CHAPTER 9
Generalized Least Squares

9.1 Introduction

This chapter considers a more general variance covariance matrix for the disturbances. In other words, $u \sim (0, \sigma^2 I_n)$ is relaxed so that $u \sim (0, \sigma^2 \Omega)$ where Ω is a positive definite matrix of dimension $(n \times n)$. First Ω is assumed known and the BLUE for β is derived. This estimator turns out to be different from $\widehat{\beta}_{OLS}$, and is denoted by $\widehat{\beta}_{GLS}$, the Generalized Least Squares estimator of β. Next, we study the properties of $\widehat{\beta}_{OLS}$ under this nonspherical form of the disturbances. It turns out that the OLS estimates are still unbiased and consistent, but their standard errors as computed by standard regression packages are biased and inconsistent and lead to misleading inference. Section 9.3 studies some special forms of Ω and derive the corresponding BLUE for β. It turns out that heteroskedasticity and serial correlation studied in Chapter 5 are special cases of Ω. Section 9.4 introduces normality and derives the maximum likelihood estimator. Sections 9.5 and 9.6 study the way in which test of hypotheses and prediction get affected by this general variance-covariance assumption on the disturbances. Section 9.7 studies the properties of this BLUE for β when Ω is unknown, and is replaced by a consistent estimator. Section 9.8 studies what happens to the W, LR and LM statistics when $u \sim N(0, \sigma^2 \Omega)$. Section 9.9 gives another application of GLS to spatial autocorrelation.

9.2 Generalized Least Squares

The regression equation did not change, only the variance-covariance matrix of the disturbances. It is now $\sigma^2 \Omega$ rather than $\sigma^2 I_n$. However, we can rely once again on a result from matrix algebra to transform our nonspherical disturbances back to spherical form, see the Appendix to Chapter 7. This result states that for every positive definite matrix Ω, there exists a *nonsingular* matrix P such that $PP' = \Omega$. In order to use this result, we transform the original model

$$y = X\beta + u \tag{9.1}$$

by premultiplying it by P^{-1}. We get

$$P^{-1}y = P^{-1}X\beta + P^{-1}u \tag{9.2}$$

Defining y^* as $P^{-1}y$ and X^* and u^* similarly, we have

$$y^* = X^*\beta + u^* \tag{9.3}$$

with u^* having 0 mean and $\text{var}(u^*) = P^{-1}\text{var}(u)P^{-1\prime} = \sigma^2 P^{-1}\Omega P'^{-1} = \sigma^2 P^{-1}PP'P'^{-1} = \sigma^2 I_n$. Hence, the variance-covariance of the disturbances in (9.3) is a scalar times an identity matrix. Therefore, using the results of Chapter 7, the BLUE for β in (9.1) is OLS on the transformed model in (9.3)

$$\widehat{\beta}_{BLUE} = (X^{*\prime}X^*)^{-1}X^{*\prime}y^* = (X'P^{-1\prime}P^{-1}X)^{-1}X'P^{-1\prime}P^{-1}y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \tag{9.4}$$

B.H. Baltagi, *Econometrics*, Springer Texts in Business and Economics, DOI 10.1007/978-3-642-20059-5_9,

with $\text{var}(\widehat{\beta}_{BLUE}) = \sigma^2(X^{*\prime}X^*)^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}$. This $\widehat{\beta}_{BLUE}$ is known as $\widehat{\beta}_{GLS}$. Define $\Sigma = E(uu') = \sigma^2\Omega$, then Σ differs from Ω only by the positive scalar σ^2. One can easily verify that $\widehat{\beta}_{GLS}$ can be alternatively written as $\widehat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$ and that $\text{var}(\widehat{\beta}_{GLS}) = (X'\Sigma^{-1}X)^{-1}$. Just substitute $\Sigma^{-1} = \Omega^{-1}/\sigma^2$ in this last expression for $\widehat{\beta}_{GLS}$ and verify that this yields (9.4).

It is clear that $\widehat{\beta}_{GLS}$ differs from $\widehat{\beta}_{OLS}$. In fact, since $\widehat{\beta}_{OLS}$ is still a linear unbiased estimator of β, the Gauss-Markov Theorem states that it must have a variance larger than that of $\widehat{\beta}_{GLS}$. Using equation (7.5) from Chapter 7, i.e., $\widehat{\beta}_{OLS} = \beta + (X'X)^{-1}X'u$ it is easy to show that

$$\text{var}(\widehat{\beta}_{OLS}) = \sigma^2(X'X)^{-1}(X'\Omega X)(X'X)^{-1} \tag{9.5}$$

Problem 1 shows that $\text{var}(\widehat{\beta}_{OLS}) -$ $\text{var}(\widehat{\beta}_{GLS})$ is a positive semi-definite matrix. Note that $\text{var}(\widehat{\beta}_{OLS})$ is no longer $\sigma^2(X'X)^{-1}$, and hence a regression package that is programmed to compute $s^2(X'X)^{-1}$ as an estimate of the variance of $\widehat{\beta}_{OLS}$ is using the wrong formula. Furthermore, problem 2 shows that $E(s^2)$ is not in general σ^2. Hence, the regression package is also wrongly estimating σ^2 by s^2. Two wrongs do not make a right, and the estimate of $\text{var}(\widehat{\beta}_{OLS})$ is biased. The direction of this bias depends upon the form of Ω and the X matrix. (We saw some examples of this bias under heteroskedasticity and serial correlation in Chapter 5). Hence, the standard errors and t-statistics computed using this OLS regression are biased. Under heteroskedasticity, one can use the White (1980) robust standard errors for OLS. In this case, $\Sigma = \sigma^2\Omega$ in (9.5) is estimated by $\widehat{\Sigma} = diag[e_i^2]$ where e_i denotes the least squares residuals. The resulting t-statistics are robust to heteroskedasticity. Similarly Wald type statistics for H_0; $R\beta = r$ can be obtained based on $\widehat{\beta}_{OLS}$ by replacing $\sigma^2(X'X)^{-1}$ in (7.41) by (9.5) with $\widehat{\Sigma} = diag[e_i^2]$. In the presence of both serial correlation and heteroskedasticity, one can use the consistent covariance matrix estimate suggested by Newey and West (1987). This was discussed in Chapter 5.

To summarize, $\widehat{\beta}_{OLS}$ is no longer BLUE whenever $\Omega \neq I_n$. However, it is still unbiased and consistent. The last two properties do not rely upon the form of the variance-covariance matrix of the disturbances but rather on $E(u/X) = 0$ and plim $X'u/n = 0$. The standard errors of $\widehat{\beta}_{OLS}$ as computed by the regression package are biased and any test of hypothesis based on this OLS regression may be misleading.

So far we have not derived an estimator for σ^2. We know however, from the results in Chapter 7, that the transformed regression (9.3) yields a mean squared error that is an unbiased estimator for σ^2. Denote this by s^{*2} which is equal to the transformed OLS residual sum of squares divided by $(n - K)$. Let e^* denote the vector of OLS residuals from (9.3), this means that $e^* = y^* - X^*\widehat{\beta}_{GLS} = P^{-1}(y - X\widehat{\beta}_{GLS}) = P^{-1}e_{GLS}$ and

$$\begin{aligned} s^{*2} &= e^{*\prime}e^*/(n-K) = (y - X\widehat{\beta}_{GLS})'\Omega^{-1}(y - X\widehat{\beta}_{GLS})/(n-K) \\ &= e'_{GLS}\Omega^{-1}e_{GLS}/(n-K) \end{aligned} \tag{9.6}$$

Note that s^{*2} now depends upon Ω^{-1}.

Necessary and Sufficient Conditions for OLS to be Equivalent to GLS

There are several necessary and sufficient conditions for OLS to be equivalent to GLS, see Puntanen and Styan (1989) for a historical survey. For pedagogical reasons, we focus on the derivation of Milliken and Albohali (1984). Note that $y = P_X y + \bar{P}_X y$. Therefore, replacing y

in $\widehat{\beta}_{GLS}$ by this expression we get

$$\widehat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}[P_Xy + \bar{P}_Xy] = \widehat{\beta}_{OLS} + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\bar{P}_Xy$$

The last term is zero for every y if and only if

$$X'\Omega^{-1}\bar{P}_X = 0 \tag{9.7}$$

Therefore, $\widehat{\beta}_{GLS} = \widehat{\beta}_{OLS}$ if and only if (9.7) is true.

Another easy necessary and sufficient condition to check in practice is the following:

$$P_X\Omega = \Omega P_X \tag{9.8}$$

see Zyskind (1967). This involves Ω rather than Ω^{-1}. There are several applications in economics where these conditions are satisfied and can be easily verified, see Balestra (1970) and Baltagi (1989). We will apply these conditions in Chapter 10 on Seemingly Unrelated Regressions, Chapter 11 on simultaneous equations and Chapter 12 on panel data. See also problem 9.

9.3 Special Forms of Ω

If the disturbances are heteroskedastic but not serially correlated, then $\Omega = \text{diag}[\sigma_i^2]$. In this case, $P = \text{diag}[\sigma_i]$, $P^{-1} = \Omega^{-1/2} = \text{diag}[1/\sigma_i]$ and $\Omega^{-1} = \text{diag}[1/\sigma_i^2]$. Premultiplying the regression equation by $\Omega^{-1/2}$ is equivalent to dividing the i-th observation of this model by σ_i. This makes the new disturbance u_i/σ_i have 0 mean and homoskedastic variance σ^2, leaving properties like no serial correlation intact. The new regression runs $y_i^* = y_i/\sigma_i$ on $X_{ik}^* = X_{ik}/\sigma_i$ for $i = 1, 2, \ldots, n$, and $k = 1, 2, \ldots, K$. Specific assumptions on the form of these σ_i's were studied in the heteroskedasticity chapter.

If the disturbances follow an AR(1) process $u_t = \rho u_{t-1} + \epsilon_t$ for $t = 1, 2, \ldots, T$; with $|\rho| \leq 1$ and $\epsilon_t \sim$IID$(0, \sigma_\epsilon^2)$, then $\text{cov}(u_t, u_{t-s}) = \rho^s\sigma_u^2$ with $\sigma_u^2 = \sigma_\epsilon^2/(1-\rho^2)$. This means that

$$\Omega = \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\ \rho & 1 & \rho & \ldots & \rho^{T-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \ldots & 1 \end{bmatrix} \tag{9.9}$$

and

$$\Omega^{-1} = \left(\frac{1}{1-\rho^2}\right)\begin{bmatrix} 1 & -\rho & 0 & \ldots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \ldots & 0 & -\rho & 1 \end{bmatrix} \tag{9.10}$$

Then

$$P^{-1} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \ldots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & -\rho & 1 \end{bmatrix} \tag{9.11}$$

is the matrix that satisfies the following condition $P^{-1\prime}P^{-1} = (1-\rho^2)\Omega^{-1}$. Premultiplying the regression model by P^{-1} is equivalent to performing the Prais-Winsten transformation. In particular the first observation on y becomes $y_1^* = \sqrt{1-\rho^2}y_1$ and the remaining observations are given by $y_t^* = (y_t - \rho y_{t-1})$ for $t = 2, 3, \ldots, T$, with similar terms for the X's and the disturbances. Problem 3 shows that the variance covariance matrix of the transformed disturbances $u^* = P^{-1}u$ is $\sigma_\epsilon^2 I_T$.

Other examples where an explicit form for P^{-1} has been derived include, (i) the MA(1) model, see Balestra (1980); (ii) the AR(2) model, see Lempers and Kloek (1973); (iii) the specialized AR(4) model for quarterly data, see Thomas and Wallis (1971); and (iv) the error components model, see Fuller and Battese (1974) and Chapter 12.

9.4 Maximum Likelihood Estimation

Assuming that $u \sim N(0, \sigma^2\Omega)$, the new likelihood function can be derived keeping in mind that $u^* = P^{-1}u = \Omega^{-1/2}u$ and $u^* \sim N(0, \sigma^2 I_n)$. In this case

$$f(u_1^*, \ldots, u_n^*; \sigma^2) = (1/2\pi\sigma^2)^{n/2} \exp\{-u^{*\prime}u^*/2\sigma^2\} \tag{9.12}$$

Making the transformation $u = Pu^* = \Omega^{1/2}u^*$, we get

$$f(u_1, \ldots, u_n; \sigma^2) = (1/2\pi\sigma^2)^{n/2}|\Omega^{-1/2}| \exp\{-u'\Omega^{-1}u/2\sigma^2\} \tag{9.13}$$

where $|\Omega^{-1/2}|$ is the Jacobian of the inverse transformation. Finally, substituting $y = X\beta + u$ in (9.13), one gets the likelihood function

$$L(\beta, \sigma^2; \Omega) = (1/2\pi\sigma^2)^{n/2}|\Omega^{-1/2}| \exp\{-(y - X\beta)'\Omega^{-1}(y - X\beta)/2\sigma^2\} \tag{9.14}$$

since the Jacobian of this last transformation is 1. Knowing Ω, maximizing (9.14) with respect to β is equivalent to minimizing $u^{*\prime}u^*$ with respect to β. This means that $\widehat{\beta}_{MLE}$ is the OLS estimate on the transformed model, i.e., $\widehat{\beta}_{GLS}$. From (9.14), we see that this RSS is a weighted one with the weight being the inverse of the variance covariance matrix of the disturbances. Similarly, maximizing (9.14) with respect to σ^2 gets $\widehat{\sigma}^2_{MLE}$ = the OLS residual sum of squares of the transformed regression (9.3) divided by n. From (9.6) this can be written as $\widehat{\sigma}^2_{MLE} = e^{*\prime}e^*/n = (n-K)s^{*2}/n$. The distributions of these maximum likelihood estimates can be derived from the transformed model using the results in Chapter 7. In fact, $\widehat{\beta}_{GLS} \sim N(\beta, \sigma^2(X'\Omega^{-1}X)^{-1})$ and $(n-K)s^{*2}/\sigma^2 \sim \chi^2_{n-K}$.

9.5 Test of Hypotheses

In order to test H_0; $R\beta = r$, under the general variance-covariance matrix assumption, one can revert to the transformed model (9.3) which has a scalar identity variance-covariance matrix and use the test statistic derived in Chapter 7

$$(R\widehat{\beta}_{GLS} - r)'[R(X^{*\prime}X^*)^{-1}R']^{-1}(R\widehat{\beta}_{GLS} - r)/\sigma^2 \sim \chi^2_g \tag{9.15}$$

Note that $\widehat{\beta}_{GLS}$ replaces $\widehat{\beta}_{OLS}$ and X^* replaces X. Replacing X^* by $P^{-1}X$, we get

$$(R\widehat{\beta}_{GLS} - r)'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\widehat{\beta}_{GLS} - r)/\sigma^2 \sim \chi^2_g \tag{9.16}$$

This differs from its counterpart in the spherical disturbances model in two ways. $\widehat{\beta}_{GLS}$ replaces $\widehat{\beta}_{OLS}$, and $(X'\Omega^{-1}X)$ takes the place of $X'X$. One can also derive the restricted estimator based on the transformed model by simply replacing X^* by $P^{-1}X$ and the OLS estimator of β by its GLS counterpart. Problem 4 asks the reader to verify that the restricted GLS estimator is

$$\widehat{\beta}_{RGLS} = \widehat{\beta}_{GLS} - (X'\Omega^{-1}X)^{-1}R'[R(X'\Omega^{-1}X)^{-1}R']^{-1}(R\widehat{\beta}_{GLS} - r) \tag{9.17}$$

Furthermore, using the same analysis given in Chapter 7, one can show that (9.15) is in fact the Likelihood Ratio statistic and is equal to the Wald and Lagrangian Multiplier statistics, see Buse (1982). In order to operationalize these tests, we replace σ^2 by its unbiased estimate s^{*2}, and divide by g the number of restrictions. The resulting statistic is an $F(g, n-K)$ for the same reasons given in Chapter 7.

9.6 Prediction

How is prediction affected by nonspherical disturbances? Suppose we want to predict one period ahead. What has changed with a general Ω? For one thing, we now know that the period $(T+1)$ disturbance is correlated with the sample disturbances. Let us assume that this correlation is given by the $(T \times 1)$ vector $\omega = E(u_{T+1}u)$, problem 5 shows that the BLUP for y_{T+1} is

$$\widehat{y}_{T+1} = x'_{T+1}\widehat{\beta}_{GLS} + \omega'\Omega^{-1}(y - X\widehat{\beta}_{GLS})/\sigma^2 \tag{9.18}$$

The first term is as expected, however it is the second term that highlights the difference between the spherical and nonspherical model predictions. To illustrate this, let us look at the AR(1) case where $\text{cov}(u_t, u_{t-s}) = \rho^s\sigma_u^2$. This implies that $\omega' = \sigma_u^2(\rho^T, \rho^{T-1}, \ldots, \rho)$. Using Ω which is given in (9.9), one can show that ω is equal to $\rho\sigma_u^2$ multiplied by the last column of Ω. But $\Omega^{-1}\Omega = I_T$, therefore, Ω^{-1} times the last column of Ω gives the last column of the identity matrix, i.e., $(0, 0, \ldots, 1)'$. Substituting for the last column of Ω its expression $(\omega/\rho\sigma_u^2)$ one gets, $\Omega^{-1}(\omega/\rho\sigma_u^2) = (0, 0, \ldots, 1)'$. Transposing and rearranging this last expression, we get $\omega'\Omega^{-1}/\sigma_u^2 = \rho(0, 0, \ldots, 1)$. This means that the last term in (9.18) is equal to $\rho(0, 0, \ldots, 1)(y - X\widehat{\beta}_{GLS}) = \rho e_{T,GLS}$, where $e_{T,GLS}$ is the T-th GLS residual. This differs from the spherical model prediction in that next year's disturbance is not independent of the sample disturbances and hence, is not predicted by its mean which is zero. Instead, one uses the fact that $u_{T+1} = \rho u_T + \epsilon_{T+1}$ and predicts u_{T+1} by $\rho e_{T,GLS}$. Only ϵ_{T+1} is predicted by its zero mean but u_T is predicted by $e_{T,GLS}$.

9.7 Unknown Ω

If Ω is unknown, the practice is to get a consistent estimate of Ω, say $\widehat{\Omega}$ and substitute that in $\widehat{\beta}_{GLS}$. The resulting estimator is

$$\widehat{\beta}_{FGLS} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y \tag{9.19}$$

and is called a feasible GLS estimator of β. Once $\widehat{\Omega}$ replaces Ω the Gauss-Markov Theorem no longer necessarily holds. In other words, $\widehat{\beta}_{FGLS}$ is not BLUE, although it is still consistent.

The finite sample properties of $\widehat{\beta}_{FGLS}$ are in general difficult to derive. However, we have the following asymptotic results.

Theorem 1: $\sqrt{n}(\widehat{\beta}_{GLS}-\beta)$ and $\sqrt{n}(\widehat{\beta}_{FGLS}-\beta)$ have the *same asymptotic distribution* $N(0,\sigma^2 Q^{-1})$, where $Q=\lim(X'\Omega^{-1}X)/n$ as $n\to\infty$, if (i) plim $X'(\widehat{\Omega}^{-1}-\Omega^{-1})X/n=0$ and (ii) plim $X'(\widehat{\Omega}^{-1}-\Omega^{-1})u/n=0$. A sufficient condition for this theorem to hold is that $\widehat{\Omega}$ is a consistent estimator of Ω and X has a satisfactory limiting behavior.

Lemma 1: If in addition plim $u'(\widehat{\Omega}^{-1}-\Omega^{-1})u/n=0$, then $s^{*2}=e'_{GLS}\Omega^{-1}e_{GLS}/(n-K)$ and $\widehat{s}^{*2}=e'_{FGLS}\widehat{\Omega}^{-1}e_{FGLS}/(n-K)$ are both *consistent* for σ^2. This means that one can perform test of hypotheses based on asymptotic arguments using $\widehat{\beta}_{FGLS}$ and $\widehat{s}^{*2}$ rather than $\widehat{\beta}_{GLS}$ and s^{*2}, respectively. For a proof of Theorem 1 and Lemma 1, see Theil (1971), Schmidt (1976) or Judge et al. (1985).

Monte Carlo evidence under heteroskedasticity or serial correlation suggest that there is gain in performing feasible GLS rather than OLS in finite samples. However, we have also seen in Chapter 5 that performing a two-step Cochrane-Orcutt procedure is not necessarily better than OLS if the X's are trended. This says that feasible GLS omitting the first observation (in this case Cochrane-Orcutt) may not be better in finite samples than OLS using all the observations.

9.8 The W, LR and LM Statistics Revisited

In this section we present a simplified and more general proof of $W\geq LR\geq LM$ due to Breusch (1979). For the general linear model given in (9.1) with $u\sim N(0,\Sigma)$ and H_0; $R\beta=r$. The likelihood function given in (9.14) with $\Sigma=\sigma^2\Omega$, can be maximized with respect to β and Σ without imposing H_0, yielding the unrestricted estimators $\widehat{\beta}_u$ and $\widehat{\Sigma}$, where $\widehat{\beta}_u=(X'\widehat{\Sigma}^{-1}X)^{-1}X'\widehat{\Sigma}^{-1}y$. Similarly, this likelihood can be maximized subject to the restriction H_0, yielding $\widetilde{\beta}_r$ and $\widetilde{\Sigma}$, where

$$\widetilde{\beta}_r=(X'\widetilde{\Sigma}^{-1}X)^{-1}X'\widetilde{\Sigma}^{-1}y-(X'\widetilde{\Sigma}^{-1}X)^{-1}R'\widetilde{\mu} \tag{9.20}$$

as in (9.17), where $\widetilde{\mu}=\widetilde{A}^{-1}(R\widehat{\beta}_r-r)$ is the Lagrange multiplier described in equation (7.35) of Chapter 7 and $\widetilde{A}=[R(X'\widetilde{\Sigma}^{-1}X)^{-1}R']$. The major distinction from Chapter 7 is that Σ is unknown and has to be estimated. Let $\widehat{\beta}_r$ denote the unrestricted maximum likelihood estimator of β conditional on the restricted variance-covariance estimator $\widetilde{\Sigma}$ and let $\widetilde{\beta}_u$ denote the restricted maximum likelihood of β (satisfying H_0) conditional on the unrestricted variance-covariance estimator $\widehat{\Sigma}$. More explicitly,

$$\widehat{\beta}_r=(X'\widetilde{\Sigma}^{-1}X)^{-1}X'\widetilde{\Sigma}^{-1}y \tag{9.21}$$

and

$$\widetilde{\beta}_u=\widehat{\beta}_u-(X'\widehat{\Sigma}^{-1}X)^{-1}R'\widehat{A}^{-1}(R\widehat{\beta}_u-r) \tag{9.22}$$

Knowing Σ, the Likelihood Ratio statistic is given by

$$\begin{aligned} LR &= -2\text{log}[\max_{R\beta=r} L(\beta/\Sigma)/\max_{\beta} L(\beta/\Sigma)]=-2\text{log}[L(\widetilde{\beta},\Sigma)/L(\widehat{\beta},\Sigma)] \qquad (9.23)\\ &= \widetilde{u}'\Sigma^{-1}\widetilde{u}-\widehat{u}'\Sigma^{-1}\widehat{u} \end{aligned}$$

where $\widehat{u} = y - X\widehat{\beta}$ and $\widetilde{u} = y - X\widetilde{\beta}$, both estimators of β are conditional on a *known* Σ.

$$R\widehat{\beta}_u \sim N(R\beta, R(X'\widehat{\Sigma}^{-1}X)^{-1}R')$$

and the Wald statistic is given by

$$W = (R\widehat{\beta}_u - r)'\widehat{A}^{-1}(R\widehat{\beta}_u - r) \quad \text{where} \quad \widehat{A} = [R(X'\widehat{\Sigma}^{-1}X)^{-1}R'] \tag{9.24}$$

Using (9.22), it is easy to show that $\widetilde{u}_u = y - X\widetilde{\beta}_u$ and $\widehat{u}_u = y - X\widehat{\beta}_u$ are related as follows:

$$\widetilde{u}_u = \widehat{u}_u + X(X'\widehat{\Sigma}^{-1}X)^{-1}R'\widehat{A}^{-1}(R\widehat{\beta}_u - r) \tag{9.25}$$

and

$$\widetilde{u}_u'\widehat{\Sigma}^{-1}\widetilde{u}_u = \widehat{u}_u'\widehat{\Sigma}^{-1}\widehat{u}_u + (R\widehat{\beta}_u - r)'\widehat{A}^{-1}(R\widehat{\beta}_u - r) \tag{9.26}$$

The cross-product terms are zero because $X'\widehat{\Sigma}^{-1}\widehat{u}_u = 0$. Therefore,

$$\begin{aligned} W &= \widetilde{u}_u'\widehat{\Sigma}^{-1}\widetilde{u}_u - \widehat{u}_u'\widehat{\Sigma}^{-1}\widehat{u}_u = -2\log[L(\widetilde{\beta}, \widehat{\Sigma})/L(\widehat{\beta}, \widehat{\Sigma})] \\ &= -2\log[\max_{R\beta=r} L(\beta/\widehat{\Sigma})/\max_{\beta} L(\beta/\widehat{\Sigma})] \end{aligned} \tag{9.27}$$

and the Wald statistic can be interpreted as a LR statistic conditional on $\widehat{\Sigma}$, the unrestricted maximum likelihood estimator of Σ.

Similarly, the Lagrange multiplier statistic, which tests that $\mu = 0$, is given by

$$LM = \mu'\widetilde{A}\mu = (R\widehat{\beta}_r - r)'\widetilde{A}^{-1}(R\widehat{\beta}_r - r) \tag{9.28}$$

Using (9.20) one can easily show that

$$\widetilde{u}_r = \widehat{u}_r + X(X'\widetilde{\Sigma}^{-1}X)^{-1}R'\widetilde{A}^{-1}(R\widehat{\beta}_r - r) \tag{9.29}$$

and

$$\widetilde{u}_r'\widetilde{\Sigma}^{-1}\widetilde{u}_r = \widehat{u}_r'\widetilde{\Sigma}^{-1}\widehat{u}_r + \mu'\widetilde{A}\mu \tag{9.30}$$

The cross-product terms are zero because $X'\widetilde{\Sigma}^{-1}\widehat{u}_r = 0$. Therefore,

$$\begin{aligned} LM &= \widetilde{u}_r'\widetilde{\Sigma}^{-1}\widetilde{u}_r - \widehat{u}_r'\widetilde{\Sigma}^{-1}\widehat{u}_r = -2\log[L(\widetilde{\beta}_r, \widetilde{\Sigma})/L(\widehat{\beta}_r, \widetilde{\Sigma})] \\ &= -2\log[\max_{R\beta=r} L(\beta/\widetilde{\Sigma})/\max_{\beta} L(\beta/\widetilde{\Sigma})] \end{aligned} \tag{9.31}$$

and the Lagrange multiplier statistic can be interpreted as a LR statistic conditional on $\widetilde{\Sigma}$ the restricted maximum likelihood of Σ. Given that

$$\max_{\beta} L(\beta/\widetilde{\Sigma}) \leq \max_{\beta,\Sigma} L(\beta, \Sigma) = \max_{\beta} L(\beta/\widehat{\Sigma}) \tag{9.32}$$

$$\max_{R\beta=r} L(\beta/\widehat{\Sigma}) \leq \max_{R\beta=r,\Sigma} L(\beta, \Sigma) = \max_{R\beta=r} L(\beta/\widetilde{\Sigma}) \tag{9.33}$$

it can be easily shown that the likelihood ratio statistic given by

$$LR = -2\log[\max_{R\beta=r,\Sigma} L(\beta, \Sigma)/\max_{\beta,\Sigma} L(\beta, \Sigma)] \tag{9.34}$$

satisfies the following inequality

$$W \geq LR \geq LM \tag{9.35}$$

The proof is left to the reader, see problem 6.

This general and simple proof holds as long as the maximum likelihood estimator of β is uncorrelated with the maximum likelihood estimator of Σ, see Breusch (1979).

9.9 Spatial Error Correlation[1]

Unlike time-series, there is typically no unique natural ordering for cross-sectional data. Spatial autocorrelation permit correlation of the disturbance terms across cross-sectional units. There is an extensive literature on spatial models in regional science, urban economics, geography and statistics, see Anselin (1988). Examples in economics usually involve spillover effects or externalities due to geographical proximity. For example, the productivity of public capital, like roads and highways, on the output of neighboring states. Also, the pricing of welfare in one state that pushes recipients to other states. Spatial correlation could relate directly to the model dependent variable y, the exogenous variables X, the disturbance term u, or to a combination of all three. Here we consider spatial correlation in the disturbances and leave the remaining literature on spatial dependence to the motivated reader to pursue in Anselin (1988, 2001) and Anselin and Bera (1998) to mention a few.

For the cross-sectional disturbances, the spatial autocorrelation is specified as

$$u = \lambda W u + \epsilon \tag{9.36}$$

where λ is the spatial autoregressive coefficient satisfying $|\lambda| < 1$, and $\epsilon \sim \text{IIN}(0, \sigma^2)$. W is a *known* spatial weight matrix with diagonal elements equal to zero. W also satisfies some other regularity conditions like the fact that $I_n - \lambda W$ must be nonsingular.

The regression model given in (9.1) can be written as

$$y = X\beta + (I_n - \lambda W)^{-1}\epsilon \tag{9.37}$$

with the variance-covariance matrix of the disturbances given by

$$\Sigma = \sigma^2\Omega = \sigma^2(I_n - \lambda W)^{-1}(I_n - \lambda W')^{-1} \tag{9.38}$$

Under normality of the disturbances, Ord (1975) derived the maximum likelihood estimators

$$\text{ln}L = -\frac{1}{2}\text{ln}|\Omega| - \frac{n}{2}\text{ln}2\pi\sigma^2 - (y - X\beta)'\Omega^{-1}(y - X\beta)/2\sigma^2 \tag{9.39}$$

The Jacobian term simplifies by using

$$\text{ln}|\Omega| = -2\text{ln}|I - \lambda W| = -2\textstyle\sum_{i=1}^{n} \ln(1 - \lambda w_i) \tag{9.40}$$

where w_i are the eigenvalues of the spatial weight matrix W. The first-order conditions yield the familiar GLS estimator of β and the associated estimator of σ^2:

$$\widehat{\beta}_{MLE} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \quad \text{and} \quad \widehat{\sigma}^2_{MLE} = e'_{MLE}\Omega^{-1}e_{MLE}/n \tag{9.41}$$

where $e_{MLE} = y - X\widehat{\beta}_{MLE}$. An estimate of λ can be obtained using the iterative solution of the first-order conditions in Magnus (1978, p. 283):

$$-\frac{1}{2}\,\text{tr}\left[\left(\frac{\partial\Omega^{-1}}{\partial\lambda}\right)\Omega\right] = e'_{MLE}\left(\frac{\partial\Omega^{-1}}{\partial\lambda}\right)e_{MLE} \tag{9.42}$$

where

$$\partial\Omega^{-1}/\partial\lambda = -W - W' + \lambda W'W \tag{9.43}$$

Alternatively, one can substitute $\widehat{\beta}_{MLE}$ and $\widehat{\sigma}^2_{MLE}$ from (9.41) into the log-likelihood in (9.39) to get the concentrated log-likelihood which will be a nonlinear function of λ, see Anselin (1988) for details.

Testing for zero spatial autocorrelation i.e., H_0; $\lambda = 0$ is usually based on the Moran I-test which is similar to the Durbin-Watson statistic in time-series. This is given by

$$MI = \frac{n}{S_0}\left(\frac{e'We}{e'e}\right) \tag{9.44}$$

where e denotes the vector of OLS residuals and S_0 is a standardization factor equal to the sum of the spatial weights $\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{w}_{ij}$. For a row-standardized weights matrix W where each row sums to one, $S_0 = n$ and the Moran I-statistic simplifies to $e'We/e'e$. In practice the test is implemented by standardizing it and using the asymptotic $N(0,1)$ critical values, see Anselin and Bera (1988). In fact, for a row-standardized W matrix, the mean and variance of the Moran I-statistic is obtained from

$$E(MI) = E\left(\frac{e'We}{e'e}\right) = \mathrm{tr}(\bar{P}_X W)/(n-k) \tag{9.45}$$

and

$$E(MI)^2 = \frac{\mathrm{tr}(\bar{P}_X W \bar{P}_X W') + \mathrm{tr}(\bar{P}_X W)^2 + \{\mathrm{tr}(\bar{P}_{\bar{X}} W)\}^2}{(n-k)(n-k+2)}$$

Alternatively, one can derive the Lagrange Multiplier test for H_0; $\lambda = 0$ using the result that $\partial \ln L/\partial\lambda$ evaluated under the null of $\lambda = 0$ is equal to $u'Wu/\sigma^2$ and the fact that the Information matrix is block-diagonal between β and (σ^2, λ), see problem 14. In fact, one can show that

$$LM_\lambda = \frac{(e'We/\widetilde{\sigma}^2)^2}{\mathrm{tr}[(W'+W)W]} \tag{9.46}$$

with $\widetilde{\sigma}^2 = e'e/n$. Under H_0, LM_λ is asymptotically distributed as χ^2_1. One can clearly see the connection between Moran's I-statistic and LM_λ. Computationally, the W and LR tests are more demanding since the require ML estimation under spatial autocorrelation.

This is only a brief introduction into the spatial dependence literature. Hopefully, it will motivate the reader to explore alternative formulations of spatial dependence, alternative estimation and testing methods discussed in this literature and the numerous applications in economics on hedonic housing, crime rates, police expenditures and R&D spillovers, to mention a few.

Note

1. This section is based on Anselin (1988, 2001) and Anselin and Bera (1998).

Problems

1. *GLS Is More Efficient than OLS.*

   (a) Using equation (7.5) of Chapter 7, verify that $\text{var}(\widehat{\beta}_{OLS})$ is that given in (9.5).

   (b) Show that $\text{var}(\widehat{\beta}_{OLS})-$ $\text{var}(\widehat{\beta}_{GLS}) = \sigma^2 A\Omega A'$ where

   $$A = [(X'X)^{-1}X' - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}].$$

   Conclude that this difference in variances is positive semi-definite.

2. *s^2 Is No Longer Unbiased for σ^2.*

   (a) Show that $E(s^2) = \sigma^2\text{tr}(\Omega\bar{P}_X)/(n-K) \neq \sigma^2$. **Hint:** Follow the same proof given below equation (7.6) of Chapter 7, but substitute $\sigma^2\Omega$ instead of $\sigma^2 I_n$.

   (b) Use the fact that P_X and Σ are non-negative definite matrices with $\text{tr}(\Sigma P_X) \geq 0$ to show that $0 \leq E(s^2) \leq \text{tr}(\Sigma)/(n-K)$ where $\text{tr}(\Sigma) = \sum_{i=1}^{n}\sigma_i^2$ with $\sigma_i^2 = \text{var}(u_i) \geq 0$. This bound was derived by Dufour (1986). Under homoskedasticity, show that this bound becomes $0 \leq E(s^2) \leq n\sigma^2/(n-K)$. In general, $0 \leq$ {mean of $n-K$ smallest characteristic roots of Σ} $\leq E(s^2) \leq$ {mean of $n-K$ largest characteristic roots of Σ} $\leq \text{tr}(\Sigma)/(n-K)$, see Sathe and Vinod (1974) and Neudecker (1977, 1978).

   (c) Show that a *sufficient condition* for s^2 to be consistent for σ^2 irrespective of X is that λ_{max} = the largest characteristic root of Ω is $o(n)$, i.e., $\lambda_{max}/n \to 0$ as $n \to \infty$ *and* plim $(u'u/n) = \sigma^2$. **Hint:** $s^2 = u'\bar{P}_X u/(n-K) = u'u/(n-K) - u'P_X u/(n-K)$. By assumption, the first term tends in probability limits to σ^2 as $n \to \infty$. The second term has expectation $\sigma^2\text{tr}(P_X\Omega)/(n-K)$. Now $P_X\Omega$ has rank K and therefore exactly K non-zero characteristic roots each of which cannot exceed λ_{max}. This means that $E[u'P_X u/(n-K)] \leq \sigma^2 K\lambda_{max}/(n-K)$. Using the condition that $\lambda_{max}/n \to 0$ proves the result. See Krämer and Berghoff (1991).

   (d) Using the same reasoning in part (a), show that s^{*2} given in (9.6) is unbiased for σ^2.

3. *The AR(1) Model.* See Kadiyala (1968).

   (a) Verify that $\Omega\Omega^{-1} = I_T$ for Ω and Ω^{-1} given in (9.9) and (9.10), respectively.

   (b) Show that $P^{-1\prime}P^{-1} = (1-\rho^2)\Omega^{-1}$ for P^{-1} defined in (9.11).

   (c) Conclude that $\text{var}(P^{-1}u) = \sigma_\epsilon^2 I_T$. **Hint:** $\Omega = (1-\rho^2)PP'$ as can be easily derived from part (b).

4. *Restricted GLS.* Using the derivation of the restricted least squares estimator for $u \sim (0, \sigma^2 I_n)$ in Chapter 7, verify equation (9.17) for the restricted GLS estimator based on $u \sim (0, \sigma^2\Omega)$. **Hint:** Apply restricted least squares results to the transformed model given in (9.3).

5. *Best Linear Unbiased Prediction.* This is based on Goldberger (1962). Consider all linear predictors of $y_{T+s} = x'_{T+s}\beta + u_{T+s}$ of the form $\widehat{y}_{T+s} = c'y$, where $u \sim (0, \Sigma)$ and $\Sigma = \sigma^2\Omega$.

   (a) Show that $c'X = x'_{T+s}$ for $\widehat{y}_{T+s}$ to be unbiased.

   (b) Show that $\text{var}(\widehat{y}_{T+s}) = c'\Sigma c + \sigma^2_{T+s} - 2c'\omega$ where $\text{var}(u_{T+s}) = \sigma^2_{T+s}$ and $\omega = E(u_{T+s}u)$.

   (c) Minimize $\text{var}(\widehat{y}_{T+s})$ given in part (b) subject to $c'X = x'_{T+s}$ and show that

   $$\widehat{c} = \Sigma^{-1}[I_T - X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}]\omega + \Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}x_{T+s}$$

   This means that $\widehat{y}_{T+s} = \widehat{c}'y = x'_{T+s}\widehat{\beta}_{GLS} + \omega'\Sigma^{-1}e_{GLS} = x'_{T+s}\widehat{\beta}_{GLS} + \omega'\Omega^{-1}e_{GLS}/\sigma^2$. For $s = 1$, i.e., predicting one period ahead, this verifies equation (9.18). **Hint:** Use partitioned inverse in solving the first-order minimization equations.

(d) Show that $\widehat{y}_{T+s} = x'_{T+s}\widehat{\beta}_{GLS} + \rho^s e_{T,GLS}$ for the stationary AR(1) disturbances with autoregressive parameter ρ, and $|\rho| < 1$.

6. *The W, LR and LM Inequality.* Using the inequalities given in equations (9.32) and (9.33) verify equation (9.35) which states that $W \geq LR \geq LM$. **Hint:** Use the conditional likelihood ratio interpretations of W and LM given in equations (9.27) and (9.31) respectively.

7. Consider the *simple linear regression*

$$y_i = \alpha + \beta X_i + u_i \quad i = 1, 2, \ldots, n$$

with $u_i \sim \text{IIN}(0, \sigma^2)$. For H_0; $\beta = 0$, derive the LR, W and LM statistics in terms of conditional likelihood ratios as described in Breusch (1979). In other words, compute $W = -2\log[\max\limits_{H_0} L(\alpha, \beta/\widehat{\sigma}^2)/\max\limits_{\alpha,\beta} L(\alpha, \beta/\widehat{\sigma}^2)]$, $LM = -2\log[\max\limits_{H_0} L(\alpha, \beta/\widetilde{\sigma}^2)/\max\limits_{\alpha,\beta} L(\alpha, \beta/\widetilde{\sigma}^2)]$ and $LR = -2\log[\max\limits_{H_0} L(\alpha, \beta, \sigma^2)/\max\limits_{\alpha,\beta,\sigma^2} L(\alpha, \beta, \sigma^2)]$ where $\widehat{\sigma}^2$ is the unrestricted MLE of σ^2 while $\widetilde{\sigma}^2$ is the restricted MLE of σ^2 under H_0. Use these results to infer that $W \geq LR \geq LM$.

8. *Sampling Distributions and Efficiency Comparison of OLS and GLS.* Consider the following regression model $y_t = \beta x_t + u_t$ for $(t = 1, 2)$, where $\beta = 2$ and x_t takes on the fixed values $x_1 = 1$, $x_2 = 2$. The u_t's have the following discrete joint probability distribution:

(u_1, u_2)	Probability
$(-1, -2)$	1/8
$(1, -2)$	3/8
$(-1, 2)$	3/8
$(1, 2)$	1/8

(a) What is the variance-covariance matrix of the disturbances? Are the disturbances heteroskedastic? Are they correlated?

(b) Find the sampling distributions of $\widehat{\beta}_{OLS}$ and $\widetilde{\beta}_{GLS}$ and verify that $\text{var}(\widehat{\beta}_{OLS}) > \text{var}(\widetilde{\beta}_{GLS})$.

(c) Find the sampling distribution of the OLS residuals and verify that the estimated $\text{var}(\widehat{\beta}_{OLS})$ is biased. Also, find the sampling distribution of the GLS residuals and verify that the MSE of the GLS regression is an unbiased estimator of the GLS regression variance. **Hint:** Read Oksanen (1991) and Phillips and Wickens (1978), pp. 3–4. This problem is based on Baltagi (1992). See also the solution by Im and Snow (1993).

9. *Equi-correlation.* This problem is based on Baltagi (1998). Consider the regression model given in (9.1) with *equi-correlated* disturbances, i.e., equal variances and equal covariances: $E(uu') = \sigma^2\Omega = \sigma^2[(1-\rho)I_T + \rho\iota_T\iota'_T]$ where ι_T is a vector of ones of dimension T and I_T is the identity matrix. In this case, $\text{var}(u_t) = \sigma^2$ and $\text{cov}(u_t, u_s) = \rho\sigma^2$ for $t \neq s$ with $t = 1, 2, \ldots, T$. Assume that the regression has a constant.

(a) Show that OLS on this model is equivalent to GLS. **Hint:** Verify Zyskind's condition given in (9.8) using the fact that $P_X\iota_T = \iota_T$ if ι_T is a column of X.

(b) Show that $E(s^2) = \sigma^2(1-\rho)$. Also, that Ω is positive semi-definite when $-1/(T-1) \leq \rho \leq 1$. Conclude that if $-1/(T-1) \leq \rho \leq 1$, then $0 \leq E(s^2) \leq [T/(T-1)]\sigma^2$. The lower and upper bounds are attained at $\rho = 1$ and $\rho = -1/(T-1)$, respectively, see Dufour (1986). **Hint:** Ω is positive semi-definite if for every arbitrary non-zero vector a we have $a'\Omega a \geq 0$. What is this expression for $a = \iota_T$?

(c) Show that for this equi-correlated regression model, the BLUP of $y_{T+1} = x'_{T+1}\beta + u_{T+1}$ is $\widehat{y}_{T+1} = x'_{T+1}\widehat{\beta}_{OLS}$ as long as there is a constant in the model.

10. Consider the simple regression with no regressors and equi-correlated disturbances:

$$y_i = \alpha + u_i \quad i = 1, \ldots, n$$

where $E(u_i) = 0$ and

$$\begin{aligned} \text{cov}(u_i, u_j) &= \rho\sigma^2 \quad \text{for} \quad i \neq j \\ &= \sigma^2 \quad \text{for} \quad i = j \end{aligned}$$

with $\frac{1}{(n-1)} \leq \rho \leq 1$ for the variance-covariance matrix of the disturbances to be positive definite.

(a) Show that the OLS and GLS estimates of α are identical. This is based on Kruskal (1968).

(b) Show that the bias in s^2, the OLS estimator of σ^2, is given by $-\rho\sigma^2$.

(c) Show that the GLS estimator of σ^2 is unbiased.

(d) Show that the E[estimated var$(\widehat{\alpha})-$ true var$(\widehat{\alpha}_{OLS})$] is also $-\rho\sigma^2$.

11. *Prediction Error Variances Under Heteroskedasticity.* This is based on Termayne (1985). Consider the t-th observation of the linear regression model given in (9.1).

$$y_t = x_t'\beta + u_t \qquad t = 1, 2, \ldots, T$$

where y_t is a scalar x_t' is $1 \times K$ and β is a $K \times 1$ vector of unknown coefficients. u_t is assumed to have zero mean, heteroskedastic variances $E(u_t^2) = (z_t'\gamma)^2$ where z_t' is a $1 \times r$ vector of observed variables and γ is an $r \times 1$ vector of parameters. Furthermore, these u_t's are not serially correlated, so that $E(u_t u_s) = 0$ for $t \neq s$.

(a) Find the var$(\widehat{\beta}_{OLS})$ and var$(\widetilde{\beta}_{GLS})$ for this model.

(b) Suppose we are forecasting y for period f in the future knowing x_f, i.e., $y_f = x_f'\beta + u_f$ with $f > T$. Let $\widehat{e}_f$ and $\widetilde{e}_f$ be the forecast errors derived using OLS and GLS, respectively. Show that the prediction error variances of the point predictions of y_f are given by

$$\text{var}(\widehat{e}_f) = x_f'(\textstyle\sum_{t=1}^T x_t x_t')^{-1}[\sum_{t=1}^T x_t x_t'(z_t'\gamma)^2](\sum_{t=1}^T x_t x_t')^{-1} x_f + (z_f'\gamma)^2$$

$$\text{var}(\widetilde{e}_f) = x_f'[\textstyle\sum_{t=1}^T x_t x_t'(z_t'\gamma)^2]^{-1} x_f + (z_f'\gamma)^2$$

(c) Show that the variances of the two forecast errors of conditional mean $E(y_f/x_f)$ based upon $\widehat{\beta}_{OLS}$ and $\widetilde{\beta}_{GLS}$ and denoted by $\widehat{c}_f$ and $\widetilde{c}_f$, respectively are the first two terms of the corresponding expressions in part (b).

(d) Now assume that $K = 1$ and $r = 1$ so that there is only one single regressor x_t and one z_t variable determining the heteroskedasticity. Assume also for simplicity that the empirical moments of x_t match the population moments of a Normal random variable with mean zero and variance θ. Show that the relative efficiency of the OLS to the GLS predictor of y_f is equal to $(T+1)/(T+3)$, whereas the relative efficiency of the corresponding ratio involving the two predictions of the conditional mean is $(1/3)$.

12. *Estimation of Time Series Regressions with Autoregressive Disturbances and Missing Observations.* This is based on Baltagi and Wu (1997). Consider the following time series regression model,

$$y_t = x_t'\beta + u_t \quad t = 1, \ldots, T,$$

where β is a $K \times 1$ vector of regression coefficients including the intercept. The disturbances follow a stationary AR(1) process, that is,

$$u_t = \rho u_{t-1} + \epsilon_t,$$

with $|\rho| < 1$, ϵ_t is IIN$(0, \sigma_\epsilon^2)$, and $u_0 \sim N(0, \sigma_\epsilon^2/(1-\rho^2))$. This model is only observed at times t_j for $j = 1, \ldots, n$ with $1 = t_1 < \ldots < t_n = T$ and $n > K$. The typical covariance element of u_t for the observed periods t_j and t_s is given by

$$\text{cov}(u_{t_j}, u_{t_s}) = \frac{\sigma_\epsilon^2}{1-\rho^2}\,\rho^{|t_j - t_s|} \quad \text{for} \quad s, j = 1, \ldots, n$$

Knowing ρ, derive a simple Prais-Winsten-type transformation that will obtain GLS as a simple least squares regression.

13. *Multiplicative Heteroskedasticity.* This is based on Harvey (1976). Consider the linear model given in (9.1) and let $u \sim N(0, \Sigma)$ where $\Sigma = \text{diag}[\sigma_i^2]$. Assume that $\sigma_i^2 = \sigma^2 h_i(\theta)$ with $\theta' = (\theta_1, \ldots, \theta_s)$ and $h_i(\theta) = \exp(\theta_1 z_{1i} + \ldots + \theta_s z_{si}) = \exp(z_i'\theta)$ with $z_i' = (z_{1i}, \ldots, z_{si})$.

    (a) Show that log-likelihood function is given by

    $$\log L(\beta, \theta, \sigma^2) = -\frac{N}{2}\log 2\pi\sigma^2 - \frac{1}{2}\sum_{i=1}^N \log h_i(\theta) - \frac{1}{2\sigma^2}\sum_{i=1}^N \frac{(y_i - x_i'\beta)^2}{h_i(\theta)}$$

    and the score with respect to θ is

    $$\partial \log L/\partial\theta = -\frac{1}{2}\sum_{i=1}^N \frac{1}{h_i(\theta)}\frac{\partial h_i}{\partial\theta} + \frac{1}{2\sigma^2}\sum_{i=1}^N \frac{(y_i - x_i'\beta)^2}{(h_i(\theta))^2}\cdot\frac{\partial h_i}{\partial\theta}$$

    Conclude that for multiplicative heteroskedasticity, equating this score to zero yields

    $$\sum_{i=1}^N \frac{(y_i - x_i'\beta)^2}{\exp(z_i'\theta)} z_i = \sigma^2 \sum_{i=1}^N z_i.$$

    (b) Show that the Information matrix is given by

    $$I(\beta, \theta, \sigma^2) = \begin{bmatrix} X'\Sigma^{-1}X & 0 & 0 \\ 0 & \frac{1}{2}\sum_{i=1}^N \frac{1}{(h_i(\theta))^2}\frac{\partial h_i}{\partial\theta}\frac{\partial h_i}{\partial\theta'} & \frac{1}{2\sigma^2}\sum_{i=1}^N \frac{1}{h_i(\theta)}\frac{\partial h_i}{\partial\theta} \\ 0 & \frac{1}{2\sigma^2}\sum_{i=1}^N \frac{1}{h_i(\theta)}\frac{\partial h_i}{\partial\theta'} & \frac{N}{2\sigma^4} \end{bmatrix}$$

    and for multiplicative heteroskedasticity this becomes

    $$I(\beta, \theta, \sigma^2) = \begin{bmatrix} X'\Sigma^{-1}X & 0 & 0 \\ 0 & \frac{1}{2}Z'Z & \frac{1}{2\sigma^2}\sum_{i=1}^N z_i \\ 0 & \frac{1}{2\sigma^2}\sum_{i=1}^N z_i' & \frac{N}{2\sigma^4} \end{bmatrix}$$

    where $Z_i' = (z_1, \ldots, z_N)$.

    (c) Assume that $h_i(\theta)$ satisfies $h_i(0) = 1$, then the test for heteroskedasticity is H_0; $\theta = 0$ versus H_1; $\theta \neq 0$. Show that the score with respect to θ and σ^2 evaluated under the null hypothesis, i.e., at $\theta = 0$ and $\widetilde{\sigma}^2 = e'e/N$ is given by

    $$\widetilde{S} = \begin{pmatrix} \frac{1}{2}\sum_{i=1}^N z_i\left(\frac{e_i^2}{\widetilde{\sigma}^2} - 1\right) \\ 0 \end{pmatrix}$$

where e denotes the vector of OLS residuals. The Information matrix with respect to θ and σ^2can be obtained from the bottom right block of $I(\beta, \theta, \sigma^2)$ given in part (b). Conclude that the score test for H_0 is given by

$$LM = \frac{\sum_{i=1}^{N} z_i'(e_i^2 - \widetilde{\sigma}^2) \left(\sum_{i=1}^{N}(z_i - \bar{z})(z_i - \bar{z})'\right)^{-1} \sum_{i=1}^{N} z_i(e_i^2 - \widetilde{\sigma}^2)}{2\widetilde{\sigma}^4}$$

This statistic is asymptotically distributed as χ_s^2 under H_0. From Chapter 5, we can see that this is a special case of the Breusch and Pagan (1979) test-statistic which can be obtained as one-half the regression sum of squares of $e^2/\widetilde{\sigma}^2$ on a constant and Z. Koenker and Bassett (1982) suggested replacing the denominator $2\widetilde{\sigma}^4$ by $\sum_{i=1}^{N}(e^2 - \widetilde{\sigma}^2)^2/N$ to make this test more robust to departures from normality.

14. *Spatial Autocorrelation.* Consider the regression model given in (9.1) with spatial autocorrelation defined in (9.36).

   (a) Verify that the first-order conditions of maximization of the log-likelihood function given in (9.39) yield (9.41).

   (b) Show that for testing H_0; $\lambda = 0$, the score $\partial \ln L/\partial\lambda$ evaluated under the null, i.e., at $\lambda = 0$, is given by $u'Wu/\sigma^2$.

   (c) Show that the Information matrix with respect to $\sigma^2$and λ, evaluated under the null of $\lambda = 0$, is given by

$$\begin{bmatrix} \dfrac{n}{2\sigma^4} & \dfrac{\text{tr}(W)}{\sigma^2} \\ \dfrac{\text{tr}(W)}{\sigma^2} & \text{tr}(W^2) + \text{tr}(W'W) \end{bmatrix}$$

   (d) Conclude from parts (b) and (c) that the Lagrange Multiplier for H_0; $\lambda = 0$ is given by LM_λ in (9.46). **Hint:** Use the fact that the diagonal elements of W are zero, hence $\text{tr}(W) = 0$.

15. *Neighborhood Effects and Housing Demand.* Ioannides and Zabel (2003) use data from the American Housing Survey to estimate a model of housing demand with neighborhood effects. The number of observations on housing units used were 1947 in 1985, 2318 in 1989 and 2909 in 1993. The housing survey has detailed information for each of these housing units and their owners, including: the owner's schooling, whether the owner is white, whether the owner is married, the number of persons in the household, household income, and whether the house has changed owners ("changed hands") in the last 5 years. In addition, the current owner's evaluation of the housing unit's market value, as well as various structural characteristics of the housing unit (such as number of bedrooms, bathrooms, and whether the house has a garage). The variable definitions are given in Table VI of Ioannides and Zabel (2003, p. 568) and the data is available from the Journal of Applied Econometrics archive:

   (a) Replicate Table VII of Ioannides and Zabel (2003, p. 569) which displays the means and standard deviations for some of the variables by year and for the pooled sample. Note that the price and income variables are different from the numbers reported in the paper.

   (b) Replicate Table VIII of Ioannides and Zabel (2003, p. 577) which reports regression results of housing demand. Note that these regressions are different from those reported in the paper.

References

Additional readings on GLS can be found in the econometric texts cited in the Preface.

Anselin, L. (2001), "Spatial Econometrics," Chapter 14 in B.H. Baltagi (ed.) *A Companion to Theoretical Econometrics* (Blackwell: Massachusetts).

Anselin, L. (1988), *Spatial Econometrics: Methods and Models* (Kluwer: Dordrecht).

Anselin, L. and A.K. Bera (1998), "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," in A. Ullah and D.E.A. Giles (eds.) *Handbook of Applied Economic Statistics* (Marcel Dekker: New York).

Balestra, P. (1970), "On the Efficiency of Ordinary Least Squares in Regression Models," *Journal of the American Statistical Association*, 65: 1330–1337.

Balestra, P. (1980), "A Note on the Exact Transformation Associated with First-Order Moving Average Process," *Journal of Econometrics*, 14: 381–394.

Baltagi, B.H. (1989), "Applications of a Necessary and Sufficient Condition for OLS to be BLUE," *Statistics and Probability Letters*, 8: 457–461.

Baltagi, B.H. (1992), "Sampling Distributions and Efficiency Comparisons of OLS and GLS in the Presence of Both Serial Correlation and Heteroskedasticity," *Econometric Theory*, Problem 92.2.3, 8: 304–305.

Baltagi, B.H. and P.X. Wu (1997), "Estimation of Time Series Regressions with Autoregressive Disturbances and Missing Observations," *Econometric Theory*, Problem 97.5.1, 13: 889.

Baltagi, B.H. (1998), "Prediction in the Equicorrelated Regression Model," *Econometric Theory*, Problem 98.3.3, 14: 382.

Breusch, T.S. (1979), "Conflict Among Criteria for Testing Hypotheses: Extensions and Comments," *Econometrica*, 47: 203–207.

Breusch, T.S. and A.R. Pagan (1979), "A Simple Test for Heteroskedasticity and Random Coefficient Variation," *Econometrica*, 47: 1287–1294.

Buse, A. (1982), "The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note," *The American Statistician*, 36: 153–157.

Dufour, J.M. (1986), "Bias of s^2 in Linear Regressions with Dependent Errors," *The American Statistician*, 40: 284–285.

Fuller, W.A. and G.E. Battese (1974), "Estimation of Linear Models with Crossed-Error Structure," *Journal of Econometrics*, 2: 67–78.

Goldberger, A.S. (1962), "Best Linear Unbiased Prediction in the Generalized Linear Regression Model," *Journal of the American Statistical Association*, 57: 369–375.

Harvey, A.C. (1976), "Estimating Regression Models With Multiplicative Heteroskedasticity," *Econometrica*, 44: 461–466.

Im, E.I. and M.S. Snow (1993), "Sampling Distributions and Efficiency Comparisons of OLS and GLS in the Presence of Both Serial Correlation and Heteroskedasticity," *Econometric Theory*, Solution 92.2.3, 9: 322–323.

Ioannides, Y.M. and J.E. Zabel (2003), "Neighbourhood Effects and Housing Demand," *Journal of Applied Econometrics* 18: 563–584.

Kadiyala, K.R. (1968), "A Transformation Used to Circumvent the Problem of Autocorrelation," *Econometrica*, 36: 93–96.

Koenker, R. and G. Bassett, Jr. (1982), "Robust Tests for Heteroskedasticity Based on Regression Quantiles," *Econometrica*, 50: 43–61.

Krämer, W. and S. Berghoff (1991), "Consistency of s^2 in the Linear Regression Model with Correlated Errors," *Empirical Economics*, 16: 375–377.

Kruskal, W. (1968), "When are Gauss-Markov and Least Squares Estimators Identical? A Coordinate-Free Approach," *The Annals of Mathematical Statistics*, 39: 70–75.

Lempers, F.B. and T. Kloek (1973), "On a Simple Transformation for Second-Order Autocorrelated Disturbances in Regression Analysis," *Statistica Neerlandica*, 27: 69–75.

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

Milliken, G.A. and M. Albohali (1984), "On Necessary and Sufficient Conditions for Ordinary Least Squares Estimators to be Best Linear Unbiased Estimators," *The American Statistician*, 38: 298–299.

Neudecker, H. (1977), "Bounds for the Bias of the Least Squares Estimator of σ^2 in Case of a First-Order Autoregressive Process (positive autocorrelation)," *Econometrica*, 45: 1257–1262.

Neudecker, H. (1978), "Bounds for the Bias of the LS Estimator in the Case of a First-Order (positive) Autoregressive Process Where the Regression Contains a Constant Term," *Econometrica*, 46: 1223–1226.

Newey, W. and K. West (1987), "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55: 703–708.

Oksanen, E.H. (1991), "A Simple Approach to Teaching Generalized Least Squares Theory," *The American Statistician*, 45: 229–233.

Ord, J.K. (1975), " Estimation Methods for Models of Spatial Interaction," *Journal of the American Statistical Association*, 70: 120–126.

Phillips, P.C.B. and M.R. Wickens (1978), *Exercises in Econometrics*, Vol. 1 (Philip Allan/Ballinger: Oxford).

Puntanen S. and G.P.H. Styan (1989), "The Equality of the Ordinary Least Squares Estimator and the Best Linear Unbiased Estimator," (with discussion), *The American Statistician*, 43: 153–161.

Sathe, S.T. and H.D. Vinod (1974), "Bounds on the Variance of Regression Coefficients Due to Heteroskedastic or Autoregressive Errors," *Econometrica*, 42: 333–340.

Schmidt, P. (1976), *Econometrics* (Marcell-Decker: New York).

Termayne, A.R. (1985), "Prediction Error Variances Under Heteroskedasticity," *Econometric Theory*, Problem 85.2.3, 1: 293–294.

Theil, H. (1971), *Principles of Econometrics* (Wiley: New York).

Thomas, J.J and K.F. Wallis (1971), "Seasonal Variation in Regression Analysis," *Journal of the Royal Statistical Society*, Series A, 134: 67–72.

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

Zyskind, G. (1967), "On Canonical Forms, Non-Negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models," *The Annals of Mathematical Statistics*, 38: 1092–1109.