Journal of Econometrics 58 (1993) 315–346. North-Holland

Seemingly unrelated regressions under additive heteroscedasticity

Theory and share equation applications

David M. Mandy
University of Tennessee, Knoxville, TN 37996-0550, USA

Carlos Martins-Filho
Oregon State University, Corvallis, OR 97331-3612, USA

Received January 1991, final version received April 1992

We derive consistent, asymptotically efficient, and asymptotically normal estimators for SUR systems that have additive heteroscedastic contemporaneous correlation. Both our estimator for the location vector and the parameters of the covariance matrix possess these properties. The procedure is superior to other methods because we use GLS to estimate the parameters of the covariance matrix. Our method also permits the use of cross-equation parameter restrictions. We discuss how this type of heteroscedasticity arises naturally in share equation systems and random coefficient models, and how these models can be uniquely estimated with our two-step estimation technique.

1. Introduction

Zellner's (1962) Seemingly Unrelated Regressions (SUR) model assumes that the contemporaneous correlation across equations is homoscedastic. Just as in single-equation models, homoscedasticity in an SUR model is sometimes an untenable assumption. For example, Chavas and Segerson (1987) show that consistent stochastic specifications of share equation systems often result in a heteroscedastic contemporaneous correlation matrix. Using their results, we argue below that heteroscedasticity is unavoidable in an efficient share equation estimation. Thus, a specification error is committed whenever share equation systems are estimated with traditional SUR, resulting in inefficient estimates and invalid inference procedures. Random coefficients models also suffer from

Correspondence to: David M. Mandy, Department of Economics, 505A Stokely Management Center, University of Tennessee, Knoxville, TN 37996-0550, USA.

0304-4076/93/$06.00

heteroscedasticity. This is the main obstacle to be overcome in estimating the Hildreth–Houck (1968) model and its successors, one of which is the Singh–Ullah (1974) SUR model with random coefficients. Random effects panel data models are essentially random coefficients models with a particular structure, and we show below how some of these models have covariance structures that fit the heteroscedastic SUR framework.

Our primary purpose is to derive an efficient SUR estimator that accommodates the 'additive' heteroscedastic structure. This structure arises naturally in share equation systems and random coefficients models. We show that our estimator for this prevalent heteroscedastic structure is asymptotically efficient, consistent, and asymptotically normal, so the usual inference procedures apply. Moreover, we propose a SUR generalization of Amemiya's (1977) efficient estimator for the parameters of the covariance matrix in a single-equation additive heteroscedastic model. We show that Amemiya's consistency and asymptotic normality persist in our SUR version of this estimator, thereby providing a basis for hypothesis tests to detect heteroscedasticity in SUR models.

Aside from Singh and Ullah, several other authors discuss heteroscedasticity in SUR models. Kmenta and Gilbert (1968) do not propose an estimator that adjusts for heteroscedasticity, but they find via Monte Carlo experiment that Zellner's SUR estimator in the presence of contemporaneous correlation remains more efficient in small samples than Ordinary Least Squares (OLS) and maximum likelihood estimators even when there is heteroscedasticity. Duncan (1983) considers estimation of a heteroscedastic SUR model when the heteroscedasticity is nonparametric. He finds that substantial efficiency gains may occur in share equation estimation if heteroscedasticity is present and accommodated by the estimation technique. Low (1983) derives the mean square error matrix of Zellner's SUR estimator under some specialized assumptions, including heteroscedasticity of two regimes. Numerical evaluation of this matrix in some simple SUR models verifies the conclusion of Kmenta and Gilbert that Zellner's estimator outperforms OLS when there is contemporaneous correlation despite the presence of heteroscedasticity. However, Low also finds that single-equation Estimated Generalized Least Squares (EGLS) that accommodates the heteroscedasticity may be superior to Zellner's estimator even though single-equation EGLS ignores the contemporaneous correlation. Whether single-equation EGLS is superior to Zellner's estimator depends on the severity of the contemporaneous correlation vis-a-vis the heteroscedasticity. Finally, Srivastava and Giles (1987) summarize the earlier work of Low (1982) in a heteroscedastic SUR model with a fixed number of regimes. Results are similar to those of Low (1983).

These results suggest the need for SUR estimators that explicitly adjust for heteroscedasticity, but only Duncan (1983) and Singh and Ullah (1974) propose such estimators. Duncan's approach is nonparametric. Thus, if theory provides insight into the correct functional form better estimates should be obtained by using the appropriate parametric model [see Judge et al. (1985, p. 455)]. Hence, we begin with a motivational example that shows how Chavas and Segerson's (1987) results can lead to share equation systems that fit the additive heteroscedastic form. General application of Chavas and Segerson's results is further discussed after our estimators are derived, and we also discuss how our two-step procedure may be used to obtain estimates that are invariant to the equation deleted in a share equation system. This suggests that our estimator is preferable to Duncan's when working with a share equation system or random coefficients model.

There are also advantages of our estimator over the Singh–Ullah estimation procedure. First, the Singh–Ullah random coefficients model is a special case of our more general additive heteroscedastic model, so our model applies to a wider variety of situations. Second, our covariance matrix estimators are more efficient asymptotically because we follow Amemiya in using Generalized Least Squares (GLS) to estimate the parameters of the covariance matrix. This means that we estimate the entire set of covariance matrix parameters as one system in the SUR model, thereby accommodating the heteroscedasticity discussed by Amemiya as well as the contemporaneous correlation present in the covariance equations. This approach also permits the use of cross-equation restrictions when estimating the covariance equations, a situation which arises in share equation systems. Of course, efficient estimation of the covariance matrix parameters cannot improve the asymptotic efficiency of an already efficient EGLS estimator of the location vectors. However, efficient covariance matrix estimators may yield better small sample properties for the EGLS location vectors estimators [see Judge et al. (1985, pp. 435–437) and Amemiya (1977)]. Moreover, efficient covariance matrix estimators are of independent interest since, if their asymptotic distributions are known, they can be used to conduct hypothesis tests designed to detect heteroscedasticity.

In contrast, Singh and Ullah follow Hildreth and Houck in using OLS to estimate each covariance equation separately. Since the covariance equations do not satisfy the OLS assumptions and there is contemporaneous correlation, this method is inefficient relative to our approach and does not accommodate cross-equation parameter restrictions. Moreover, Crockett (1985) argues that Singh and Ullah's proof of asymptotic normality is flawed. Since the Singh–Ullah model is a special case of the model considered here, our results verify their conclusions even if their arguments are in error. Our results are obtained using the general method suggested by Crockett to prove asymptotic equivalence of GLS and EGLS estimators in heteroscedastic models. This method relies on a theorem originally due to Carroll and Ruppert (1982) and is useful in the present context for proving asymptotic normality of both our EGLS estimator of the location vectors and our EGLS estimator for the parameters of the covariance matrix.

2. A motivational example

Contemporaneous additive heteroscedasticity in SUR models may be a reasonable assumption in a number of situations, but it arises naturally in the econometric modeling of producer behavior. Accordingly, this section presents a simple cost function estimation to motivate the need for an SUR estimator that adjusts for additive heteroscedasticity. The example is designed to demonstrate that even with simple stochastic assumptions the equations to be estimated form an SUR system with heteroscedasticity of the additive form when a consistent stochastic specification is explicitly incorporated.

Assume for simplicity that firm t produces a single output using two inputs. Further, follow Chavas and Segerson (1987) in assuming that, at least from the viewpoint of the econometrician, each firm's cost function consists of additively separable deterministic and stochastic components and that the stochastic component is linear in the random variables. Hence, if the deterministic part of the production technology for firm t is represented by a translog cost function, we have

$$y_{t0} = \beta_0 + \beta_1 p_{t1} + \beta_2 p_{t2} + \beta_z z_t + \tfrac{1}{2}[\beta_{11} p_{t1}^2 + \beta_{22} p_{t2}^2 + 2\beta_{12} p_{t1} p_{t2}]$$
$$+ \beta_{1z} p_{t1} z_t + \beta_{2z} p_{t2} z_t + \tfrac{1}{2}\beta_{zz} z_t^2 + \theta_t' H(p_{t1}, p_{t2}, z_t),$$

where y_{t0} is logarithmic cost, p_{t1} and p_{t2} are the logarithmic input prices, z_t is logarithmic output, and $\theta_t' H(p_{t1}, p_{t2}, z_t)$ is Chavas and Segerson's stochastic component. Here, θ_t is a random vector containing all of the random variables in the model, while H is a vector-valued function of conformable dimension.

The share equations follow from the cost function in the usual manner by applying Shephard's Lemma, but since the shares are related to the cost function in this manner the 'error' terms in the share equations must derive from the only stochastic term present in the cost function, $\theta_t' H(p_{t1}, p_{t2}, z_t)$, through logarithmic differentiation with respect to the input prices. Hence, H must depend on p_1 and p_2 as denoted. Chavas and Segerson prove this formally, and call the resulting error structure for the whole system a consistent stochastic specification. Thus, the share equations are given by

$$y_{t1}(p_{t1}, p_{t2}, z_t) = \beta_1 + \beta_{11} p_{t1} + \beta_{12} p_{t2} + \beta_{1z} z_t + \theta_t' \left[\frac{\partial H(p_{t1}, p_{t2}, z_t)}{\partial p_{t1}} \right],$$

$$y_{t2}(p_{t1}, p_{t2}, z_t) = \beta_2 + \beta_{12} p_{t1} + \beta_{22} p_{t2} + \beta_{2z} z_t + \theta_t' \left[\frac{\partial H(p_{t1}, p_{t2}, z_t)}{\partial p_{t2}} \right],$$

where y_{ti} is the share of input i in total cost and $\partial H / \partial p_{ti}$ is the vector of derivatives of components of H with respect to p_{ti}. Each observation (firm) of the system can be written $y_t = X_t\beta + u_t$, where

$$y_t = (y_{t1} \quad y_{t2} \quad y_{t0})',$$

$$X_t = \begin{pmatrix} 0 & 1 & 0 & 0 & p_{t1} & 0 & p_{t2} & z_t & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & p_{t2} & p_{t1} & 0 & z_t & 0 \\ 1 & p_{t1} & p_{t2} & z_t & \frac{1}{2}p_{t1}^2 & \frac{1}{2}p_{t2}^2 & p_{t1}p_{t2} & p_{t1}z_t & p_{t2}z_t & \frac{1}{2}z_t^2 \end{pmatrix},$$

$$\beta = (\beta_0 \quad \beta_1 \quad \beta_2 \quad \beta_z \quad \beta_{11} \quad \beta_{22} \quad \beta_{12} \quad \beta_{1z} \quad \beta_{2z} \quad \beta_{zz})',$$

$$u_t = \left(\theta_t' \frac{\partial H}{\partial p_{t1}} \quad \theta_t' \frac{\partial H}{\partial p_{t2}} \quad \theta_t' H \right)'.$$

To examine whether the covariance matrix of this system takes the SUR form with additive heteroscedasticity, assumptions are required on the stochastic component $\theta_t' H$. It is customary in these models to assume that there is one random variable for each equation and that these random variables are contemporaneously correlated but independent across observations. Thus, assume θ_t is three-dimensional with $\theta_t \sim \mathrm{IID}(0, \{\sigma_{ij}\}_{i,j=1}^3)$. If the share equations are to be stochastic (i.e., $\partial H / \partial p_{ti} \neq 0$), then the simplest available assumptions on H are linearity in the prices and independence from output. An H function with these properties that also satisfies the cost exhaustion condition discussed in section 6 below is $H = (\gamma_0, \gamma_1(p_{t1} - p_{t2}), \gamma_2(p_{t1} - p_{t2}))'$, and this yields an error vector for the three equations of firm t given by

$$u_t = (\theta_{t2}\gamma_1 + \theta_{t3}\gamma_2, -\theta_{t2}\gamma_1 - \theta_{t3}\gamma_2,$$
$$\theta_{t1}\gamma_0 + (\theta_{t2}\gamma_1 + \theta_{t3}\gamma_2)(p_{t1} - p_{t2}))',$$

where θ_{ti} is the ith component of θ_t. With this specification $\mathrm{E}(u_t u_\tau') = 0$ for $t \neq \tau$ and the contemporaneous correlation matrix for observation t is

$$\mathrm{E}(u_t u_t') = \Omega_t = \begin{pmatrix} \alpha_1 & -\alpha_1 & \alpha_2 + \alpha_1(p_{t1} - p_{t2}) \\ -\alpha_1 & \alpha_1 & -\alpha_2 - \alpha_1(p_{t1} - p_{t2}) \\ \alpha_2 + \alpha_1(p_{t1} - p_{t2}) & -\alpha_2 - \alpha_1(p_{t1} - p_{t2}) & \alpha_3 + 2\alpha_2(p_{t1} - p_{t2}) + \alpha_1(p_{t1} - p_{t2})^2 \end{pmatrix},$$

where $\alpha_1 = \sigma_{22}\gamma_1^2 + 2\sigma_{23}\gamma_1\gamma_2 + \sigma_{33}\gamma_2^2$, $\alpha_2 = \sigma_{12}\gamma_0\gamma_1 + \sigma_{13}\gamma_0\gamma_2$, and $\alpha_3 = \sigma_{11}\gamma_0^2$. Thus, even with this simple consistent stochastic specification the contemporaneous correlation matrix is heteroscedastic because it depends on the input prices, which vary across observations. This occurs whenever the cost function is included in a consistent stochastic specification, because H must depend on the prices if the shares are to be stochastic. Hence, an input share system including the cost function is unavoidably heteroscedastic. While these systems are sometimes estimated excluding the cost function, this practice is inefficient because of the heteroscedasticity and the contemporaneous correlation between the share equations and the cost function. Therefore, efficient estimation requires a heteroscedastic SUR estimation even in the simplest case. Moreover, for simple specifications of H the heteroscedasticity assumes the 'additive' form since each element of the covariance matrix can be expressed as a linear function of an unknown parameter vector like the vector $\alpha = (\alpha_1, \alpha_2, \alpha_3)'$ for the present example.

The restrictions needed to ensure a well-behaved cost function may imply relationships between elements of α as well as the usual restrictions among the elements of β, in addition to the restrictions arising due to the occurrence of a particular parameter in more than one element of Ω_t. These relationships generally include cross-equation parameter restrictions, so joint estimation of β and joint estimation of α are required both because of parameter restrictions and because of the heteroscedasticity, even if there are no efficiency gains to traditional SUR (for example, if the matrix of independent variables is the same in each equation). These restrictions also imply that Ω_t is singular, a familiar problem that is discussed in section 6 below.

3. The SUR model under additive heteroscedasticity

Arrange T observations on m equations as

$$y_t = X_t\beta + u_t, \quad t = 1, \dots, T, \tag{1}$$

where y_t is an observable m-dimensional dependent vector,

$$X_t = \begin{bmatrix} x'_{t1} & 0' & \cdots & 0' \\ 0' & x'_{t2} & \cdots & 0' \\ \vdots & \vdots & \ddots & \vdots \\ 0' & \cdots & 0' & x'_{tm} \end{bmatrix}, \tag{2}$$

x'_{ti} is a K_i-dimensional vector of nonstochastic explanatory variables for equation i in observation t and the zero vectors are of conformable dimensions, β is a $K = \sum_{i=1}^m K_i$-dimensional vector of unknown nonstochastic parameters containing m subvectors of dimensions K_i, and the u_t's are independent m-dimensional unobservable random vectors. Zellner's stochastic specification is

$E(u_t) = 0$, $E(u_t u_\tau') = 0$ for $t \neq \tau$, and $E(u_t u_t')$ a fixed but unknown matrix for $t = 1, 2, \ldots$. We adopt the first two assumptions but depart from Zellner by assuming

$$E(u_t u_t') = \Omega_t = \begin{bmatrix} \sigma_{11}^t & \sigma_{12}^t & \cdots & \sigma_{1m}^t \\ \sigma_{21}^t & \sigma_{22}^t & \cdots & \sigma_{2m}^t \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}^t & \sigma_{m2}^t & \cdots & \sigma_{mm}^t \end{bmatrix}, \tag{3}$$

so that the contemporaneous correlation matrix is permitted to vary across observations. We also assume that the vectors u_t possess a multivariate normal distribution. As Jobson and Fuller (1980) discuss, our estimators can be constructed and possess the same asymptotic properties even if the error vectors u_t are not normal, but some known relationship between the first, second, third, and fourth moments is needed for estimation purposes.

Letting $y' = [y_1' \ldots y_T']$, $X' = [X_1' \ldots X_T']$, $u' = [u_1' \ldots u_T']$, and $\Omega =$ block diag$\{\Omega_1 \ldots \Omega_T\}$, we may write the entire system as

$$y = X\beta + u, \qquad E(u) = 0, \qquad E(uu') = \Omega. \tag{4}$$

The BLUE estimator for β is the usual GLS estimator

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y,$$

but in most cases Ω is unknown, and so we must replace it with an estimator $\hat{\Omega}$ to obtain the EGLS estimator

$$\hat{\beta}_{\text{EGLS}} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y.$$

It is well-known that no efficiency gain results from joint estimation in the traditional SUR context if $x_{t1} = x_{t2} = \ldots = x_{tm}$ for all t. However, the joint estimator $\hat{\beta}_{GLS}$ for model (4) is superior to individual equation estimation even if the independent variables are the same in each equation, since GLS corrects for the heteroscedastic covariance matrix.

The heteroscedasticity assumes the 'additive' functional form $\sigma_{ij}^t = \alpha_{ij}' z_{ij}^t$ for $t = 1, 2, \ldots$ and $i, j = 1, \ldots, m$, where α_{ij} is a G_{ij}-dimensional vector of unknown nonstochastic parameters and z_{ij}^t is a conformable vector of nonstochastic explanatory variables. Since Ω_t is symmetric, there are only $m(m+1)/2$ distinct σ_{ij}^t elements, implying that $\alpha_{ij} = \alpha_{ji}$ unless different explanatory variables are involved in σ_{ij}^t than in σ_{ji}^t. However, if $z_{ij}^t \neq z_{ji}^t$, then restrictions that involve the explanatory variables would generally be required on the estimation of α_{ij} and α_{ji} to assure symmetry of $\hat{\Omega}_t$. In all applications that we are aware of this situation does not arise, so henceforth we assume that $z_{ij}^t = z_{ji}^t$ and $\alpha_{ij} = \alpha_{ji}$, implying that there are only $m(m+1)/2$ distinct α_{ij} vectors. Hence, we may define the $G = \sum_{i=1}^{m}\sum_{j=i}^{m} G_{ij}$-dimensional vector of covariance matrix parameters by

$$\alpha' = [\alpha_{11}' \; \alpha_{21}' \; \alpha_{22}' \cdots \alpha_{mm-1}' \; \alpha_{mm}'].$$

Denoting an arbitrary estimator for α by $\hat{\alpha}$, we can estimate σ_{ij}^t by $\hat{\sigma}_{ij}^t = \hat{\alpha}_{ij}' z_{ij}^t$ and use these estimators to construct $\hat{\Omega}_t$ and $\hat{\Omega}$. Under well-known regularity conditions [see Schmidt (1976, chs. 1, 2)], satisfied by our assumption A.2 below, the GLS estimator has asymptotic distribution

$$T^{1/2}(\hat{\beta}_{\mathrm{GLS}} - \beta) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \lim_{T\to\infty} (T^{-1}X'\Omega^{-1}X)^{-1}\right).$$

Since the GLS estimator is efficient, the asymptotic efficiency, consistency, and asymptotic normality of the EGLS estimator based on any $\hat{\alpha}$ can be established simultaneously by showing that $\hat{\beta}_{\mathrm{EGLS}}$ converges in probability to $\hat{\beta}_{\mathrm{GLS}}$. Given the regularity conditions for asymptotic normality of $\hat{\beta}_{\mathrm{GLS}}$, sufficient conditions for the convergence in probability of $\hat{\beta}_{EGLS}$ and $\hat{\beta}_{GLS}$ are [see Judge et al. (1985, p. 176), Schmidt (1976, p. 71), or Theil (1971, p. 399)]:

$$\operatorname*{plim}_{T\to\infty} T^{-1}X'\hat{\Omega}^{-1}X = \lim_{T\to\infty} T^{-1}X'\Omega^{-1}X, \tag{5}$$

$$\operatorname*{plim}_{T\to\infty} T^{-1/2}X'(\hat{\Omega}^{-1} - \Omega^{-1})u = 0. \tag{6}$$

Thus, our main problem is to find an estimator $\hat{\alpha}$ such that eqs. (5) and (6) hold, in which case there is no cost, asymptotically, to using EGLS instead of GLS. Among the class of estimators $\hat{\alpha}$ that satisfy (5) and (6), a secondary problem is to find an efficient estimator since, as Amemiya and Judge observe, this should lead to more efficient small sample estimates for β. Moreover, as mentioned at the outset, efficient estimators for α are of independent interest since, if their asymptotic distributions are known, they can be used to conduct hypothesis tests designed to detect heteroscedasticity.

4. The estimation procedure

Estimate α by first obtaining the OLS residuals $\hat{u} = u - X(X'X)^{-1}X'u$. This can be accomplished by performing OLS on eq. (4) or by rearranging the model by equations and performing OLS on each equation separately. However, estimation of eq. (4) permits the use of cross-equation parameter restrictions if appropriate, as is the case with share equation systems. The residual subvector for each observation is

$$\hat{u}_t = u_t - X_t(X'X)^{-1}X'u, \qquad t = 1, \dots, T. \tag{7}$$

Next, denote the ith component of u_t by u_{ti} and analogously for $\hat{u}_t$, and form the $m(m+1)/2$-dimensional vectors of cross-products

$$\begin{aligned}
\hat{e}_t' &= [\hat{u}_{t1}\hat{u}_{t1}\ \hat{u}_{t2}\hat{u}_{t1}\ \hat{u}_{t2}\hat{u}_{t2} \cdots \hat{u}_{tm}\hat{u}_{tm-1}\ \hat{u}_{tm}\hat{u}_{tm}],\\
e_t' &= [u_{t1}u_{t1}\ u_{t2}u_{t1}\ u_{t2}u_{t2} \cdots u_{tm}u_{tm-1}\ u_{tm}u_{tm}],\\
\sigma_t' &= [\sigma^t_{11}\ \sigma^t_{21}\ \sigma^t_{22} \cdots \sigma^t_{mm-1}\ \sigma^t_{mm}],\\
v_t &= e_t - \sigma_t,\\
\varepsilon_t &= \hat{e}_t - e_t.
\end{aligned} \tag{8}$$

Denoting

$$Z_t = \begin{bmatrix}
z^{t'}_{11} & 0' & 0' & \cdots & 0' & 0'\\
0' & z^{t'}_{21} & 0' & \cdots & 0' & 0'\\
0' & 0' & z^{t'}_{22} & 0' & \cdots & 0'\\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\
0' & 0' & \cdots & 0' & z^{t'}_{mm-1} & 0'\\
0' & 0' & \cdots & 0' & 0' & z^{t'}_{mm}
\end{bmatrix}, \tag{9}$$

we have $\sigma_t = Z_t\alpha$ so that

$$\hat{e}_t = Z_t\alpha + v_t + \varepsilon_t. \tag{10}$$

Letting $\hat{e}' = [\hat{e}_1' \dots \hat{e}_T']$, $e' = [e_1' \dots e_T']$, $\sigma' = [\sigma_1' \dots \sigma_T']$, $v = e - \sigma$, $\varepsilon = \hat{e} - e$, and $Z' = [Z_1' \dots Z_T']$, we have $\sigma = Z\alpha$ and may therefore write the entire system as

$$\hat{e} = Z\alpha + v + \varepsilon. \tag{11}$$

Note that $\mathrm{E}(v) = 0$ and $\mathrm{E}(vv') = S = \text{block diag}\{S_1, \dots, S_T\}$, where

$$S_t = \mathrm{E}((e_t - \sigma_t)(e_t - \sigma_t)') = \mathrm{E}(e_t e_t') - \sigma_t\sigma_t'.$$

The multivariate normal distribution yields

$$\mathrm{E}(u_{ti}u_{tj}u_{tk}u_{tl}) = \sigma^t_{ij}\sigma^t_{kl} + \sigma^t_{ik}\sigma^t_{jl} + \sigma^t_{jk}\sigma^t_{il},$$

so

$$S_t = \begin{bmatrix} \sigma^t_{11}\sigma^t_{11} + \sigma^t_{11}\sigma^t_{11} & \sigma^t_{12}\sigma^t_{11} + \sigma^t_{12}\sigma^t_{11} & \sigma^t_{12}\sigma^t_{12} + \sigma^t_{12}\sigma^t_{12} & \cdots & \sigma^t_{1m}\sigma^t_{1m} + \sigma^t_{1m}\sigma^t_{1m} \\ \sigma^t_{21}\sigma^t_{11} + \sigma^t_{11}\sigma^t_{21} & \ddots & \cdots & \cdots & \sigma^t_{2m}\sigma^t_{1m} + \sigma^t_{1m}\sigma^t_{2m} \\ \sigma^t_{21}\sigma^t_{21} + \sigma^t_{21}\sigma^t_{21} & \cdots & \ddots & \cdots & \sigma^t_{2m}\sigma^t_{2m} + \sigma^t_{2m}\sigma^t_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma^t_{m1}\sigma^t_{m1} + \sigma^t_{m1}\sigma^t_{m1} & \cdots & \cdots & \cdots & \sigma^t_{mm}\sigma^t_{mm} + \sigma^t_{mm}\sigma^t_{mm} \end{bmatrix}.$$

Eq. (11) provides a basis for estimating α. Unfortunately, the 'error' term $v + \varepsilon$ has neither zero mean nor scalar identity covariance, so OLS is biased and inefficient. Amemiya overcomes these problems in the single-equation context, at least asymptotically, by applying EGLS to eq. (11), ignoring the effects of ε. This method corrects for the heteroscedasticity of v. He shows that ε does not affect the asymptotic distribution of the estimator, so it can be ignored if asymptotic properties are the only concern.

Singh and Ullah organize a variant of (11) by equation in the random coefficients SUR context, rather than by observation, and then apply OLS to each equation individually. This method disregards the heteroscedasticity of v. Moreover, in the SUR context eq. (11) constitutes a second heteroscedastic SUR system and should therefore be estimated jointly rather than by individual equations since joint EGLS corrects for both heteroscedasticity and contemporaneous correlation, provided that the effects of ε remain negligible as $T \to \infty$. Joint estimation also permits the introduction of cross-equation restrictions of the type mentioned in section 2.

The GLS estimator for eq. (11) obtained by ignoring ε is

$$\hat{\alpha}_{\mathrm{GLS}} = (Z'S^{-1}Z)^{-1}Z'S^{-1}\hat{e},$$

while an EGLS estimator is obtained by substituting an estimator $\hat{S}$ for S:

$$\hat{\alpha}_{\mathrm{EGLS}} = (Z'\hat{S}^{-1}Z)^{-1}Z'\hat{S}^{-1}\hat{e}.$$

To obtain $\hat{S}$, we follow Amemiya in using the OLS estimator applied to eq. (11),

$$\hat{\alpha}_{\text{OLS}} = (Z'Z)^{-1}Z'\hat{e},$$

which leads to preliminary estimates of the σ_{ij}^t's and hence $\hat{S}$. Both $\hat{\alpha}_{\text{EGLS}}$ and $\hat{\alpha}_{\text{OLS}}$ can incorporate cross-equation restrictions, if appropriate. We show in the next section that the effects of ε remain negligible in the SUR context, so $\hat{\alpha}_{\text{GLS}}$ is consistent and asymptotically normal. Then we show that $\hat{\alpha}_{\text{EGLS}}$ converges in probability to $\hat{\alpha}_{\text{GLS}}$, so EGLS in the SUR context retains the desirable properties discussed by Amemiya for the single-equation case. Finally, we show that eqs. (5) and (6) hold when $\hat{\beta}_{\text{EGLS}}$ is based on $\hat{\alpha}_{\text{EGLS}}$.

One potential problem warrants mention before proceeding to the asymptotic properties. Monte Carlo studies using single-equation models of additive heteroscedasticity show that these models may produce estimated covariance matrices that are not positive definite [see, for example, Raj (1975)], in which case a substantial degradation of the EGLS estimates for β occurs. Unfortunately, there is no analytic solution to this problem within a linear estimation framework. Applied researchers should check $\hat{S}$ and $\hat{\Omega}$ to assure that they are positive definite. If a problem is encountered, the first solution should be to obtain additional data since the asymptotic results assure that the estimates approach the true positive definite matrices as the sample size tends to infinity. The Monte Carlo results verify that the problem rarely occurs with large samples. If additional data is unobtainable, it may be possible to solve the problem analytically with nonlinear constraints or a nonlinear reparameterization of the model. However, in this case, an approach that accommodates nonlinear covariance equations would be needed. Magnus (1978) discusses a maximum likelihood approach and Judge et al. (1985, pp. 435–437, 808) provide a summary of the methods available to assure positive definiteness.

5. Properties of the estimators

Denote the kth element of x_{ti} by x_{tik} and the kth element of z_{ij}^t by z_{ijk}^t for $i, j = 1, \ldots, m$ and $t = 1, 2, \ldots$ Our results rely on the following standard regularity assumptions:

A.1. There exist upper bounds $B_x < \infty$ and $B_z < \infty$ such that $|x_{tik}| < B_x$ and $|z_{ijk}^t| < B_z$ for $t = 1, 2, \ldots$, $i, j = 1, \ldots, m$, and $k = 1, \ldots, K_i$ (or G_{ij}, as appropriate).

A.2. $T^{-1}X'X$, $T^{-1}X'\Omega X$, $T^{-1}X'\Omega^{-1}X$, $T^{-1}Z'Z$, $T^{-1}Z'SZ$, and $T^{-1}Z'S^{-1}Z$ all converge to finite positive definite matrices as $T \to \infty$. These limits are denoted by Q_X, $Q_{\Omega X}$, $Q_{\Omega^{-1}X}$, etc.

A.3. There exists a lower bound $\varepsilon_0 > 0$ such that Det $\Omega_t \geq \varepsilon_0$ and Det $S_t \geq \varepsilon_0$ for $t = 1, 2, \ldots$.

Asymptotic normality of $\hat{\alpha}_{\mathrm{GLS}}$ and $\hat{\alpha}_{\mathrm{OLS}}$ follow at once under these conditions from the following generalization of Amemiya's theorem (all proofs are contained in the appendix).

Theorem 1. Let $\hat{\alpha} = (A'Z)^{-1}A'\hat{e}$ *be an estimator for* α, *where* A *is a* $(Tm(m+1)/2 \times G)$ *matrix with elements bounded by* B_a *as* $T \to \infty$. *Assume further that* $T^{-1}A'Z$ *converges to a finite positive definite matrix* Φ *as* $T \to \infty$. *Then*

$$T^{1/2}(\hat{\alpha} - \alpha) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \Phi^{-1}\left(\lim_{T\to\infty} T^{-1}A'SA\right)\Phi^{-1}\right),$$

provided this covariance limit exists.

Letting $A = Z$ in Theorem 1 shows that

$$T^{1/2}(\hat{\alpha}_{\mathrm{OLS}} - \alpha) \xrightarrow{\mathrm{d}} \mathrm{N}(0, Q_Z^{-1}Q_{SZ}Q_Z^{-1}), \tag{12}$$

while letting $A = S^{-1}Z$ shows that

$$T^{1/2}(\hat{\alpha}_{\mathrm{GLS}} - \alpha) \xrightarrow{\mathrm{d}} \mathrm{N}(0, Q_{S^{-1}Z}^{-1}). \tag{13}$$

Eq. (13) also provides the limiting distribution of $\hat{\alpha}_{\mathrm{EGLS}}$ if $\hat{\alpha}_{\mathrm{EGLS}}$ converges in probability to $\hat{\alpha}_{\mathrm{GLS}}$. Given assumption A.2, sufficient conditions for this convergence, analogous to eqs. (5) and (6), are

$$\operatorname*{plim}_{T\to\infty} T^{-1}Z'\hat{S}^{-1}Z = Q_{S^{-1}Z}, \tag{14}$$

$$\operatorname*{plim}_{T\to\infty} T^{-1/2}Z'(\hat{S}^{-1} - S^{-1})(v + \varepsilon) = 0. \tag{15}$$

These conditions, as well as eqs. (5) and (6), involve the limits of the inverses of covariance matrices. The elements of these matrices can be expressed as the ratio of a cofactor to the determinant, but we need to consider these elements evaluated at various possible values for α. Thus, let $H_t(\gamma)$ be the matrix that results from substituting γ in place of α in either Ω_t or S_t, depending upon whether we are examining $\hat{\beta}_{\text{EGLS}}$ or $\hat{\alpha}_{\text{EGLS}}$. Furthermore, let $\Delta_t(\gamma) = \text{Det}H_t(\gamma)$ and $C^t_{ij}(\gamma)$ be the cofactor of the (i, j) element of $H_t(\gamma)$. The following fundamental properties of these functions are used to establish eqs. (5), (6), (14), and (15).

Lemma 1. *Δ_t and C^t_{ij} are uniformly (in t) continuous in γ.*

Corollary. *For every $C \in [0, \infty)$ there exists T_C (independent of t) such that*

$$T > T_C \quad \text{and} \quad \|\gamma\| \le C \Rightarrow |\Delta_t(T^{-1/2}\gamma + \alpha)| > \varepsilon_0/2.$$

Lemma 2. *For every $C \in [0, \infty)$ there exists a bound $B_{H,C} > 0$ such that*

$$\|\gamma\| \le C \Rightarrow \begin{cases} |\Delta_t(T^{-1/2}\gamma + \alpha)| \le B_{H,C} \\ |C^t_{ij}(T^{-1/2}\gamma + \alpha)| \le B_{H,C} \end{cases}$$

for every $t, T = 1, 2, \ldots$ and $i, j = 1, \ldots, m$ [or $m(m+1)/2$, as appropriate].

Eqs. (5) and (14) are an immediate consequence of the following theorem.

Theorem 2. *Let A be a $(Tr \times q)$ matrix with elements bounded by B_a as $T \to \infty$, and $H_t(\gamma)$ be $(r \times r)$ matrices satisfying Lemmas 1 and 2 and the Corollary. Moreover, assume*

$$\lim_{T\to\infty} T^{-1} A'[\text{block diag}\{H_1(\alpha), \ldots, H_T(\alpha)\}]^{-1} A = \Phi,$$

a finite positive definite matrix. If $\hat{\alpha}$ is a consistent estimator for α, then

$$\operatorname*{plim}_{T\to\infty} T^{-1} A'[\text{block diag}\{H_1(\hat{\alpha}), \ldots, H_T(\hat{\alpha})\}]^{-1} A = \Phi.$$

Verify eq. (5) when $\hat{\Omega}$ is based on $\hat{\alpha}_{\text{EGLS}}$ by assuming for the moment that $\hat{\alpha}_{\text{EGLS}}$ is consistent. Set $r = m$, $A = X$, $B_a = B_x$, $\Phi = Q_{\Omega^{-1}X}$, and let $H_t(\gamma)$ be the matrix that results from substituting γ in place of α in Ω_t, so that $H_t(\alpha) = \Omega_t$ and $H_t(\hat{\alpha}_{\text{EGLS}}) = \hat{\Omega}_t$. Then the conclusion of Theorem 2 yields eq. (5) once we show that $\hat{\alpha}_{\text{EGLS}}$ is consistent (below). Eq. (14) is verified by noting that $\hat{\alpha}_{\text{OLS}}$ is consistent by (12). Hence, set $r = m(m+1)/2$, $A = Z$, $B_a = B_z$, $\Phi = Q_{S^{-1}Z}$, and let $H_t(\gamma)$ be the matrix that results from substituting γ in place of α in S_t, so that $H_t(\alpha) = S_t$ and $H_t(\hat{\alpha}_{\text{OLS}}) = \hat{S}_t$. Then Theorem 2 confirms eq. (14).

Eqs. (6) and (15) present more formidable problems. Schmidt (1976, pp. 68–70) shows that consistent estimators for the error covariance matrix in heteroscedastic models need not produce EGLS estimators that have the same asymptotic distribution as the GLS estimator. This difficulty is sometimes overlooked, as Crockett (1985) notes that several proposed 'proofs' of the asymptotic equivalence of EGLS and GLS estimators in random coefficients models are flawed. However, Crockett states a special case of a theorem originally due to Carroll and Ruppert (1982), and then uses it to establish asymptotic equivalence of EGLS and GLS in the Hildreth–Houck model. We show below that this approach can be used in the present context to establish eqs. (6) and (15). As Schmidt's example suggests, consistency of $\hat{\alpha}$ is not sufficient for an application of the Carroll–Ruppert Theorem, but if $\hat{\alpha}$ satisfies the stronger condition $\hat{\alpha} - \alpha = O_p(T^{-1/2})$, then the Carroll–Ruppert Theorem can be applied to establish eqs. (6) and (15). It is still possible that consistency of $\hat{\alpha}$ is sufficient to establish (6) and (15) using some other approach, because Schmidt's example does not fit the assumption that the heteroscedasticity takes the additive form. Hence, in the present context Schmidt's example is only suggestive that a condition stronger than consistency is needed, but we know of no correct proof that establishes (6) or (15) using only consistency of $\hat{\alpha}$. Currently, the Carroll–Ruppert Theorem and its stronger requirement appear to be the only correct approach. Alternatively, we know of no counter-example that demonstrates insufficiency of consistency in an additive heteroscedastic model. We first restate Crockett's version of the Carroll–Ruppert Theorem [for a proof, see Crockett (1983)], and then apply it to the current problem.

Theorem 3. (Carroll–Ruppert, Crockett). For $T = 1, 2, \ldots$ and $t = 1, \ldots, T$, and for every $\gamma \in \mathscr{R}^G$, let $A_{tT}(\gamma)$ be a $(q \times r)$ matrix. Suppose:

1. $A_{tT}(0) = 0$ for every t, T.

2. For every $C \in (0, \infty)$ there exists $T_C, B_C < \infty$ such that

$$\|\gamma_0\|, \|\gamma_1\| \leq C \quad \text{and} \quad T \geq T_C \Rightarrow$$

$$|||A_{tT}(\gamma_0) - A_{tT}(\gamma_1)||| \leq T^{-1/2} \|\gamma_0 - \gamma_1\| B_C,$$

where $|||A_{tT}||| = \max_{i,j} |a_{ij}|$.

Let $\gamma_T \in \mathscr{R}^G$ be random vectors for $T = 1, 2, \ldots$ such that

3. $\gamma_T = O_p(1)$.

Finally, let $w_t \in \mathscr{R}^r$ be independent random vectors for $t = 1, 2, \ldots$ such that

4. $E(w_t) = 0$ for every t.

5. $\sup_t \mathrm{E}(\|w_t\|^2) < \infty$.

Then, $\operatorname{plim}_{T\to\infty} T^{-1/2}\sum_{t=1}^{T} A_{tT}(\gamma_T)w_t = 0$.

Theorem 4. Let A be a $(Tr \times q)$ matrix with elements bounded by B_a as $T \to \infty$, and let $H_t(\gamma)$ be $(r \times r)$ matrices satisfying Lemmas 1 and 2 and the Corollary. Let w_t be independent r-dimensional random vectors with $\mathrm{E}(w_t) = 0$ *and* $\mathrm{E}(\|w_t\|^2) \le B_w < \infty$ *for every t. If $\hat{\alpha} - \alpha = \mathrm{O}_p(T^{-1/2})$, then*

$$\operatorname*{plim}_{T\to\infty} T^{-1/2} A'([\text{block diag}\{H_1(\hat{\alpha}), \dots, H_T(\hat{\alpha})\}]^{-1}$$

$$- [\text{block diag}\{H_1(\alpha), \dots, H_T(\alpha)\}]^{-1})[w'_1, \dots, w'_T]' = 0.$$

Verify eq. (6) when $\hat{\Omega}$ is based on $\hat{\alpha}_{\text{EGLS}}$ by assuming for the moment that $\hat{\alpha}_{\text{EGLS}} - \alpha = \mathrm{O}_p(T^{-1/2})$. Let $w_t = u_t$ and all other definitions be as in the verification of eq. (5). Since $\mathrm{E}(\|u_t\|^2) \le m\|\alpha\| G^{1/2} B_z$, the conclusion of Theorem 4 yields eq. (6) once we show that $\hat{\alpha}_{\text{EGLS}} - \alpha = \mathrm{O}_p(T^{-1/2})$ (below). By eqs. (12) and (13), eqs. (5) and (6) also hold when $\hat{\Omega}$ is based on $\hat{\alpha}_{\text{OLS}}$ or $\hat{\alpha}_{\text{GLS}}$. Eq. (15) is verified by noting that $\hat{\alpha}_{\text{OLS}} - \alpha = \mathrm{O}_p(T^{-1/2})$ by (12). Hence, let $w_t = v_t$ and all other definitions be as in the verification of eq. (14). Since $\mathrm{E}(\|v_t\|) \le m(m+1) \times \|\alpha\|^2 G B_z^2$, Theorem 4 yields

$$\operatorname*{plim}_{T\to\infty} Z'(\hat{S}^{-1} - S^{-1})v = 0.$$

However, the ε term does not satisfy the assumptions of the Carroll–Ruppert Theorem (and hence Theorem 4). Following Amemiya, we show that the effect of ε vanishes as $T \to \infty$ with the following theorem.

Theorem 5. $\operatorname{plim}_{T\to\infty} T^{-1/2} Z'(\hat{S}^{-1} - S^{-1})\varepsilon = 0$.

Since $\hat{\alpha}_{\text{EGLS}}$ satisfies eqs. (14) and (15), we have

$$T^{1/2}(\hat{\alpha}_{\text{EGLS}} - \alpha) \xrightarrow{\text{d}} \mathrm{N}(0, Q_{S^{-1}Z}^{-1}),$$

which verifies that $\hat{\alpha}_{\text{EGLS}} - \alpha = \mathrm{O}_p(T^{-1/2})$.

6. Applications

The model and estimation techniques described above may be applied in several settings. First, it should be noted that whenever heteroscedasticity is suspected one option for the researcher is to assume a parametric structure and then estimate the assumed structure. Our model provides one candidate structure. Of course, if this structure is assumed incorrectly, then efficiency loss may occur, and it is frequently the case that a researcher has little theoretical guidance on whether a particular parametric structure is appropriate. However, Monte Carlo evidence for single-equation heteroscedastic models suggests that assuming the wrong parametric structure may not entail large efficiency losses [see Surekha and Griffiths (1984)]. Hence, our method provides at least one parametric solution, that may be useful in a variety of settings, to the problem of heteroscedasticity in an SUR model.

General Share Equation Estimation. Another application for our model and estimation techniques is share equation systems, particularly models of producer behavior such as those discussed by Christensen and Greene (1976) and Christensen et al. (1975). Often, classical SUR is used to estimate these models, and parameter restrictions are used to assure that the dictates of economic theory are satisfied. However, the stochastic error terms may be incorrectly assumed to satisfy Zellner's specification, as discussed in section 2.

Chavas and Segerson's (1987) general approach begins with an objective function $F(X_t, \beta, \alpha, \theta_t)$, where X_t is a matrix of observable parameters to both the researcher and the economic agent, β and α are vectors of parameters that are unobservable by the researcher, and θ_t is a J-dimensional vector of unobservable zero-mean shift parameters that capture the behavior not explained by X_t, β, and α. The matrix X_t usually consists of a price vector p_t and output or income.

Optimization of F leads to $m-1$ share equations, which form a system with F that can be used to estimate β and α. These models are usually assumed to be linear in the errors θ_t, so Chavas and Segerson assume that the share equations take the separable form

$$y_{ti} = q_i(p_t, \beta, \alpha) + \theta_t' h_i(p_t, \beta, \alpha), \tag{16}$$

where y_{ti} is the share of the ith demand in total outlay for the tth observation. When the economic agent is a cost-minimizing firm, Chavas and Segerson show that (16) and homogeneity of F (the cost function) in p_t require the logarithmic cost function to take the form

$$y_{tm} = \ln F(X_t, \beta, \alpha, \theta_t) = Q(p_t, \beta, \alpha) + \theta_t' H(p_t, \beta, \alpha) + \Phi_t(\beta, \alpha, \theta_t),$$

where $q_i(p_t, \beta, \alpha) = \partial Q(p_t, \beta, \alpha)/\partial \ln p_{ti}$, H is a vector-valued function with components H_j for $j = 1, \ldots, J$, $h_{ij}(p_t, \beta, \alpha) = \partial H_j(p_t, \beta, \alpha)/\partial \ln p_{ti}$ is the jth component of h_i, and Φ_t depends on t because it may depend on X_t exclusive of p_t. Hence, Q and q_i are the deterministic parts of the logarithmic objective function and share equations, respectively, $\theta_t' H$ and $\theta_t' h_i$ are the stochastic parts, and Φ_t is a constant of integration that was ignored in section 2 for simplicity.

There are two traditional simplifying assumptions in this cost minimization model. First, the deterministic part of the system is usually written as a conventional linear function of β, denoted $X_t\beta$. This is essentially an assumption that $[q_1 \ldots q_{m-1} Q + \Phi_t]'$ can be written as $X_t\beta$, and holds for the translog form considered in section 2 as well as most other familiar function forms. Second, the stochastic part of the logarithmic objective function, $\theta_t' H$, is normally assumed to be independent of β, although Jobson and Fuller (1980) consider models in which the location vector is not separable from the covariance matrix parameters. With these assumptions we have an SUR system with separable location and covariance matrix parameters. The error vector for observation t including all share equations and the cost function is $u_t = [u_{t1} \ldots u_{tm}]' = [\theta_t' h_1 \ldots \theta_t' h_{m-1} \theta_t' H]'$, and is heteroscedastic in general even if $\theta_t \sim \mathrm{IID}(0, \sigma^2 I_J)$, because in this case $\mathrm{E}(u_t u_t')$ is given by

$$\Omega_t = \sigma^2 \begin{bmatrix} h_1' h_1 & \cdots & h_1' h_{m-1} & h_1' H \\ \vdots & \ddots & \vdots & \vdots \\ h_{m-1}' h_1 & \cdots & h_{m-1}' h_{m-1} & h_{m-1}' H \\ H' h_1 & \cdots & H' h_{m-1} & H' H \end{bmatrix},$$

where h_i and H depend on p_t. To obtain $\hat{\beta}_{\mathrm{EGLS}}$, σ^2 can be ignored and we need only obtain estimates of the elements of this matrix. These elements take the additive form discussed above whenever the (i, j) element can be written in the form $\sum_{k=1}^{G_{ij}} \alpha_{ijk} f_{ijk}(p_t)$, where f_{ijk} are arbitrary functions of the price vector p_t.

The usual cost exhaustion, homogeneity, symmetry, nonnegativity, and concavity restrictions may apply to the estimation of β. Chavas and Segerson discuss these restrictions and note that the cost exhaustion constraint on the stochastic parts of the share equations ($\sum_{i=1}^{m-1} h_i = 0$) results in the familiar singularity problem of applying GLS or EGLS to share equation systems. Barnett (1976) and others have argued that dropping a share equation in a finite-step Aitken estimation is an unacceptable solution to the singularity problem because the resulting estimates depend on which equation is omitted. Hence, most researchers have adopted iterative techniques that converge to maximum likelihood estimates under certain conditions (see Barnett).

Since our methods for estimating α and β are both two-step methods, Barnett's criticism warrants some comment in the present context. It is *not* necessary for finite-step estimates like ours to vary with the equation deleted, and since unique two-step estimates are obtainable, two-step procedures may be preferable to iterative procedures both because two-step methods are easier to apply (no convergence problems) and because Kmenta and Gilbert (1968, p. 1196) find evidence that iterating may be inefficient in small samples. Chavas and Segerson present one method for obtaining unique two-step estimates. This method entails deleting a share equation and then performing a first-stage GLS estimation with an assumed covariance matrix that is constructed to yield estimates invariant to the equation deleted. The resulting unique residuals are used to obtain a unique estimate of the covariance matrix, which is singular like the true covariance matrix if the constraints are imposed on the first-stage estimation. Then, Theil's (1971, p. 281) result on invariance of GLS estimates to omission of linearly dependent observations shows that a second-stage EGLS estimation without the omitted equation, but including any appropriate constraints, yields unique two-step estimates. Schmidt's example (see section 5 above) shows that the properties of this estimator depend on the exact method used to estimate the covariance matrix from the residuals, which in turn depends on the form assumed by the covariance matrix.

The key observation of Chavas and Segerson in defense of finite-step procedures is that any dependence on which equation is omitted arises because the estimated covariance matrix may depend on the deleted equation, not because the second-stage estimation depends on the deleted equation. If the constraints are imposed on the first-stage estimation so that any estimated covariance matrix is singular, as it should be, then Theil's result shows that dropping any share equation in a second-stage EGLS estimation yields unique estimates for the location vector, for the given estimate of the covariance matrix. Hence, we need only obtain a unique estimate of the covariance matrix in order to obtain unique two-step estimates in share equation systems. But there is no difficulty in obtaining a unique singular estimate of the covariance matrix without the constructed first-stage covariance matrix of Chavas and Segerson, and we can also utilize all of the data in the process. Simply perform joint OLS on the complete set of equations with the cost exhaustion and any other appropriate constraints imposed. This yields unique linearly dependent residuals. Then, an assumption concerning the form of the covariance matrix, like the additive structure considered above, yields a unique singular estimate of the true singular covariance matrix. Our results show that if the additive structure is appropriate the estimator obtained from this algorithm has desirable asymptotic properties. Note, however, that our estimator for α in a share equation system involves all of the share equations, so the singularity problem arises again in the estimation of α. Thus, the algorithm for obtaining unique two-step estimates must be applied twice, once in the estimation of α and once in the estimation of β. Other covariance structures may only require the algorithm in the estimation of β. In fact, this algorithm was used by Caves et al. (1980) with a traditional SUR covariance matrix.

Existing computer programs may not implement this technique in a straightforward manner. For example, the SAS®[1] SYSLIN procedure with the SUR

[1] SAS is a registered trademark of the SAS Institute Inc., Cary, NC.

option does not impose cross-equation restrictions in the first-stage OLS estimation. Hence, the resulting estimated covariance matrix is nonsingular and gives constrained EGLS estimates of the location vector even if all share equations are included, but these are not the estimates of interest since the estimated covariance matrix does not satisfy the dictates of theory. SYSLIN also does not provide the flexibility to estimate more equations in the first stage than in the second stage, as required by the two-step algorithm for solving the singularity problem. However, the equations can be stacked as in (4) and (11), and single-equation procedures along with a matrix language can then be used to implement the procedure.

Random Coefficients and Panel Data Models. Some models that assume a particular covariance structure across regimes or economic units are special cases of the model presented above, although our estimation technique usually differs from the methods suggested by the original authors. Letting t denote one dimension of the data for $t = 1, \ldots, T$ and i denote the other dimension for $i = 1, \ldots, m$, a general structure for such models is

$$y_{ti} = \sum_{k=1}^{K_i} \beta_{tik} x_{tik} + \mu_{ti}, \tag{17}$$

where $\beta_{tik} = \bar{\beta}_{ik} + \varepsilon_{tik}$. The stochastic terms satisfy

(i) $\mathrm{E}(\mu_{ti}) = \mathrm{E}(\varepsilon_{tik}) = 0 \quad \text{for all} \quad t, i, k,$

(ii) $$\mathrm{E}(\mu_{ti}\mu_{\tau j}) = \begin{cases} \sigma_{ij} & \text{(independent of } t\text{)} \quad \text{if} \quad t = \tau, \\ 0 & \text{otherwise,} \end{cases}$$

(iii) $$\mathrm{E}(\varepsilon_{tik}\varepsilon_{\tau j\kappa}) = \begin{cases} \delta_{ijk\kappa} & \text{(independent of } t\text{)} \quad \text{if} \quad t = \tau, \\ 0 & \text{otherwise,} \end{cases}$$

(iv) $\mathrm{E}(\mu_{ti}\varepsilon_{\tau jk}) = 0 \quad \text{for all} \quad t, \tau, i, j, k.$

This can be written in the form of eqs. (1)–(4) of section 3 through the following definitions. First, let $u_{ti} = \mu_{ti} + \sum_{k=1}^{K_i} x_{tik}\varepsilon_{tik}$ so that

$$\mathrm{E}(u_{ti}u_{\tau j}) = \begin{cases} \sigma_{ij} + \sum_{k=1}^{K_i} \sum_{\kappa=1}^{K_j} x_{tik}x_{tj\kappa}\delta_{ijk\kappa} & \text{if} \quad t = \tau, \\ 0 & \text{otherwise.} \end{cases}$$

Then, letting $y_t = (y_{t1}, \ldots, y_{tm})'$, $x_{ti} = (x_{ti1}, \ldots, x_{tiK_i})'$, $\beta = (\bar{\beta}_{11}, \ldots, \bar{\beta}_{1K_1}, \ldots, \bar{\beta}_{m1}, \ldots, \bar{\beta}_{mK_m})'$, and $u_t = (u_{t1}, \ldots, u_{tm})$ shows that this model is a special case of the additive heteroscedastic SUR model since in the present case $E(u_t u_\tau') = 0$ for $t \neq \tau$ and Ω_t is given by

$$\begin{bmatrix} \sigma_{11} + \sum_{k=1}^{K_1} \sum_{\kappa=1}^{K_1} x_{t1k} x_{t1\kappa} \delta_{11k\kappa} & \cdots & \sigma_{1m} + \sum_{k=1}^{K_1} \sum_{\kappa=1}^{K_m} x_{t1k} x_{tm\kappa} \delta_{1mk\kappa} \\ \vdots & \ddots & \vdots \\ \sigma_{m1} + \sum_{k=1}^{K_m} \sum_{\kappa=1}^{K_1} x_{tmk} x_{t1\kappa} \delta_{m1k\kappa} & \cdots & \sigma_{mm} + \sum_{k=1}^{K_m} \sum_{\kappa=1}^{K_m} x_{tmk} x_{tm\kappa} \delta_{mmk\kappa} \end{bmatrix}. \tag{18}$$

The elements of Ω_t take the additive form with

$$\alpha_{ij} = (\sigma_{ij}, \delta_{ij11}, \delta_{ij21}, \delta_{ij22}, \ldots, \delta_{ijK_j(K_j-1)}, \delta_{ijK_jK_j}, \delta_{ij(K_j+1)1}, \ldots,$$
$$\delta_{ij(K_j+1)K_j}, \ldots, \delta_{ijK_i1}, \ldots, \delta_{ijK_iK_j})',$$

$$z_{ij}^t = (1, x_{ti1}x_{tj1}, x_{ti2}x_{tj1}, x_{ti2}x_{tj2}, \ldots, x_{tiK_j}x_{tj(K_j-1)}, x_{tiK_j}x_{tjK_j},$$
$$x_{ti(K_j+1)}x_{tj1}, \ldots, x_{ti(K_j+1)}x_{tjK_j}, \ldots, x_{tiK_i}x_{tj1}, \ldots, x_{tiK_i}x_{tjK_j})',$$

since $\delta_{ijk\kappa} = \delta_{ij\kappa k}$ by symmetry, where we have assumed $K_i \geq K_j$ for simplicity.

This model reduces to the SUR random coefficients model of Singh and Ullah (1974) when $\delta_{ijk\kappa} = 0$ for $k \neq \kappa$ and the data is interpreted as comprising T observation on m equations. Note that if $x_{ti1} = 1$ as in Singh and Ullah, then $\sigma_{ij} + \delta_{ij11}$ must be estimated as one intercept parameter. As mentioned in the introduction, our method for estimating α is more efficient asymptotically than the method suggested by Singh and Ullah.

Eqs. (17) and (18) reduce to the random effects panel data model of Swamy and Metha (1975, 1977) under the following conditions:

(a) t and i are interpreted as the two dimensions of the panel, with $T \to \infty$ and m fixed for all asymptotic results.

(b) $K_i = K$ for all i, so that the number of independent variables is the same for every observation in the panel.

(c) $\bar{\beta}_{ik} = \bar{\beta}_k$ for all i and k, so that the mean parameter vector is the same for every observation in the panel.

(d) $\sigma_{ij} = 0$ for all i, j, so that μ_{ti} does not enter the model.

(e) $\delta_{ijk\kappa} = \lambda_{k\kappa}$ for all $i \neq j$, so that the covariances of ε_{tik} and $\varepsilon_{tj\kappa}$ do not vary across i and j when $i \neq j$.

Note that (b) and (c) imply that the β vector contains only K distinct elements, so β must be constrained appropriately when estimating. This constraint can be implemented by simply writing X_t as $X_t = (x_{t1}, \ldots, x_{tm})'$ rather than the specification given in (2). Also, parameter restrictions are required in the estimation of α in this model since α_{ij} is the same vector for all i, j provided $i \neq j$. Since our method for estimating α is asymptotically efficient, it is at least as efficient in large samples as the method suggested by Swamy and Metha.

7. Summary

We derived consistent, asymptotically efficient, and asymptotically normal estimators for SUR systems that have additive heteroscedastic contemporaneous correlation. Both our estimator for the location vector and the parameters of the covariance matrix possess these properties. The procedure presented above is superior to the method proposed by Singh and Ullah (1974), since we followed Amemiya (1977) in using GLS to estimate the parameters of the covariance matrix. Our method also permits the use of cross-equation parameter restrictions. We discuss how this type of heteroscedasticity arises naturally in share equation systems and random coefficient models, and how these models can be uniquely estimated with our two-step estimation technique.

Appendix

Proof of Theorem 1: Substituting eq. (11) into $\hat{\alpha}$ yields

$$T^{1/2}(\hat{\alpha} - \alpha) = (T^{-1}A'Z)^{-1}T^{-1/2}A'(v + \varepsilon).$$

Since $\lim_{T\to\infty}(T^{-1}A'Z)^{-1} = \Phi^{-1}$, we need only show that $T^{-1/2}A'(v+\varepsilon) \xrightarrow{\text{d}} \text{N}(0, \lim_{T\to\infty}T^{-1}A'SA)$. Partition A into T submatrices of dimension $(m(m+1)/2 \times G)$, $A' = [A_1' \ldots A_T']$, and note that $A'(v+\varepsilon) = \sum_{t=1}^{T} A_t'(v_t + \varepsilon_t)$. Moreover $\text{E}(A_t'v_t) = 0$ and $\text{V}(A_t'v_t) = A_t'S_tA_t$ for $t = 1, 2, \ldots,$ and v_t is independent of v_τ for $t \neq \tau$. Normality assures that the higher moments of v_t exist, so [see Judge et al. (1985, p. 189)]

$$T^{-1/2}A'v \xrightarrow{\text{d}} \text{N}\left(0, \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} A_t'S_tA_t\right) = \text{N}\left(0, \lim_{T\to\infty} T^{-1}A'SA\right).$$

Thus, it only remains to show that $\text{plim}_{T\to\infty}T^{-1/2}A'\varepsilon = 0$. Let a_{tn}' denote the G-dimensional nth row of A_t and ε_{tn} the nth element of ε_t. Then

$$T^{-1/2}A'\varepsilon = T^{-1/2}\sum_{t=1}^{T}\sum_{n=1}^{m(m+1)/2} a_{tn}\varepsilon_{tn}.$$

Since the inner sum is of finite length, we need only show that

$$\operatorname*{plim}_{T\to\infty} T^{-1/2} \sum_{t=1}^{T} a_{tn}\varepsilon_{tn} = 0 \quad \text{for arbitrary } n.$$

To each n there corresponds an i and j indexing the ε_t vector, so from eqs. (7) and (8) we may write

$$\begin{aligned}\varepsilon_{tn} &= \hat{u}_{ti}\hat{u}_{tj} - u_{ti}u_{tj}\\ &= (u_{ti} - \tilde{x}'_{ti}(X'X)^{-1}X'u)(u_{tj} - \tilde{x}'_{tj}(X'X)^{-1}X'u) - u_{ti}u_{tj}\\ &= -(u_{ti}\tilde{x}_{tj} + u_{tj}\tilde{x}_{ti})'(X'X)^{-1}X'u\\ &\quad + u'X(X'X)^{-1}\tilde{x}_{ti}\tilde{x}'_{tj}(X'X)^{-1}X'u,\end{aligned}$$

for some i and j, where $\tilde{x}'_{ti}$ is the ith row of X_t including the zeros. Thus, it is sufficient to show

$$\operatorname*{plim}_{T\to\infty} T^{-1/2} \sum_{t=1}^{T} a_{tn}(u_{ti}\tilde{x}_{tj} + u_{ti}\tilde{x}_{ti})'(X'X)^{-1}X'u = 0, \tag{A.1}$$

$$\operatorname*{plim}_{T\to\infty} T^{-1/2} \sum_{t=1}^{T} a_{tn}u'X(X'X)^{-1}\tilde{x}_{ti}\tilde{x}'_{tj}(X'X)^{-1}X'u = 0. \tag{A.2}$$

For (A.1) we have

$$\left[T^{-3/4}\sum_{t=1}^{T} a_{tn}(u_{ti}\tilde{x}_{tj} + u_{tj}\tilde{x}_{ti})'\right](T^{-1}X'X)^{-1}(T^{-3/4}X'u),$$

where the hth column of the $(G\times K)$ term in brackets is

$$T^{-3/4}\sum_{t=1}^{T} a_{tn}(u_{ti}\tilde{x}_{tjh} + u_{tj}\tilde{x}_{tih}).$$

This random vector has zero mean and covariance

$$T^{-3/2}\sum_{t=1}^{T} a_{tn}a'_{tn}(\sigma^t_{ii}\tilde{x}^2_{tjh} + 2\sigma^t_{ij}\tilde{x}_{tjh}\tilde{x}_{tih} + \sigma^t_{jj}\tilde{x}^2_{tih}),$$

where the absolute value of one element of this covariance matrix is bounded by

$$B_a^2B_x^2T^{-3/2}\sum_{t=1}^{T}(|\sigma^t_{ii}| + |2\sigma^t_{ij}| + |\sigma^t_{jj}|) \le T^{-1/2}B_a^2B_x^2 4\|\alpha\|B_zG^{-1/2},$$

which approaches zero as $T \to \infty$. Hence, the hth column has zero probability limit and the entire $(G \times K)$ matrix vanishes in probability. Moreover, $(T^{-1}X'X)^{-1}$ converges by assumption and $(T^{-3/4}X'u)$ has zero mean and covariance $T^{-1/2}(T^{-1}X'\Omega X)$, which approaches zero since $T^{-1}X'\Omega X$ converges. Thus,

$$\operatorname*{plim}_{T \to \infty} (T^{-1}X'X)^{-1}(T^{-3/4}X'u) = 0, \tag{A.3}$$

which establishes (A.1). For (A.2), we have

$$\left[T^{-1} \sum_{t=1}^{T} a_{tn}(T^{-3/4}u'X(T^{-1}X'X)^{-1}\tilde{x}_{ti}\tilde{x}'_{tj} \right](T^{-1}X'X)^{-1}(T^{-3/4}X'u),$$

where by (A.3) we need only show that the $(G \times K)$ matrix in brackets is bounded in probability. The hth column is

$$T^{-1} \sum_{t=1}^{T} a_{tn}(T^{-3/4}u'X)(T^{-1}X'X)^{-1}\tilde{x}_{ti}\tilde{x}_{tjh}$$

$$= \left[T^{-1} \sum_{t=1}^{T} a_{tn}\tilde{x}_{tjh}\tilde{x}'_{ti} \right](T^{-1}X'X)^{-1}(T^{-3/4}X'u).$$

Once again applying (A.3), we need only show that the $(G \times K)$ matrix in brackets is bounded. Each element is clearly bounded by $B_a B_x^2$, which establishes (A.2). ■

Proof of Lemma 1: We shall prove the result for Δ_t when $H_t(\gamma)$ is the matrix that results from substituting γ in place of α in Ω_t. Since C^t_{ij} is a determinant that takes the same form as Δ_t, the proofs are identical. When $H_t(\gamma)$ results from substituting γ in place of α in S_t, the arguments are the same except that there are more terms involved. Fix $\gamma \in \mathscr{R}^G$ and suppose that $\|\tilde{\gamma} - \gamma\| < \delta$. We may write

$$|\Delta_t(\tilde{\gamma}) - \Delta_t(\gamma)| = |\textstyle\sum \pm [\tilde{\gamma}'_{1j_1}z^t_{1j_1} \cdots \tilde{\gamma}'_{mj_m}z^t_{mj_m} - \gamma'_{1j_1}z^t_{1j_1} \cdots \gamma'_{mj_m}z^t_{mj_m}]|,$$

where the sum ranges over all permutations $(j_1, \ldots, j_m)$. Let $\tilde{d} = \tilde{\gamma}'_{2j_2}z^t_{1j_2} \ldots \tilde{\gamma}'_{mj_m}z^t_{mj_m}$ and define d analogously. Then

$$\begin{aligned} |\Delta_t(\tilde{\gamma}) - \Delta_t(\gamma)| &\le \textstyle\sum |(\tilde{\gamma}_{1j_1} - \gamma_{1j_1})'z^t_{1j_1}d + \tilde{\gamma}'_{1j_1}z^t_{1j_1}(\tilde{d} - d)| \\ &\le \textstyle\sum [\, \|\tilde{\gamma}_{1j_1} - \gamma_{1j_1}\| B_z G^{1/2}_{1j_1} \|\gamma\|^{m-1} G^{(m-1)/2} B_z^{m-1} \\ &\qquad + \|\tilde{\gamma}_{1j_1}\| B_z G^{1/2}_{1j_1} |\tilde{d} - d| \,] \\ &\le n_m \delta B_z^m G^{m/2} \|\gamma\|^{m-1} + [\, \|\gamma\| + \delta] B_z G^{1/2} \textstyle\sum |\tilde{d} - d|, \end{aligned}$$

where n_m is the number of terms in the sum. Clearly the first term is arbitrarily small through an appropriate choice of δ. The second term is a constant (independent of t) multiplied by a sum that is analogous to what we began with, but with one fewer term in each product. Hence, repeating the argument m times gives the result. ■

Proof of Corollary: By Lemma 1, there exists $\delta > 0$ (independent of t) such that

$$\| T^{-1/2}\gamma \| < \delta \Rightarrow |\Delta_t(T^{-1/2}\gamma + \alpha) - \Delta_t(\alpha)| < \varepsilon_0/2.$$

But this implies $\Delta_t(\alpha) - \varepsilon_0/2 < \Delta_t(T^{-1/2}\gamma + \alpha) \leq |\Delta_t(T^{-1/2}\gamma + \alpha)|$. Since $|\Delta_t(\alpha)| \geq \varepsilon_0$ for every t,

$$\| T^{-1/2}\gamma \| < \delta \Rightarrow \varepsilon_0/2 < |\Delta_t(T^{-1/2}\gamma + \alpha)|.$$

Now, let $\gamma * \in \mathscr{R}^G$ be any vector satisfying $\| \gamma * \| = C$. Clearly there exists T_C such that $T > T_C \Rightarrow \| T^{-1/2}\gamma * \| < \delta$. But for any vector satisfying $\| \gamma \| \leq C$ we have $\| T^{-1/2}\gamma \| \leq \| T^{-1/2}\gamma * \|$, so

$$T > T_C \text{ and } \| \gamma \| \leq C \Rightarrow \| T^{-1/2}\gamma \| < \delta \Rightarrow \varepsilon_0/2 < |\Delta_t(T^{-1/2}\gamma + \alpha)|. \quad ■$$

Proof of Lemma 2: As in Lemma 1, we shall prove the result for $\Delta_t(\gamma)$ when $H_t(\gamma)$ is obtained by substituting γ for α in Ω_t. From Lemma 1,

$$\begin{aligned}
&|\Delta_t(T^{-1/2}\gamma + \alpha)| \\
&\leq \sum |(T^{-1/2}\gamma_{1j_1} + \alpha_{1j_1})' z^t_{1j_1}| \cdots |(T^{-1/2}\gamma_{mj_m} + \alpha_{mj_m})' z^t_{mj_m}| \\
&\leq \sum (T^{-1/2}\|\gamma_{1j_1}\| + \|\alpha_{1j_1}\|) B_z G^{1/2}_{1j_1} \cdots (T^{-1/2}\|\gamma_{mj_m}\| + \|\alpha_{mj_m}\|) B_z G^{1/2}_{mj_m} \\
&\leq \sum (C + \|\alpha\|)^m B_z^m G^{m/2} \\
&= n_m (C + \|\alpha\|)^m B_z^m G^{m/2}. \quad ■
\end{aligned}$$

Proof of Theorem 2: First we use the uniform continuity of Δ_t and C^t_{ij} to show that for every $\delta, \varepsilon > 0$ there exists T^* such that

$$T > T^* \Rightarrow \begin{cases} \mathrm{P}(|\Delta_t(\hat{\alpha})| \leq \varepsilon_0/4, \exists t) < \varepsilon, & \text{(A.4)} \\ \mathrm{P}(|\Delta_t(\hat{\alpha}) - \Delta_t(\alpha)| \geq \delta, \exists t) < \varepsilon, & \text{(A.5)} \\ \mathrm{P}(|C^t_{ij}(\hat{\alpha}) - C^t_{ij}(\alpha)| \geq \delta, \exists t) < \varepsilon \quad \text{for all} \quad i,j. & \text{(A.6)} \end{cases}$$

By Lemma 1 there exists $\delta_0 > 0$ such that $\|\hat{\alpha} - \alpha\| < \delta_0 \Rightarrow |\Delta_t(\hat{\alpha}) - \Delta_t(\alpha)| < \varepsilon_0/4$ for every t. Since $|\Delta_t(\alpha)| - \varepsilon_0/4 \leq |\Delta_t(\hat{\alpha})| + |\Delta_t(\alpha) - \Delta_t(\hat{\alpha})| - \varepsilon_0/4$, we have $\|\hat{\alpha} - \alpha\| < \delta_0 \Rightarrow |\Delta_t(\alpha)| - \varepsilon_0/4 < |\Delta_t(\hat{\alpha})|$. Set $C = 0$ in the Corollary and note that the conclusion is then independent of T. Thus, $|\Delta_t(\alpha)| > \varepsilon_0/2$ and we have $\|\hat{\alpha} - \alpha\| < \delta_0 \Rightarrow \varepsilon_0/4 < |\Delta_t(\hat{\alpha})|$ for every t. By consistency of $\hat{\alpha}$ there exists T^* such that $T > T^* \Rightarrow \mathrm{P}(\|\hat{\alpha} - \alpha\| \geq \delta_0) < \varepsilon$. Thus,

$$\begin{aligned}
T > T^* \Rightarrow \mathrm{P}(|\Delta_t(\hat{\alpha})| \leq \varepsilon_0/4, \exists t) \\
&= \mathrm{P}(|\Delta_t(\hat{\alpha})| \leq \varepsilon_0/4, \exists t, \text{ and } \|\hat{\alpha} - \alpha\| \geq \delta_0) \\
&\quad + \mathrm{P}(|\Delta_t(\hat{\alpha})| \leq \varepsilon_0/4, \exists t, \text{ and } \|\hat{\alpha} - \alpha\| \geq \delta_0) \\
&\leq \mathrm{P}(\|\hat{\alpha} - \alpha\| \geq \delta_0) \\
&\quad + \mathrm{P}(|\Delta_t(\hat{\alpha})| \leq \varepsilon_0/4, \exists t, \text{ and } \|\hat{\alpha} - \alpha\| < \delta_0) \\
&< \varepsilon,
\end{aligned}$$

which establishes (A.4). Now choose δ_0 such that $\|\hat{\alpha} - \alpha\| < \delta_0 \Rightarrow |\Delta_t(\hat{\alpha}) - \Delta_t(\alpha)| < \delta$ for every t. Then (A.5) follows since

$$\begin{aligned}
T > T^* \Rightarrow \mathrm{P}(|\Delta_t(\hat{\alpha}) - \Delta_t(\alpha)| \geq \delta, \exists t) \\
&= \mathrm{P}(|\Delta_t(\hat{\alpha}) - \Delta_t(\alpha)| \geq \delta, \exists t, \text{ and } \|\hat{\alpha} - \alpha\| \geq \delta_0) \\
&\quad + \mathrm{P}(|\Delta_t(\hat{\alpha}) - \Delta_t(\alpha)| \geq \delta, \exists t, \text{ and } \|\hat{\alpha} - \alpha\| < \delta_0) \\
&< \varepsilon.
\end{aligned}$$

Finally, since (A.5) relies only upon uniform continuity and consistency, (A.6) holds by the same arguments. Now, as in Theorem 1 partition A into T matrices of dimension $(r \times q)$, $A' = [A_1' \cdots A_T']$, and note that

$$T^{-1} A'[\text{block diag}\{H_1(\gamma), \ldots, H_T(\gamma)\}]^{-1} A = T^{-1} \sum_{t=1}^{T} A_t' H_t(\gamma)^{-1} A_t.$$

The (l, k) element of $H_t(\gamma)^{-1}$ is $C_{kl}^t(\gamma)/\Delta_t(\gamma)$. Hence, the (i, j) element of the above matrix is

$$T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{r} \sum_{k=1}^{r} a_{til} C_{kl}^t(\gamma) a_{tkj} / \Delta_t(\gamma),$$

and we need only show that the probability limit of this expression when $\gamma = \hat{\alpha}$ is the (i,j) element of Φ, denoted by ϕ_{ij}. Fix $\delta, \varepsilon > 0$, let $\varepsilon_1 = \varepsilon\varepsilon_0^2/16B_a^2 r^2$, and use (A.4), (A.5), (A.6), and the definition of Φ to select T^* such that

$$T > T^* \Rightarrow \begin{cases} |T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{r} \sum_{k=1}^{r} a_{til} C_{kl}^t(\alpha) a_{tkj}/\Delta_t(\alpha) - \phi_{ij}| < \varepsilon/2, \\ \mathrm{P}(|\Delta_t(\hat{\alpha})| \leq \varepsilon_0/4, \exists t) < \delta/3, \\ \mathrm{P}(|\Delta_t(\hat{\alpha}) - \Delta_t(\alpha)| \geq \max\{\sqrt{\varepsilon_1/4}, \varepsilon_1/4B_{H,0}\}, \exists t) < \delta/3, \\ \mathrm{P}(|C_{kl}^t(\hat{\alpha}) - C_{kl}^t(\alpha)| \geq \max\{\sqrt{\varepsilon_1/4}, \varepsilon_1/4B_{H,0}\}, \exists t) < \delta/3. \end{cases}$$

Then,

$$\begin{aligned} T > T^* \Rightarrow \mathrm{P}&\left(\left| T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{r} \sum_{k=1}^{r} a_{til} C_{kl}^t(\hat{\alpha}) a_{tkj}/\Delta_t(\hat{\alpha}) - \phi_{ij} \right| > \varepsilon \right) \\ &\leq \mathrm{P}\left(T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{r} \sum_{k=1}^{r} |a_{til} a_{tkj}| \, |C_{kl}^t(\hat{\alpha})/\Delta_t(\hat{\alpha}) - C_{kl}^t(\alpha)/\Delta_t(\alpha)| \right. \\ &\qquad \left. + \left| T^{-1} \sum_{t=1}^{T} \sum_{l=1}^{r} \sum_{k=1}^{r} a_{til} C_{kl}^t(\alpha) a_{tkj}/\Delta_t(\alpha) - \phi_{ij} \right| > \varepsilon \right) \\ &\leq \mathrm{P}\left(T^{-1} B_a^2 \sum_{t=1}^{T} \sum_{l=1}^{r} \sum_{k=1}^{r} |C_{kl}^t(\hat{\alpha})/\Delta_t(\hat{\alpha}) \right. \\ &\qquad \left. - C_{kl}^t(\alpha)/\Delta_t(\alpha)| > \varepsilon/2 \right) \\ &\leq \mathrm{P}\left(T^{-1} B_a^2 8/\varepsilon_0^2 \sum_{t=1}^{T} \sum_{l=1}^{r} \sum_{k=1}^{r} |C_{kl}^t(\hat{\alpha}) \Delta_t(\alpha) \right. \\ &\qquad \left. - C_{kl}^t(\alpha) \Delta_t(\hat{\alpha})| > \varepsilon/2 \right) \\ &\qquad + \mathrm{P}(|\Delta_t(\hat{\alpha})| \leq \varepsilon_0/4, \exists t) \\ &\leq \mathrm{P}\left(\max_{\substack{1 \leq t \leq T \\ 1 \leq l,k \leq r}} |\Delta_t(\alpha) C_{kl}^t(\hat{\alpha}) - \Delta_t(\hat{\alpha}) C_{kl}^t(\alpha)| > \varepsilon_1 \right) + \delta/3. \end{aligned}$$

For arbitrary l, k, and t, we have

$$
\begin{aligned}
T > T^* \Rightarrow\ & \mathrm{P}(|\Delta_t(\alpha)C^t_{kl}(\hat{\alpha}) - \Delta_t(\hat{\alpha})C^t_{kl}(\alpha)| > \varepsilon_1) \\
& \leq \mathrm{P}(|C^t_{kl}(\hat{\alpha})||\Delta_t(\alpha) - \Delta_t(\hat{\alpha})| \\
& \quad + |\Delta_t(\hat{\alpha})||C^t_{kl}(\hat{\alpha}) - C^t_{kl}(\alpha)| > \varepsilon_1) \\
& \leq \mathrm{P}(|C^t_{kl}(\hat{\alpha}) - C^t_{kl}(\alpha)||\Delta_t(\alpha) - \Delta_t(\hat{\alpha})| \\
& \quad + |C^t_{kl}(\alpha)||\Delta_t(\alpha) - \Delta_t(\hat{\alpha})| \\
& \quad + |\Delta_t(\hat{\alpha}) - \Delta_t(\alpha)||C^t_{kl}(\hat{\alpha}) - C^t_{kl}(\alpha)| \\
& \quad + |\Delta_t(\alpha)||C^t_{kl}(\hat{\alpha}) - C^t_{kl}(\alpha)| > \varepsilon_1) \\
& \leq \mathrm{P}(2|C^t_{kl}(\hat{\alpha}) - C^t_{kl}(\alpha)||\Delta_t(\hat{\alpha}) - \Delta_t(\alpha)| \\
& \quad + B_{H,0}|\Delta_t(\hat{\alpha}) - \Delta_t(\alpha)| + B_{H,0}|C^t_{kl}(\hat{\alpha}) - C^t_{kl}(\alpha)| > \varepsilon_1) \\
& \quad \text{by Lemma 2} \\
& \leq \mathrm{P}(2\sqrt{\varepsilon_1/4}\sqrt{\varepsilon_1/4} + \varepsilon_1/4 + \varepsilon_1/4 > \varepsilon_1) + 2\delta/3 \\
& = 2\delta/3. \quad \blacksquare
\end{aligned}
$$

Proof of Theorem 4: As in Theorem 2, partition A into T submatrices of dimension $(r \times q)$, $A' = [A'_1 \ \dots \ A'_T]$. Then

$$
\begin{aligned}
& T^{-1/2}A'([\text{block diag}\{H_1(\hat{\alpha}), \dots, H_T(\hat{\alpha})\}]^{-1} \\
& \quad - \text{block diag}\{H_1(\alpha), \dots, H_T(\alpha)\}]^{-1})[w'_1 \ \dots \ w'_T] \\
& = T^{-1/2}\sum_{t=1}^{T} A'_t(H_t(\hat{\alpha})^{-1} - H_t(\alpha)^{-1})w_t.
\end{aligned}
$$

Make the following definitions in Theorem 3:

(1) $A_{tT}(\gamma) = A'_t(H_t(T^{-1/2}\gamma + \alpha)^{-1} - H_t(\alpha)^{-1})$.

(2) For every $C < \infty$, T_C satisfies the Corollary and $B_C = 4rB_a G B_{\partial,C} \times (2B_{H,C} + 1)/\varepsilon_0^2$, where $B_{\partial,C}$ is to be defined below.

(3) $\gamma_T = T^{-1/2}(\hat{\alpha} - \alpha)$.

Clearly $A_{tT}(0) = 0$, $\gamma_T = O_p(1)$, and w_t satisfies hypothesis 4 and 5 of Theorem 3. Hence, we need only verify the Lipschitz Condition, hypothesis 2. Fix C and $\|\gamma_0\|, \|\gamma_1\| < C$, and note that

$$|||A_{tT}(\gamma_0) - A_{tT}(\gamma_1)||| \leq rB_a|||H_t(T^{-1/2}\gamma_0 + \alpha) - H_t(T^{-1/2}\gamma_1 + \alpha)|||$$

$$= rB_a|C^t_{ij}(T^{-1/2}\gamma_0 + \alpha)/\Delta_t(T^{-1/2}\gamma_0 + \alpha)$$

$$- C^t_{ij}(T^{-1/2}\gamma_1 + \alpha)/\Delta_t(T^{-1/2}\gamma_1 + \alpha)|,$$

for some i, j between 1 and r. Thus, for $T > T_C$ we have

$$|||A_{tT}(\gamma_0) - A_{tT}(\gamma_1)|||$$

$$\leq (4rB_a/\varepsilon_0^2)|\Delta_t(T^{-1/2}\gamma_1 + \alpha)C^t_{ij}(T^{-1/2}\gamma_0 + \alpha)$$

$$- \Delta_t(T^{-1/2}\gamma_0 + \alpha)C^t_{ij}(T^{-1/2}\gamma_1 + \alpha)|$$

$$\leq (4rB_a/\varepsilon_0^2)\{|\Delta_t(T^{-1/2}\gamma_1 + \alpha)C^t_{ij}(T^{-1/2}\gamma_0 + \alpha)$$

$$- \Delta_t(T^{-1/2}\gamma_1 + \alpha)C^t_{ij}(T^{-1/2}\gamma_1 + \alpha)|$$

$$+ |\Delta_t(T^{-1/2}\gamma_0 + \alpha)C^t_{ij}(T^{-1/2}\gamma_0 + \alpha)$$

$$- \Delta_t(T^{-1/2}\gamma_0 + \alpha)C^t_{ij}(T^{-1/2}\gamma_1 + \alpha)|$$

$$+ |\Delta_t(T^{-1/2}\gamma_1 + \alpha)C^t_{ij}(T^{-1/2}\gamma_1 + \alpha)$$

$$- \Delta_t(T^{-1/2}\gamma_0 + \alpha)C^t_{ij}(T^{-1/2}\gamma_0 + \alpha)|\}$$

$$\leq (4rB_a/\varepsilon_0^2)\{2B_{H,C}|C^t_{ij}(T^{-1/2}\gamma_0 + \alpha) - C^t_{ij}(T^{-1/2}\gamma_1 + \alpha)|$$

$$+ |\Delta_t(T^{-1/2}\gamma_1 + \alpha)C^t_{ij}(T^{-1/2}\gamma_1 + \alpha)$$

$$- \Delta_t(T^{-1/2}\gamma_0 + \alpha)C^t_{ij}(T^{-1/2}\gamma_0 + \alpha)|\}$$

by Lemma 2.

Suppose $C^t_{ij}(T^{-1/2}\gamma + \alpha)$ and $\Delta_t(T^{-1/2}\gamma + \alpha)$ are differentiable functions of γ on $\mathscr{R}^G$, and that there exists a bound $B_{\partial,C}$ such that $\|\gamma\| \leq C$ implies every partial derivative is less than $B_{\partial,C}$ in absolute value, for every t, T, i, and j. Then, by the

Mean Value Theorem there exists $\gamma_{tT}*$ on the line segment between γ_0 and γ_1 (and hence satisfying $\|\gamma_{tT}*\| \le C$) such that

$$|C_{ij}^t(T^{-1/2}\gamma_0 + \alpha) - C_{ij}^t(T^{-1/2}\gamma_1 + \alpha)| = T^{-1/2}|\nabla C_{ij}^t(\gamma_{tT}*)\cdot(\gamma_0 - \gamma_1)|$$
$$\le G^{1/2}B_{\partial,C}T^{-1/2}\|\gamma_0 - \gamma_1\|.$$

Similarly,

$$|\Delta_t(T^{-1/2}\gamma_1 + \alpha)C_{ij}^t(T^{-1/2}\gamma_1 + \alpha) - \Delta_t(T^{-1/2}\gamma_0 + \alpha)C_{ij}^t(T^{-1/2}\gamma_0 + \alpha)|$$
$$\le G^{1/2}B_{\partial,C}T^{-1/2}\|\gamma_0 - \gamma_1\|,$$

since the bound on Δ_t and C_{ij}^t (Lemma 2) assures that the partials of $\Delta_t(T^{-1/2}\gamma + \alpha)C_{ij}^t(T^{-1/2}\gamma + \alpha)$ are bounded. Thus,

$$|||A_{tT}(\gamma_0) - A_{tT}(\gamma_1)||| \le (4rB_aGB_{\partial,C}/\varepsilon_0^2)(2B_{H,C} + 1)T^{-1/2}\|\gamma_0 - \gamma_1\|$$
$$= B_C T^{-1/2}\|\gamma_0 - \gamma_1\|.$$

For differentiability, note that $C_{ij}^t(T^{-1/2}\gamma + \alpha)$ and $\Delta_t(T^{-1/2}\gamma + \alpha)$ are polynomials in terms like $T^{-1/2}\gamma_{kl} + \alpha_{kl})'z_{kl}^t$, where γ_{kl} and α_{kl} are G_{kl}-dimensional subvectors of γ and α. The boundedness of the partial derivatives follows from the same arguments contained in Lemmas 1 and 2. ■

Proof of Theorem 5: Note that

$$T^{-1/2}Z'(\hat{S}^{-1} - S^{-1})\varepsilon = T^{-1/2}\sum_{t=1}^{T} Z_t'(\hat{S}_t^{-1} - S_t^{-1})\varepsilon_t.$$

Since each row of Z_t' has only one nonzero element, one element of this $(G \times 1)$ vector can be written as

$$T^{-1/2}\sum_{t=1}^{T} z_{lsk}^t \sum_{n=1}^{m(m+1)/2} (C_{sn}^t(\hat{\alpha}_{\mathrm{OLS}})/\Delta_t(\hat{\alpha}_{\mathrm{OLS}}) - C_{sn}^t(\alpha)/\Delta_t(\alpha))\varepsilon_{tn},$$

where z_{lsk}^t is the kth element of z_{ls}^t. Since the inner sum is of finite length it suffices to show

$$\operatorname*{plim}_{T\to\infty} T^{-1/2}\sum_{t=1}^{T} z_{lsk}^t(C_{sn}^t(\hat{\alpha}_{\mathrm{OLS}})/\Delta_t(\hat{\alpha}_{\mathrm{OLS}}) - C_{sn}^t(\alpha)/\Delta_t(\alpha))\varepsilon_{tn} = 0,$$

for arbitrary l, s, k, and n. Substituting for ε_{tn} as in Theorem 1 yields the sufficient conditions

$$\operatorname*{plim}_{T\to\infty}\left[T^{-1/2}\sum_{t=1}^{T} z^t_{lsk}(C^t_{sn}(\hat{\alpha}_{\mathrm{OLS}})/\Delta_t(\hat{\alpha}_{\mathrm{OLS}}) - C^t_{sn}(\alpha)/\Delta_t(\alpha)) \times (u_{ti}\tilde{x}_{tj} + u_{tj}\tilde{x}_{ti})'\right](X'X)^{-1}X'u = 0, \tag{A.7}$$

$$\operatorname*{plim}_{T\to\infty} T^{-1/2}\sum_{t=1}^{T} z^t_{lsk}(C^t_{sn}(\hat{\alpha}_{\mathrm{OLS}})/\Delta_t(\hat{\alpha}_{\mathrm{OLS}}) - C^t_{sn}(\alpha)/\Delta_t(\alpha)) \times u'X(X'X)^{-1}\tilde{x}_{ti}\tilde{x}'_{tj}(X'X)^{-1}X'u = 0. \tag{A.8}$$

For (A.7), the hth element of the K-dimensional vector in brackets is

$$\begin{aligned} &T^{-1/2}\sum_{t=1}^{T} z^t_{\ell sk}(C^t_{sn}(\hat{\alpha}_{\mathrm{OLS}})/\Delta_t(\hat{\alpha}_{\mathrm{OLS}}) - C^t_{sn}(\alpha)/\Delta_t(\alpha))(u_{ti}\tilde{x}_{tjh} + u_{tj}\tilde{x}_{ith}) \\ &= T^{-1/2}A'([\text{block diag}\{\Delta_1(\hat{\alpha}_{\mathrm{OLS}})/C^1_{sn}(\hat{\alpha}_{\mathrm{OLS}}), \dots, \\ &\quad \Delta_T(\hat{\alpha}_{\mathrm{OLS}})/C^T_{sn}(\hat{\alpha}_{\mathrm{OLS}})\}]^{-1} \\ &\quad - [\text{block diag}\{\Delta_1(\alpha)/C^1_{sn}(\alpha), \dots, \\ &\quad \Delta_T(\alpha)/C^T_{sn}(\alpha)\}]^{-1})[w'_1 \cdots w'_T]', \end{aligned}$$

where $A' = [z^1_{lsk} \dots z^T_{lsk}]$ and $w_t = u_{ti}\tilde{x}_{tjh} + u_{tj}x_{tih}$ for $t = 1, \dots, T$. Setting $q = r = 1$ and $B_a = B_z$ in Theorem 4 shows that the hth element vanishes in probability since $\mathrm{E}(w_t) = 0$ and

$$\mathrm{E}(w_t^2) = \sigma^t_{ii}\tilde{x}^2_{tjh} + 2\sigma^t_{ij}\tilde{x}_{tjh}\tilde{x}_{tih} + \sigma^t_{jj}\tilde{x}^2_{tih} \le B_x^2 4\|\alpha\| B_z G^{1/2}.$$

Thus, the entire vector in brackets vanishes in probability. Combining with (A.3) demonstrates (A.7). For (A.8), recall from Theorem 2 that consistency of $\hat{\alpha}_{\mathrm{OLS}}$ implies

$$\operatorname*{plim}_{T\to\infty}(C^t_{sn}(\hat{\alpha}_{\mathrm{OLS}})/\Delta_t(\hat{\alpha}_{\mathrm{OLS}}) - C^t_{sn}(\alpha)/\Delta_t(\alpha)) = 0,$$

uniformly in t. Following the methodology of Theorem 1, we need only show that the K-dimensional vector

$$T^{-1}\sum_{t=1}^{T} z^t_{\ell sk}(C^t_{sn}(\hat{\alpha}_{\mathrm{OLS}})/\Delta_t(\hat{\alpha}_{\mathrm{OLS}}) - C^t_{sn}(\alpha)/\Delta_t(\alpha))\tilde{x}_{tjh}\tilde{x}'_{ti} \tag{A.9}$$

is bounded in probability. Fix $\varepsilon > 0$ and select T^* such that

$$T > T^* \Rightarrow \mathrm{P}(|C_{sn}^t(\hat{\alpha}_{\mathrm{OLS}})/\Delta_t(\hat{\alpha}_{\mathrm{OLS}}) - C_{sn}^t(\alpha)/\Delta_t(\alpha)| \geq 1, \exists t) < \varepsilon.$$

Then, for the pth element of (A.9) we have

$$\begin{aligned} T > T^* \Rightarrow \mathrm{P}\Bigg(T^{-1} \Bigg| \sum_{t=1}^{T} z_{isk}^t (C_{sn}^t(\hat{\alpha}_{\mathrm{OLS}})/\Delta_t(\hat{\alpha}_{\mathrm{OLS}}) \\ - C_{sn}^t(\alpha)/\Delta_t(\alpha))\tilde{x}_{tjh}\tilde{x}_{tip} \Bigg| \geq B_z B_x^2 \Bigg) \\ \leq \mathrm{P}(B_z B_x^2 > B_z B_x^2) + \mathrm{P}(|C_{sn}^t(\hat{\alpha}_{\mathrm{OLS}})/\Delta_t(\hat{\alpha}_{\mathrm{OLS}}) \\ - C_{sn}^t(\alpha)/\Delta_t(\alpha)| \geq 1, \exists t) < \varepsilon. \quad \blacksquare \end{aligned}$$

References

Amemiya T., 1977, A note on a heteroscedastic model, Journal of Econometrics 6, 365–370.

Barnett, W.A., 1976, Maximum likelihood and iterated Aitken estimation of nonlinear systems of equations, Journal of the American Statistical Association 71, 354–360.

Carroll, R.J. and D. Ruppert, 1982, Robust estimation in heteroscedastic linear models, Annals of Statistics 10, 429–441.

Caves, D.W., L.R. Christensen, and M.W. Tretheway, 1980, Flexible cost functions for multiproduct firms, Review of Economics and Statistics 62, 477–481.

Chavas, J.-P. and K. Segerson, 1987, Stochastic specification and estimation of share equation systems, Journal of Econometrics 35, 337–358.

Christensen, L.R. and W.H. Greene, 1976, Economies of scale in U.S. electric power generation, Journal of Political Economy 84, 655–676.

Christensen, L.R. et al., 1975, Transcendental logarithmic utility functions, American Economic Review 65, 367–383.

Crockett, P.W., 1983, Nonparametric density and moment estimation in a random coefficients model, Unpublished Ph.D. dissertation (University of North Carolina, Chapel Hill, NC).

Crockett, P.W., 1985, Asymptotic distribution of the Hildreth–Houck estimator, Journal of the American Statistical Association 80, 202–204.

Duncan, G.M., 1983, Estimation and inference for heteroscedastic systems of equations, International Economic Review 24, 559–566.

Hildreth, C. and J.P. Houck, 1968, Some estimators for a linear model with random coefficients, Journal of the American Statistical Association 63, 584–595.

Jobson, J.D. and W.A. Fuller, 1980, Least squares estimation when the covariance matrix and parameter vector are functionally related, Journal of the American Statistical Association 75, 176–181.

Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl, and T.-C. Lee, 1985, The theory and practice of econometrics, 2nd ed. (Wiley, New York, NY).

Kmenta, J. and R. Gilbert, 1968, Small sample properties of alternative estimators of seemingly unrelated regressions, Journal of the American Statistical Association 63, 1180–1200.

Low, C., 1982, Seemingly unrelated regressions with heteroskedastic disturbances: Some finite sample results, Ph.D. thesis (Department of Econometrics and Operations Research, Monash University, Melbourne).

Low, C., 1983, Exact finite sample properties of Zellner's SUR estimators when the disturbances are heteroscedastic, Working paper no. 01/83 (Department of Economics and Statistics, National University of Singapore, Singapore).
Magnus, J.R., 1978, Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix, Journal of Econometrics 7, 281–312.
Raj, B., 1975, Linear regression with random coefficients: The finite sample and convergence properties, Journal of the American Statistical Association 70, 127–137.
Schmidt, P., 1976, Econometrics (Dekker, New York, NY).
Singh, B. and A. Ullah, 1974, Estimation of seemingly unrelated regression with random coefficients, Journal of the American Statistical Association 69, 191–195.
Srivastava, V. and D.E.A. Giles, 1987, Seemingly unrelated regression equations models: Estimation and inference (Dekker, New York, NY).
Surehka, K. and W.E. Griffiths, 1984, A Monte Carlo comparison of some Bayesian and sampling theory estimators in two heteroscedastic error models, Communications in Statistics B 13, 85–105.
Swamy, P.A.V.B. and J.S. Mehta, 1975, Bayesian and non-Bayesian analysis of switching regressions and of random coefficient regression models, Journal of the American Statistical Association 70, 593–602.
Swamy, P.A.V.B. and J.S. Mehta, 1977, Estimation of linear models with time and cross-sectionally varying components, Journal of the American Statistical Association 72, 890–898.
Theil, H., 1971, Principles of econometrics (Wiley, New York, NY).
Zellner, A., 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, Journal of the American Statistical Association 57, 348–368.