Time Series Analysis for the Social Sciences: Online Appendix

Janet M. Box-Steffensmeier, John R. Freeman, Matthew P. Hitt,
and Jon C. Pevehouse

1 Step 3 of Pseudo-IV GLS Estimation, the Exchange Rate Model of British Voting Intentions in Chapter 3

Figure 1 displays the acf and the pacf for the residuals from step three of the procedure for the bivariate workhorse model with exchange rates.

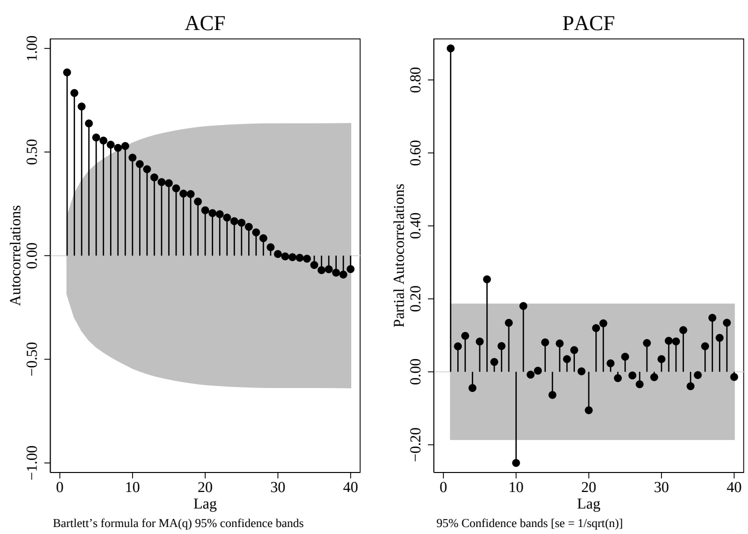

Figure 1: ACF and PACF of Residuals from Step 3, Exchange Rate Model of Voting Intentions, Britain, 1997:5-2006:9

1.1 Step 3 of Pseudo-IV GLS Esitmation, the Short Term Interest Rate Model of British Voting Intentions

Figure 1.1 displays the acf and pacf for the residuals from step three of the procedure for the bivariate model with short-term interest rates.

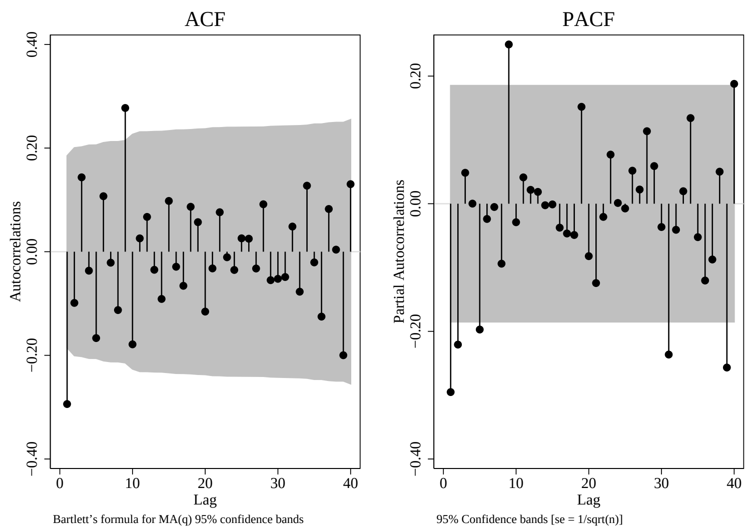

Figure 2: ACF of Residuals from Step 3, Short Term Interest Rate Model of Voting Intentions, Britain, 1997:5-2006:9

2 A Brief Review of Regression Simultaneous Equation Estimation

2.1 Terminology

We present this primer because analysts should know that simply estimating some specifications of a simultaneous equation model with OLS can result in biased coefficients and badly mistaken inferences, as we will demonstrate below.

Simultaneity is synonymous with interdependence or feedback between two variables. Put another way, it means some variables are a cause and a consequence of other variables. These variables are called *endogenous*. Since each such variable has its own equation–an equation in which it alone is the left hand side variable, the number of equations in the system is equal to the number of endogenous variables.[1]

The other variables in the system are call *predetermined*. They are independent of all current and subsequent disturbances in the system (Greene 2003: 382). There are two types of such variables:

1. Exogenous Variables: variables that are completely determined outside of the system of equations.

2. Lagged Endogenous Variables: variables that represent *past* values of the endogenous variables in the system.[2]

Once more, the key feature of simultaneous equation models is that the dependent variable in one equation may be an explanatory variable in another. So, in our British political economy example in equation (4.2),consumer prices,(CPI_t), and interest rates, (IR_t), are interdependent and hence endogenous. The exchange rate (XR_{t-1})and prime ministerial approval (PM_{t-1}) are predetermined and exogenous. In political science, the focus is on the coefficient on the latter variable. As we explain below, there are debates about whether central banks are forced to bolster public support for elected officials like Prime Ministers. Tests of the competing hypotheses in these debates would emphasize the statistical significance of the β_2 coefficient in the model.[3]

2.2 The pitfalls of using OLS; simultaneous equation bias

Simply put, OLS estimation of the coefficients in a simultaneous equation system yields results that are biased and inconsistent. This is because the endogenous variables on the right side of the equations are correlated with the respective error terms. To illustrate this problems, consider a very

[1]The system of equations also is said to be "complete" insofar as there are as many equations in the system as there are endogenous variables. See Greene 2003: Section 15.2.1.

[2]For a deeper classification of variables especially exogenous variables see Hendry and Richards (1983). An application of Hendry and Richards typology is Granato and Smith 1994.

[3]There also are different forms of the basic model, such as block diagonal, recursive, and block triangular. See Kmenta 1986: 655-7, for more details.

simple (structural) Keynesian model of consumption[4]

$$c_t = \beta y_t + \epsilon_t$$
$$y_t = c_t + i_t + g_t$$

where;

$c_t =$ aggregate consumption at time t
$y_t =$ national income
$i_t =$ investment
$g_t =$ government spending
$u =$ Stochastic Disturbance
$\beta =$ marginal propensity to consume (MPC; $0 < \beta < 1$).

Recall that the ordinary least squares estimate of β is simply:

$$\hat{\beta} = \frac{\sum c_t y_t}{\sum y_t^2} = \frac{\sum y_t(\beta y_t + \epsilon_t)}{\sum y_t^2} = \beta + \frac{\sum y_t \epsilon_t}{\sum y_t^2}. \quad (1)$$

Moreover, as regards consistency, we know, for OLS,

$$plim \hat{\beta} = \beta + plim \frac{\sum y_t \epsilon_t}{\sum y_t^2.} \quad (2)$$

Now, note that we can write both endogenous variables, c_t and y_t, in terms of the exogenous variables alone. To obtain the expression for c_t we substitute the identity for y_t into the right hand side of (4.4), move the c_t terms to the left hand side and then multiply by the reciprocal of $1 - \beta$. This yields the desired expression for aggregate consumption:

$$c_t = \frac{\beta}{1-\beta} i_t + \frac{\beta}{1-\beta} g_t + \frac{\epsilon_t}{1-\beta}. \quad (3)$$

Similarly we substitute for c_t on the right side of (4.5), move the y_t terms to the left side and multiple again by the reciprocal of $1 - \beta$ to obtain the desired expression for national income:

$$y_t = \frac{1}{1-\beta} i_t + \frac{1}{1-\beta} g_t + \frac{\epsilon_t}{1-\beta}. \quad (4)$$

But return to our equation for the probability limit of the OLS estimate of β. Consider the second term on the right side of equation (4.7). Note that we now have

$$plim \frac{\sum y_t \epsilon_t}{\sum y_t^2} = \frac{Cov(i_t, \epsilon_t) + Cov(g_t, \epsilon_t) + Var(\epsilon_t)}{(1-\beta) Var(y_t)} \quad (5)$$

where Cov and Var denote the covariance and variance, respectively. We assumed that i_t and g_t are exogenous. So by assumption, their covariances with ϵ_t will be zero. However, the third term in the numerator on the right side of (4.10) will not be zero; it is always a positive number. The

[4] This passage follows Pindyck and Rubinfeld 1998: 341-342. Note that the variables are in deviation form, for instance, $q_t = Q_t - \overline{Q}$.

denominator in this quotient always will be positive as well, $Var(y_t)$. From these facts and our assumption that the MPC is always between zero and unity, we know that

$$plim \frac{\sum y_t \epsilon_t}{\sum y_t^2} = \frac{Var(\epsilon_t)}{(1-\beta)Var(y_t)} > 0. \tag{6}$$

Therefore, the OLS estimate of β will be biased and consistent. In this example, the bias will be positive. This is called "simultaneous equation bias."[5] The risk of simultaneous equation bias necessitates that analysts choose a method other than OLS to estimate structural models like the one discussed above. We now turn one of these alternatives to OLS: Indirect Least Squares.

2.3 The Indirect Least Squares Method of Estimation

From a system of structural equations we can derive another, related set of equations that can be estimated with OLS. This derived set of equations is called the *reduced form* of the system. In the reduced form the endogenous variables appear only on the left side of the equal sign; the right side is composed only of predetermined variables and stochastic disturbances. These predetermined variables are, *by construction (theory)*, not correlated with the disturbance terms. Hence, OLS will yield consistent estimates of the coefficients in the reduced form. Equations (4.8) and (4.9) from the example in the previous subsection constitute the reduced form of our simple, structural Keynesian model. Note that in both these equations, the endogenous variables appear alone on the left hand sides and the exogenous variables, i_t and g_t, appear alone on the right hand side.

Let us work through another example. Consider another structural economic model in deviations form:[6]

$$q_t^S = \alpha_2 p_t + \epsilon_t \tag{7}$$

$$q_t^D = \beta_2 p_t + \beta_3 y_t + \mu_t. \tag{8}$$

where q_t^S is the quantity supplied, q_t^D is the quantity demanded, p_t is the price, y_t is national income, and ϵ_t, mu_t are error terms. The reduced form of this systems of equations can be obtained by assuming that the market clears, $q^D = q^S$. With this assumption we set the two equations equal and solve for p_t:

$$p_t = \pi_{22} y_t + v_{2t} \tag{9}$$

where $\pi_{22} = \frac{\beta_3}{\alpha_2 - \beta_2}$ and $v_{2t} = \frac{\mu_t - \epsilon_2}{\alpha_2 - \beta_2}$. Substituting this value for p_t back into (4.13) and solving for q_t we obtain the other equation in the reduced form:

$$q_t = \pi_{12} y_t + v_{1t}. \tag{10}$$

where $\pi_{12} = \frac{\alpha_2 \beta_3}{\alpha_2 - \beta_2}$ and $v_{1t} = \frac{\alpha_2 \mu_t - \beta_2 \epsilon_t}{\alpha_2 - \beta_2}$.

Once again, equations (4.14) and (4.15) are called the reduced form because price and quantity are both expressed only in terms of the exogenous variable, y_t. OLS therefore can be used to estimate both reduced form equations because y_t is (assumed to be) uncorrelated with the disturbance

[5] The direction of the bias in the OLS esimtates is clear here. But, as Pindyck and Rubinfeld (1998) explain, this is not always the case.

[6] This example also is taken from Pindyck and Rubinfeld, 1998: 340, 346-7

terms, v_{1t}, v_{2t}.
To recover the structural coefficient in the original system in the supply equation, (4.12), we can simply manipulate the reduced form coefficients. For example, $\frac{\pi_{12}}{\pi_{22}} = \alpha_2$. In terms of our notation, using OLS, we can estimate each of the reduced form coefficients and then solve for the structural coefficient in the supply equation: $\hat{\alpha}_2 = \frac{\hat{\pi_{12}}}{\hat{\pi_{22}}}$. This procedure is called *indirect least squares estimation*. It can yield consistent estimates of the structural coefficients.

But will this method always work? Can the structural coefficients always be retrieved from the reduced form coefficients? Unfortunately, the answer is no. For example, there is no way to use indirect least squares to obtain estimates of the structural coefficients in our demand equation, (4.13). Whether or not we can use indirect least squares depends upon the structure of our system of equations. There are three possibilities, and whether or not indirect least squares can be used depends on which of the following three cases a particular system of equations falls into.

2.4 The Identification Problem

There are three cases. We may have a system that is unidentified. In this case, it is impossible to derive the structural coefficients from the (indirect least squares) estimates of the reduced form coefficients (as with our demand equation, (4.13) above). Second, we might have a system that is exactly identified. OLS estimation of the reduced form, by the method of indirect least squares, then yields consistent, unique estimates of the structural coefficients. Finally, our system could be over-identified. We then have a problem of "observational equivalence" (Greene 2004: Section 15.3); more than one numerical value is produced for some of the structural coefficients. This occurs because a particular reduced form equation may represent more than one set of structural equations.[7]

Above we illustrated an exactly identified system, showing how a structural coefficient in one equation can be recovered from a reduced form. To illustrate the other two cases, consider some other simple economic models in which price is determined simultaneously by supply and demand.

- Unidentified

Demand Equation: $Q^d_t = \alpha_0 + \alpha_1 P_t + u_{1t}, \quad \alpha_1 < 0$

Supply Equation: $Q^s_t = \beta_0 + \beta_1 P_t + u_{2t}, \quad \beta_1 > 0$

Equilibrium Condition: $Q^d_t = Q^s_t$

Equating supply and demand we have:

$$\alpha_0 + \alpha_1 P_t + u_{1t} = \beta_0 + \beta_1 P_t + u_{2t} \tag{11}$$

Collecting terms we have:

$$(\alpha_1 - \beta_1)P_t = \beta_0 - \alpha_0 + \mu_{2t} \tag{12}$$

[7]It is important to stress that these three cases are the result of the structure of our equation system and not the data sample. In this sense, identification is not an estimation problem per se. It is an issue of whether the coefficients are estimable. On this important point see Manksi 1995; see also Greene 2003: Section 15.1

Multiplying on both sides of (4.17) by the reciprocal of the coefficient on P_t yields an equation of the form:

$$P_t = \pi_0 + \nu_t \tag{13}$$

where,

$$\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$$

$$\nu_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$$

Now, substitute (4.18) into either the demand equation or the supply equation and solve for the equilibrium quantity. We obtain another equation of the form:

$$Q_t = \pi_1 + \omega_t \tag{14}$$

where,

$$\pi_1 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}$$

$$\omega_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

Equations (4.18) and (4.19) are the reduced form equations for this model. Each equation has only one coefficient. There are, however, four structural coefficients, α_0, α_1, β_0, β_1. Therefore, the structural coefficients cannot be uniquely determined (i.e. there are more unknowns than equations).

- Over-Identified

Demand Function:	$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t}$
Supply Equation:	$Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 P_{t-2} + u_{2t}$
Equilibrium Condition:	$Q_t^d = Q_t^s$

where all the variables are as previously defined, and:
$I =$ Income
$R =$ Wealth.
We assume that in this economy production depends on two lags of the previous price. Equating the supply and demand yields the equilibrium price and quantities:

$$P_t = \pi_{10} + \pi_{11} I_t + \pi_{12} R_t + \pi_{13} P_{t-1} \pi_{14} P_{t-2} + \nu_t \tag{15}$$

$$Q_t = \pi_{20} + \pi_{21} I_t + \pi_{22} R_t + \pi_{23} P_{t-1} + \pi_{24} P_{t-2} + \omega_t \tag{16}$$

where,

$$\pi_{10} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \qquad \pi_{11} = \frac{-\alpha_2}{\alpha_1 - \beta_1}$$

$$\pi_{12} = \frac{-\alpha_3}{\alpha_1 - \beta_1} \qquad \pi_{13} = \frac{\beta_2}{\alpha_1 - \beta_1}$$

$$\pi_{14} = \frac{\beta_3}{\alpha_1 - \beta_1}$$

$$\pi_{20} = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \qquad \pi_{21} = \frac{-\alpha_2\beta_1}{\alpha_1 - \beta_1}$$

$$\pi_{22} = \frac{-\alpha_3\beta_1}{\alpha_1 - \beta_1} \qquad \pi_{23} = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}$$

$$\pi_{24} = \frac{\alpha_1\beta_3}{\alpha_1 - \beta_1}$$

$$\nu_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1} \qquad \omega_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

Unique estimation of the structural coefficients is not possible. For instance, note that there are two ways to obtain an estimate of α_1:

$\alpha_1 = \frac{\pi_{23}}{\pi_{13}}$ and $\alpha_1 = \frac{\pi_{24}}{\pi_{14}}$

In other words, there now are two estimates of the price coefficient in the demand equation. Since α_1 cannot be uniquely determined, neither can any of the other coefficients. The reason is simply that α_1 appears in the denominator of *all* the other reduced form coefficients.

It should be obvious at this point that both unidentified and over-identified systems of equations present significant challenges to analysts who might wish to estimate a structural model of those systems. How can an analyst know without whether a system is exactly identified or not without going through an analysis like that presented above? There are more straightforward procedures for determining whether or not a structural equation is identified. One is to apply the **Order Condition**:

"A necessary condition for a structural equation to be identified is that the number of predetermined variables excluded from a given equation is at least as large as the number of endogenous variables included, *minus one*."[8]

Notice that this is a necessary, but not sufficient condition. It is not sufficient because the predetermined variables excluded from an equation, but not the model, may not be independent. There is not a one-to-one correspondence between reduced from and structural coefficients. A necessary *and* sufficient condition for identification is the **Rank Order Condition**:

For a model with M equations in M endogenous variables, a structural equation is identified *if*

[8] This is the version of the Order Condition presented by Pindyck and Rubinfeld (1998: 340). Greene (2003: 392) gives an alternative version: The number of exogenous variables excluded from equation j must be at least as large as the number of endogenous variables included in j.

and only if at least one non-zero determinant of order $(M-1)(M-1)$ can be constructed from the coefficients of the variables (endogenous and predetermined) excluded from that particular equation, but included in the other equations of the model. [9]

2.5 More on Estimating Systems of Equations

Much has been written about how to estimate systems of equations. Space does not allow a full review of this literature here. Simply put there are two approaches: single equation methods and systems methods. These sometimes are called limited and full information methods (Greene 2003: Section 15.4-5).

Single equation methods include Indirect Least Squares and Two Stage Least Squares. When equations are exactly identified, the former can be used. We derive the reduced form coefficients and work backwards to the structural coefficients. Two-Stage Least Squares (2SLS) estimation is used for equations that are over-identified. It is called two stage because it involves two steps. First one regresses the endogenous variable on all the predetermined variables and saves the fitted values of the respective endogenous variable. This step replaces the stochastic endogenous variables by a linear combination of non-stochastic predetermined variables. The structural equation then is estimated with the fitted values from first regression inserted in place of the original endogenous variable. The results will be consistent but not efficient, because they does not take into account the correlations between the disturbances across the equations.[10]

System (full information) methods take these cross-equation correlations into account. A common method is Three-Stage Least Squares (3SLS). The general estimation approach is as follows:

- Estimate the reduced form equations.
- Use the fitted values of the endogenous variables to obtain 2SLS estimates.
- Use residuals to estimate the cross-equations variance & covariance.
- Then produce GLS estimates.

The 3SLS parameter estimates will be more efficient because they take into account the cross-equation variance and covariance.[11] Regardless of which method chosen, readers should remember that a simple application of OLS to a structural model with endogenous variables (e.g., equations 4.4 and 4.5) is prone to simultaneous equation bias, rendering parameter estimates unusable. Thus, a method like those listed above must be employed for this type of data.

[9] Recall that the rank of a matrix is given by the largest number of linearly independent rows or columns. See the Appendix to Pindyck and Rubinfeld's (1998) Chapter 12 for an explanation of the identification conditions in terms of matrix algebra.

[10] 2SLS uses a weighted average of the predetermined variables to create an instrument for the endogenous variable. Pindyck and Rubenfeld 1998: 299. Also, if an equation system satisfies the order but not the rank condition, 2SLS is possible but not efficient. Greene 2003: 15.5.3

[11] A good exposition of 3SLS can be found in Greene 2004: 405ff. STATA references Greene's passage in its exposition of 3SLS.

3 Seemingly Unrelated Regression & Simultaneous Equations

Seemingly Unrelated Regression (SURE) is a natural analytical step from the single equation setup to a set of equations that are directly related to simultaneous equations. Seemingly Unrelated Regression is a set of equations that may be related not because they interact, but because their error terms are related.

Examples include demand functions for a commodity — exogenous shocks affect the demand for both X and Y (i.e. there is a spill-over effect of demand of X on Y). The equations are estimated as a "set" to increase efficiency or the behavior of two political parties, which is determined by certain exogenous variables. The connection between the parties lies only in the shocks that both parties experience.

In simultaneous equations, the dependent variables are determined by the simultaneous interaction of several relationships. So typically there is an endogenous variable as an independent variable, so the classical regression model cannot be used, re: it is not fixed in repeated samples.

SURE is based on the idea of a ***set*** of equations of the form:

$$\mathsf{y} = \mathsf{X}\beta + \epsilon$$

where the disturbances are correlated across equations (e.g. countries, parties).

Various methods have been employed to estimate such a set of equations. All attempt to exploit the information in the correlated errors, either contemporaneously or autoregressively, in order to achieve greater efficiency in the estimates.

OLS will yield unbiased & consistent estimates for each separate equation. However, because the approach ignores the correlation of the disturbances the estimates will ***not*** be efficient (i.e. because this approach ignores information about the mutual correlation of disturbances).

3.1 Estimation via Generalized Least Squares

This approach takes into account the variability across equations and will yield BLU (best, linear, unbiased) estimates.

The system has the following general form (see Kmenta 1986: 636):

$$\begin{aligned} y_{1t} &= \beta_{11}x_{1t,1} + \beta_{12}x_{1t,2} + \cdots + \beta_{1k_1}x_{1t,k_1} + \epsilon_{1t} \\ &\vdots \\ y_{mt} &= \beta_{m1}x_{mt,1} + \beta_{m2}x_{mt,2} + \cdots + \beta_{mk_m}x_{mt,k_m} + \epsilon_{mt} \end{aligned} \tag{17}$$

for $t = 1, 2, 3, \ldots, T$

The system can be expressed more parsimoniously using matrix notation:

$$\begin{aligned} \mathsf{y_1} &= \mathsf{X_1}\beta_\mathsf{1} + \epsilon_\mathsf{1} \\ &\vdots \\ \mathsf{y_m} &= \mathsf{X_m}\beta_\mathsf{m} + \epsilon_\mathsf{m} \end{aligned} \tag{18}$$

Each equation is expected to satisfy the assumptions of the Classical Linear Regression model (CLRM). However, if the regression disturbances in the different equations are mutually correlated then we have that:

$$E[\epsilon_m, \epsilon_p'] = \sigma_{mp}\mathsf{I}_T$$

for $m, p = 1, 2, 3, \ldots, m$

Therefore, σ_{mp} is the covariance of the disturbances of the m^{th} and p^{th} equations, which is assumed to be constant over all observations, and is the only link between the m^{th} and p^{th} equations. This is a subtle link and thus referred to as seemingly unrelated regression.

Equation (2) can be re-written as:

$$\begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{vmatrix} = \begin{vmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & X_m \end{vmatrix} \begin{vmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{vmatrix} + \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{vmatrix}$$

Each equation is expected to satisfy the assumptions of classic linear regression. Where each,
y_i is a $T \times 1$ vector of sample values on the dependent variable(s)
X_i is a $T \times k_i$ matrix of sample values on the k_i independent variables
β_i is a $k_i \times 1$ vector of coefficients

Assume that ϵ_i is normally distributed with:
$E[\epsilon_i] = 0$
$E[\epsilon_i\epsilon_i'] = \sigma_{ii}\mathsf{I}_T$

Then, the variance-covariance matrix ($\Omega = E[\epsilon\epsilon']$) can be defined as:

$$\Omega = \left| \begin{array}{cccc} E[\epsilon_1\epsilon_1'] & E[\epsilon_1\epsilon_2'] & \cdots & E[\epsilon_1\epsilon_m'] \\ E[\epsilon_2\epsilon_1'] & E[\epsilon_2\epsilon_2'] & & E[\epsilon_2\epsilon_m'] \\ \vdots & & \ddots & \vdots \\ E[\epsilon_m\epsilon_1'] & E[\epsilon_m\epsilon_2'] & \cdots & E[\epsilon_m\epsilon_m'] \end{array} \right| \left| \begin{array}{cccc} \sigma_{11}\mathsf{I}_T & \sigma_{12}\mathsf{I}_T & \cdots & \sigma_{1m}\mathsf{I}_T \\ \sigma_{21}\mathsf{I}_T & \sigma_{22}\mathsf{I}_T & & \sigma_{2m}\mathsf{I}_T \\ \vdots & & \ddots & \vdots \\ \sigma_{m1}\mathsf{I}_T & \sigma_{m2}\mathsf{I}_T & \cdots & \sigma_{mm}\mathsf{I}_T \end{array} \right|$$

Where the relevant information is contained in:

$$E[\epsilon_m\epsilon_p'] = \sigma_{mp}\mathsf{I}_T$$

σ_{mp} is the covariance of disturbances between the m^{th} and p^{th} equations, contemporaneously (which is assumed to be constant across all equations).

Estimating via GLS yields:

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

$$VCV = (\tilde{\beta} - \beta)(\tilde{\beta} - \beta)^{-1} = (\mathsf{X}'\Omega^{-1}\mathsf{X})^{-1}$$

The inclusion of Ω^{-1} (which is smaller) improves the efficiency of the estimates, especially when the disturbances are highly correlated, but the independent variables are not.

The GLS estimator will equal the OLS estimator when:

1. $\sigma_{mp} = 0$
2. The equations have exactly the same values on all the independent variables (i.e. $x_m = x_p$ $\forall$ m,p)

In these instances the regressions are not "seemingly," but actually, unrelated.

Calculation proceeds in 2-stages:

1. Estimate via OLS, obtain residuals.
2. Estimate $\hat{\Omega}$ (must estimate because the variance-covariance of regression disturbances will generally be unknown).

$$\hat{\Omega} = \left| \begin{array}{cccc} s_{11}\mathsf{I}_T & s_{12}\mathsf{I}_T & \cdots & s_{1m}\mathsf{I}_T \\ s_{21}\mathsf{I}_T & s_{22}\mathsf{I}_T & & s_{2m}\mathsf{I}_T \\ \vdots & & \ddots & \vdots \\ s_{m1}\mathsf{I}_T & s_{m2}\mathsf{I}_T & \cdots & s_{mm}\mathsf{I}_T \end{array} \right|$$

The most likely case will be that the variance-covariance matrix is unknown. Kmenta (1986: 645) tells us that "there appears to be some - though by no means a great - gain in efficiency by going from ordinary least squares to two-stage Aitken or maximum likelihood estimation. The reason for this relatively low gain in efficiency is, at least in part, the high degree of correlation between the explanatory variables in the two equations".

3.2 AR Disturbances

Returning to the basic model:

$$\begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{vmatrix} = \begin{vmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & X_m \end{vmatrix} \begin{vmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{vmatrix} + \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{vmatrix}$$

Where we now have added AR disturbances:

$$\begin{vmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{mt} \end{vmatrix} = \begin{vmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{21} & \rho_{22} & & \rho_{2m} \\ \vdots & & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & \rho_{mm} \end{vmatrix} \begin{vmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \\ \vdots \\ \epsilon_{m,t-1} \end{vmatrix} + \begin{vmatrix} \nu_{1t} \\ \nu_{2t} \\ \vdots \\ \nu_{mt} \end{vmatrix} \tag{19}$$

Which may be re-written as:

$$\epsilon_{\mathsf{t}} = \mathsf{R}\epsilon_{\mathsf{t-1}} + \nu_{\mathsf{t}} \tag{20}$$

where, ν_t are IID random vectors with mean zero (i.e. $E[\nu_t] = 0$)

And, the covariance matrix is given by:

$$E[\nu_t \nu_t'] = \Sigma = \begin{vmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2m} \\ \vdots & & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{vmatrix} \tag{21}$$

That is,

$E[\nu_{it}] = 0$
$E[\nu_{it}\nu_{js}] = \sigma_{ij}$ for $t = s$ and 0 otherwise.

The AR(1) process is represented in Equations (4.21) and (4.22) and extends the single equation case by allowing the current disturbance for a given equation to depend on previous periods' disturbances in all equations. This is known as ***Vector Autoregressive Errors***.

In the common case where R is diagonal, Equation (22) may e re-written as:

$$\epsilon_{it} = \rho_{ii}\epsilon_{i,t-1} + \nu_{it}$$

So, the value of the the disturbance ϵ_{it} does not depend on the lagged disturbance in the other equations.

The goal remains the same — estimation using GLS.

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

where,

$$E[\epsilon_i\epsilon_j'] = \Omega_{ij} = \frac{\sigma_{ij}}{1-\rho_i\rho_j}\begin{vmatrix} 1 & \rho_j & \cdots & \rho_j^{T-1} \\ \rho_i & 1 & & \rho_j^{T-2} \\ \vdots & & \ddots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \cdots & 1 \end{vmatrix}$$

There are certain computational costs associated with this approach. You need to transform each variable using standard procedures. Basically performing first differences on all but the first observation. The transformation of the first observation is slightly more complex requiring a 5-step procedure (see Judge et al., pp.488-9). Higher order AR processes can also be modelled, again, see Judge et al., Section 12.3.2)

General Observations About SURE

1. Finite Sample Properties
   (a) There is some evidence that asymptotics hold in finite samples.
   (b) The omission of the first observation does not appear to matter a great deal.
2. There exists a test of $\mathsf{R} = 0$
   (a) Assuming a diagonal R appears to be the best approach even if you know that it is not.
3. General Recommendations
   (a) If there are a sufficient number of observations it is always best to allow for contemporaneous correlation of the disturbances.
   (b) Some form of autocorrelation process may also be included. However, if the sample has less than 20 observations it will be difficult to get accurate parameter estimates (even if you are sure an AR process exists).

Franklin (1991) provides a clear example of Maximum Likelihood Heteroscedastic SURE. He emphasizes the substantive interpretation of the correlations and covariances in the model. He looks at individuals in a survey who are judging two candidates and the disturbance terms are correlated in those two separate evaluations. That is, exogenous shocks seem to impact candidate evaluations in a way that is correlated. So a substantive interpretation is put on rho (as opposed to the gain in efficiency), the correlation between the two equations as being the unaccounted for commonality between judgements. The heteroscedasticity term further introduces some variation in how the errors terms are treated and some substance to that. Asymptotically, it is the same as a 2-step GLS, but the interpretation is different – efficiency vs. capturing more substance.

4 The Critical Values for Cointegration Tests

N	Variant	Size(%)	Observations	$\beta_{infinity}$	(SE)	β_1	β_2
1	No constant	1	600	-2.5658	(.0023)	-1.960	-10.04
		5	600	-1.9393	(.0008)	-0.398	0.0
		10	560	-1.6156	(.0007)	-0.181	0.0
1	No trend	1	600	-3.4335	(.0024)	-5.999	-29.25
		5	600	-2.8621	(.0011)	-2.738	-8.36
		10	600	-2.5671	(.0009)	-1.438	-4.48
1	With trend	1	600	-3.9638	(.0019)	-8.353	-47.44
		5	600	-3.4126	(.0012)	-4.039	-17.83
		10	600	-3.1279	(.0009)	-2.418	-7.58
2	No trend	1	600	-3.9001	(.0022)	-10.534	-30.03
		5	600	-3.3377	(0012)	-5.967	-9.98
		10	600	-3.0462	(-0009)	-4.069	-5.73
2	With trend	1	600	-4.3266	(-0022)	-15.531	-34.03
		5	560	-3.7809	(.0013)	-9421	-15.06
		10	600	-3.4959	(.0009)	-7.203	-4.01
3	No trend	1	560	4.2981	(.0023)	-13.790	-46.37
		5	560	3.7429	(.0012)	-8.352	-13.41
		10	600	-3.4518	(.0010)	-6.241	-2.79
3	With trend	1	600	-4.6676	(.0022)	-18.492	-49.35
		5	600	-4.1193	(.0011)	-12.024	-13.13
		10	600	-3.8344	(.0009)	-9.188	-4.85
4	No trend	1	560	-4.6493	(.0023)	-17.188	-59.20
		5	560	-4.1000	(.0012)	-10.745	-21.57
		10	600	-3.8110	(.0009)	8.317	-5.19
4	With trend	1	600	-4.9695	(.0021)	-22.504	-50.22
		5	560	-4.4294	(.0012)	-14.501	-19.54
		10	560	-4.1474	(.0010)	-11.165	-9.88
5	No trend	1	520	-4.9587	(.0026)	-22.140	-37.29
		5	560	-4.4185	(.0013)	-13.641	-21.16
		10	600	-4.1327	(.0009)	-10.638	-5.48
5	With trend	1	600	-5.2497	(.0024)	-26.606	-49.56
		5	600	-5.2497	(.0024)	-26.606	-4956
		10	600	-4.4345	(.0010)	-13.654	-5.77
6	No trend	1	408	-5.2400	(.0029)	-26.278	-41.65
		5	480	-4.7048	(.0018)	-17.120	-11.17
		10	480	-4.4242	(.0010)	-13.347	0.0
6	With trend	1	480	-5.5127	(.0033)	-30.735	-52.50
		5	480	-4.9767	(.0017)	-20.883	-9.05
		10	480	-4.6999	(.0011)	-16.445	0.0

The Critical Values for Cointegration Tests. Source: MacKinnon 1991.