlandse Industrie, 1958–1976," Report AE10/78, Institute of Actuarial Science and Econometrics, University of Amsterdam (1978).

—— AND A. D. WOODLAND, "Some Matrix Results Useful for Multivariate Two Error Components Analysis," unpublished report (1987).

—— AND ——, "Interfuel Substitution in Dutch Manufacturing," *Applied Economics* 19 (1987b), 1639–1664.

MORAN, P. A. P., "Maximum-Likelihood Estimation in Non-standard Conditions," *Proceedings of the Cambridge Philosophical Society* 70 (1971), 441–450.

NETHERLANDS CENTRAL BUREAU OF STATISTICS, *De Nederlandse Energiehuis-houding, uitkomsten van maand- en kwartaaltellingen*, published quarterly since 1960.

PAGANO, M., "Estimation of Models of Autoregressive Signal Plus White Noise," *The Annals of Statistics* 2 (1974), 99–108.

PINDYCK, R. S., "Interfuel Substitution and the Industrial Demand for Energy: An International Comparison," *Review of Economics and Statistics* 61 (1979), 169–179.

REVANKAR, N. S., "Error Component Models with Serially Correlated Time Effects," *Journal of the Indian Statistical Association* 17 (1979), 137–160.

——, "Analysis of Regressions Containing Serially Correlated and Serially Uncorrelated Error Components," *International Economic Review* 21 (1980), 185–199.

SEVESTRE, P. AND A. TROGNON, "Propriétés de grands échantillons d'une classe d'estimateurs des modèles autorégressifs à erreurs composées," *Annales de l'INSEE* 50 (1983), 25–47.

—— AND ——, "A Note on Autoregressive Error Components Models," *Journal of Econometrics* 28 (1985), 231–245.

SHEPHARD, R. W., *Production and Cost Functions* (Princeton: Princeton University Press, 1970).