BOSTON COLLEGE  
Department of Economics  
EC771: Econometrics  
Spring 2004  
Prof. Baum, Ms. Uysal

SOLUTION KEY FOR PROBLEM SET 6  

1. What is the covariance matrix, \( \text{cov}[\hat{\beta}, \hat{\beta} - b] \), of the GLS estimator \( \hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \) and the difference between it and the OLS estimator, \( b = (X'X)^{-1}X'y \)? The result plays a pivotal role in the development of specification tests in Hausman (1978).

Write the two estimators as \( \hat{\beta} = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\epsilon \) and \( b = \beta + (X'X)^{-1}X'\epsilon \). Then, \( (\hat{\beta} - b) = [(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} - (X'X)^{-1}X']\epsilon \) has \( E[\hat{\beta} - b] = 0 \) since both estimators are unbiased. Therefore, \( \text{Cov}[\hat{\beta}, \hat{\beta} - b] = E[(\hat{\beta} - \beta)(\hat{\beta} - b)'] \).

Then,

\[
E\{ (X'\Omega^{-1}X)^{-1}X'\epsilon'[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} - (X'X)^{-1}X'] \}
\]

\[
= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(\sigma^2\Omega)[\Omega^{-1}X(X'\Omega^{-1}X)^{-1} - X(X'X)^{-1}]
\]

\[
= \sigma^2(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X(X'\Omega^{-1}X)^{-1} - (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X(X'X)^{-1}
\]

once the inverse matrices are multiplied.

2. Suppose that the regression model is \( y = \mu + \epsilon \), where \( \epsilon \) has a zero mean, constant variance, and equal correlation \( \rho \) across observations. Then \( \text{cov}[\epsilon_i, \epsilon_j] = \sigma^2 \rho \) if \( i \neq j \). Prove that the least squares estimator of \( \mu \) is inconsistent. Find the characteristic roots of \( \Omega \) and show that Condition 2 after Theorem 10.2 is violated.

The covariance matrix is

\[
\sigma^2\Omega = \sigma^2 \begin{bmatrix}
  1 & \rho & \rho & \cdots & \rho \\
  \rho & 1 & \rho & \cdots & \rho \\
  \rho & \rho & 1 & \cdots & \rho \\
  \cdots & \cdots & \cdots & \cdots & \cdots \\
  \rho & \rho & \rho & \cdots & 1
\end{bmatrix}
\]

The matrix \( X \) is a column of 1s, so the least squares estimator of \( \mu \) is \( \bar{y} \). Inserting this \( \Omega \) into (10-5), we obtain \( \text{var}[\hat{y}] = \frac{\sigma^2}{n}(1 - \rho + \rho n) \). The limit of this expression is \( \rho \sigma^2 \), not zero. Although ordinary least squares is unbiased, it is not consistent. For this model, \( (X'\Omega X)/n = 1 + \rho(n-1) \), which does not converge. Using theorem 10.2 instead, \( X \) is a column of 1s, so \( (X'X) = n \), a scalar, which satisfies condition 1. To find the characteristic roots, multiply out the equation \( \Omega X = \lambda x = (1 - \rho)1x + \rho ii'x = \lambda x \). Since \( i'x = \sum x_i \), consider any
vector \( x \) whose elements sum to zero. If so, then it’s obvious that \( \lambda = \rho \). There are \( n - 1 \) such roots. Finally, suppose that \( x = i \). Plugging this into the equation produces \( \lambda = 1 - \rho + np \). The characteristic roots of \( \Omega \) are \( (1 - \rho) \) with multiplicity \( n - 1 \) and \( (1 - \rho + np) \), which violates condition 2.

3. Suppose that the regression model is \( y_i = \mu + \epsilon_i \), where \( E[\epsilon_i|x_i]=0 \), but \( \text{var}[\epsilon_i|x_i]=\sigma^2 x_i^2 \), \( x_i > 0 \).
(a) Given a sample of observations on \( y_i \) and \( x - i \), what is the most efficient estimator of \( \epsilon \)? What is its variance?
(b) What is the ordinary least squares estimator of \( \mu \) and what is the variance of the ordinary least squares estimator?
(c) Prove that the estimator in (a) is at least as efficient as the estimator in (b).

This is a heteroskedastic regression model in which the matrix \( X \) is a column of ones. The efficient estimator is the GLS estimator, \( \hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = \left[ \sum \frac{y_i x_i^2}{(x_i^2)} \right] / \left[ \sum (1/x_i^2) \right] \). As always, the variance of the estimator is \( \text{var}[\hat{\beta}] = \sigma^2 (X'\Omega^{-1}X)^{-1} \). The OLS estimator is \( (X'X)^{-1}X'y = \tilde{y} \). The variance of \( \tilde{y} \) is \( \sigma^2/n \). To show that the variance of the OLS estimator is greater than or equal to that of the GLS estimator, we must show that \( \sigma^2/n \sum x_i^2 \geq \sigma^2/(\sum (1/x_i^2)) \) or \( (1/n^2)(\sum x_i^2)/(\sum (1/x_i^2)) \geq 1 \) or \( \sum_j x_j^2/x_j^2 \geq n^2 \). The double sum contains \( n \) terms equal to one. There remain \( n(n-1)/2 \) pairs of the form \( (x_i^2/x_j^2 + x_j^2/x_i^2) \). If it can be shown that each of these sums is greater than or equal to 2, the result is proved. Just let \( z_i = x_i^2 \). Then, we require \( z_i/z_j + z_j/z_i - 2 \geq 0 \). But this is equivalent to \( (z_i^2 + z_j^2 - 2z_i z_j)/(z_i z_j) \geq 0 \) or \( (z_i - z_j)^2/(z_i z_j) \geq 0 \), which is certainly true if \( z_i \) and \( z_j \) are positive. They are since \( z_i \) equals \( x_i^2 \). This completes the proof.

5. Does first differencing reduce autocorrelation? Consider the models \( y_i = \beta'z_i + \epsilon_i \), where \( \epsilon_i = \rho \epsilon_{i-1} + u_i \) and \( \epsilon_i = u_i - \lambda u_{i-1} \). Compare the autocorrelation of \( \epsilon_i \) in the original model to that of \( \nu_t \) in \( y_i - y_{i-1} = \beta'(x_i - x_{i-1}) + \nu_t \) where \( \nu_t = \epsilon_t - \epsilon_{t-1} \).

For the first order autoregressive model, the autocorrelation is \( \rho \). Consider the first difference, \( \nu_t = \epsilon_t - \epsilon_{t-1} \) which has \( \text{var}[\nu_t] = \text{var}[\epsilon_t] - 2 \text{cov}([\epsilon_t, \epsilon_{t-1}]) = 2 \sigma^2(1/(1 - \rho^2) - \rho/(1 - \rho^2)) \). The variance of the differenced process is \( \text{var}[
u_t] = \sigma^2(1/(1 - \rho^2) - (2\rho - 1 - \rho^2)/(1 + \rho)) \). Therefore, first differencing reduces the absolute value of the autocorrelation coefficient when \( \rho \) is greater than 1/3. For economic data, this is likely to be fairly common.

For the moving average process, the first order autocorrelation is \( \text{cov}([\epsilon_t, \epsilon_{t-1}])/\text{var}([\epsilon_t]) = -\lambda/(1 + \lambda^2) \). To obtain the autocorrelation of the first difference, write \( \epsilon_t - \epsilon_{t-1} = u_t - (1 + \lambda)u_{t-1} + \lambda u_{t-2} \). The variance of the difference is \( \text{var}[\epsilon_t - \epsilon_{t-1}] = \sigma^2(1 + \lambda)^2 + (1 + \lambda^2) \). The covariance can be found by taking the expected product of terms with equal
subscripts. Thus, $\text{cov}[\epsilon_t - \epsilon_{t-1}, \epsilon_{t-1} - \epsilon_{t-2}] = -\sigma^2 (1 + \lambda)^2$. The autocorrelation is $\text{cov}[\epsilon_t - \epsilon_{t-1}, \epsilon_{t-1} - \epsilon_{t-2}] / \text{var}[\epsilon_t - \epsilon_{t-1}] = -(1 + \lambda)^2 / [(1 + \lambda)^2 + (1 + \lambda^2)]$. For most of the range of the autocorrelation of the original series, differences increases autocorrelation. But, for most of the range of values that are economically meaningful, differencing reduces autocorrelation.