Econometrics 616—Spring 2004
Exam II—Answer Key

1. a) Heteroskedasticity is the problem of a non-spherical error variance-covariance matrix. In layman's terms it suggests that when we analyze our data and try to use statistical inference procedures, the manner in which the random shocks affect our endogenous variables are not equivalent. By ignoring this fact a researcher can be mislead and make erroneous conclusions drawn from hypothesis testing due to the fact that the variances of the estimators are incorrect as well as any covariance that exists between parameters.

   b) $Y = \beta X + u \quad u \sim \left(0, \sigma^2 X^2\right)$, the variance of the OLS estimator gives us the following: $V\left(\hat{\beta}\right) = E\left(\left(\sum Xu\right)^2 / \left(\sum X^2\right)^2\right) = \left(\sum X^2 E\left(u^2\right)\right) / \left(\sum X^2\right)^2$ which equals $V\left(\hat{\beta}\right) = \left(\sum X^4 \sigma^2\right) / \left(\sum X^2\right)^2 = \sigma^2 \left(\sum X^4 / \left(\sum X^2\right)^2\right)$ while the variance of the GLS estimator equals:
   $V\left(\tilde{\beta}\right) = E\left(\left(\sum u/X\right)/n\right)^2 = n^{-2}\left(\sum E\left(u^2/X^2\right)\right) = n^{-2}\left(\sum E\left(u^2\right)/X^2\right) = \sigma^2/n$

   To show that the GLS estimator is more efficient than the OLS estimator we will use the Cauchy-Schwartz inequality. Let $w = X^2, \quad z = 1$ then the inequality states that $\sum w^2 \sum z^2 \geq \left(\sum wz\right)^2 \Rightarrow \sum \left(X^2\right)^2 \sum 1^2 \geq \left(\sum X^2 \cdot 1\right)^2$ and this implies that $n \cdot \sum X^4 \geq \left(\sum X^2\right)^2$ which allows us to show that GLS is more efficient than OLS.

2. $Y = \beta X + u \quad u \sim N\left(0, \sigma^2\right)$ First we will define the OLS estimate and its variance before proceeding to the LR, LM, and W test statistics.
   $\hat{\beta} = \left(\sum XY\right)/\left(\sum X^2\right) = 25/20 = 1.25, \quad V\left(\hat{\beta}\right) = \sigma^2 / \sum X^2 = \sigma^2/20$
   $e = Y - \hat{\beta}X, \quad e'e = \sum\left(Y - \hat{\beta}X\right)^2 = \sum Y^2 - 2\hat{\beta}\sum XY + \hat{\beta}^2 \sum X^2 = 40 - 2.5 \cdot 25 + 1.5625 \cdot 20 = 8.75$

   $e_* = Y - X, \quad e_*' e_* = \sum (Y - X)^2 = \sum Y^2 - 2\sum XY + \sum X^2 = 40 - 2 \cdot 25 + 20 = 10$

   Using the formulas given on the exam we compute the test statistics as follows:

   $LR = 50 \cdot (\ln 10 - \ln 8.75) = 50 \cdot (2.30259 - 2.16905) = 50 \cdot 0.133531 = 6.67657$
   $LM = 50(10 - 8.75)/10 = 5 \cdot 1.25 = 6.25$
   $W = 50 \cdot (10 - 8.75)/8.75 = 5.71429 \cdot 1.25 = 7.14286$
   And we can check to see that $W \geq LR \geq LM$. To show the next part we simply computer the t-statistic for $H_0 : \beta = 1$ which is $t = \left(\hat{\beta} - 1\right)/s.e.\left(\hat{\beta}\right)$. To find the standard error all we need is the estimate of the variance.

$V\left(\hat{\hat{\beta}}\right)=\hat{\sigma}^2/20$, $\hat{\sigma}^2 = e'e/n$. This suggests that $\hat{\sigma}^2 = 8.75/50 = 0.175$ and so our standard error becomes $\sqrt{0.175/20} = .093541$ which makes our t-statistic $t = .25/.093541 = 2.67261$ and we see that the square of this number is equivalent to our Wald statistic proving our desired result.

3. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \quad u \sim iid\left(0, \sigma_u^2\right) \quad (1)$
   $Y = \beta_0 + \beta_1 X_1 + v \quad v \sim iid\left(0, \sigma_v^2\right) \quad (2)$

   a) If we erroneously assume that model 2 is the correct model then we can find that the bias of our OLS estimator is:
      $E\left(\breve{\beta}_1\right) = E\left(\sum\left(X_1 - \bar{X}_1\right)\left(Y - \bar{Y}\right)\right) / \left(\sum\left(X_1 - \bar{X}_1\right)^2\right)$
      $= \beta_1 + \beta_2\left(\left(\sum\left(X_1 - \bar{X}_1\right)X_2\right) / \left(\sum\left(X_1 - \bar{X}_1\right)^2\right)\right)$ and we note that the OLS estimator is biased upwards is the correlation between the included and the excluded variable has the same sign as the excluded variable's associated parameter.

   b) Show that $V\left(\hat{\beta}\right) = \sigma_u^2 / \left(\left(1 - r^2\right)\left(\sum\left(X_1 - \bar{X}_1\right)^2\right)\right)$. Using the hint, we find that $V\left(\hat{\beta}\right) = \sigma_u^2 / X_1' M_{-1} X_1$ and we note that the scalar in the denominator is the residual sum of squares from a regression of X_1 on a constant and X_2. This means that
      $V\left(\hat{\beta}\right) = \sigma_u^2 / X_1' M_{-1} X_1 = \sigma_u^2 / \left(\sum\left(X_1 - \bar{X}_1\right)^2 - \left[\sum\left(X_1 - \bar{X}_1\right)\left(X_2 - \bar{X}_2\right)\right]^2 / \sum\left(X_2 - \bar{X}_2\right)^2\right)$
      which becomes the desired result upon removing $\sum\left(X_1 - \bar{X}_1\right)^2$ from the denominator.

   c) In the over specified model
      $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + w \quad w \sim iid\left(0, \sigma_w^2\right) \quad (3)$
      Show that $V\left(\breve{\beta}\right) = \sigma_w^2 / \left(\left(1 - R_1^2\right)\left(\sum\left(X_1 - \bar{X}_1\right)^2\right)\right)$ where R_1^2 is the coefficient of determination from the regression of X_1 on a constant, X_2, and X_3. Using the hint again we have $V\left(\breve{\beta}\right) = \sigma_w^2 / X_1' M_{-1} X_1$ and the denominator of this fraction is just the residual sum of squares of the regression of X_1 on a constant, X_2, and X_3. Multiplying and dividing the denominator by the total sum of squares yields the following:
      $V\left(\breve{\beta}\right) = \sigma_w^2 / \left(TSS \cdot \left(RSS/TSS\right)\right) = \sigma_w^2 / \left(\sum\left(X_1 - \bar{X}_1\right)^2 \cdot \left(1 - R_1^2\right)\right)$.

   d) To compare the variances between the over specified model and the under specified we must note that each model does not have identical errors and so does not necessarily have identical error variances. We would believe that the over specified model would have a lower error variance than that from the under specified model. However because the over specified model has more variables it must also have a smaller denominator due to the fact that $R_1^2 \geq r^2$. Since both the denominator and the numerator are smaller for the over specified model there is no sure way to say which parameter variance is larger. However an over specified is unbiased while and under specified model is not which leads to biased parameter estimates when the model being estimated is not a special case of the true Data Generating Process as is the case with a model that is under specified.

4. $Y_t = \alpha + u_t, \quad u_t = \rho u_t + \varepsilon_t \quad \varepsilon_t \sim iid\left(0, \sigma_\varepsilon^2\right), u_t \sim \left(0, \sigma_u^2 \Omega\right)$

a) For the AR(1) model above either the Cochrane-Orcutt procedure or the Prais-Winsten procedure is applicable. The only difference between the two is whether or not the first observation is included in all of the auxiliary regressions or not. For both procedures you estimate the model as if there was no AR(1) process governing the movement of the random error term. This would yield a preliminary estimate of the intercept from which thee vector of residuals could be found. The next step would be to use these residuals in the AR(1) model to estimate rho. Once rho is found then the data can be transformed with C-O making sure to drop the first observation or with P-W where the first observation is different from C-O and the rest are equivalent. This procedure can be iterated until some desired tolerance level is achieved between successive rhos. To obtain a final estimate of the intercept we have the following $\hat{\alpha} = \alpha\left(1 \hat{-} \rho\right) / \left(1 - \hat{\rho}\right)$.

b) If instead of the standard AR(1) process we introduce heteroskedasticity into the innovations of the AR(1) then we must use a slight variant for the procedure in part a. We first develop a preliminary guess to the AR(1) movement by choosing a rho. We next transform the model as we would in the C-O procedure. Once we have done this only the AR(1) innovation will be left and we can then correct for heteroskedasticity in our model. We then find the residuals from this regression and run them in the AR(1) regression to find an estimate of rho. This is possible because the residuals have already been corrected for heteroskedasticity in the first step. Once we find our new estimate of rho we again use the C-O transformation of the data and correct for heteroskedasticity before running the regression. This procedure can be run as many times as one wishes until some desired tolerance level is achieved. Due to the fact that there will be an x-variable associated with the intercept we can find it exactly the same way we did in part a.

c) $Y_t = \alpha + u_t \quad u_t \sim iid\left(0, \sigma_u^2\right)$ and so $\hat{\alpha} = \overline{Y}$. Some manipulation yields us the following:

$$Y_t - \alpha = u_t \sim iid\left(0, \sigma_u^2\right) \Rightarrow \sqrt{n}\left(\overline{Y} - \alpha\right) = \sqrt{n}\left(\hat{\alpha} - \alpha\right) = \sqrt{n}\overline{u} \overset{a}{\sim} N\left(0, \sigma_u^2\right)$$

d) If we had an MA(1), $u_t = \varepsilon_t + \phi\varepsilon_{t-1} \quad \varepsilon_t \sim iid\left(0, \sigma_\varepsilon^2\right)$, instead of an AR(1) we would estimate the model ignoring any type of correlation between the variables. We would obtain an estimate of the intercept and use it to find the residuals of the model. Once these were found we compute the variance and the covariance of them. These estimates allow us to find the coefficient of the MA(1) model as follows:
$V(\hat{u}) = \left(1 + \phi^2\right)\sigma_\varepsilon^2, \quad Cov\left(\hat{u}_t, \hat{u}_{t-1}\right) = \phi\sigma_\varepsilon^2$. This gives us a quadratic formula in the variance and the covariance of the residuals that will allow us to find an estimate of the innovation coefficient. We see that $Cov\left(\hat{u}_t, \hat{u}_{t-1}\right)/\phi = \sigma_\varepsilon^2 \quad V(\hat{u}) = \left(1 + \phi^2\right)Cov\left(\hat{u}_t, \hat{u}_{t-1}\right)/\phi$ and we can simplify this as $\phi^2 Cov\left(\hat{u}_t, \hat{u}_{t-1}\right) - \phi V(\hat{u}) + Cov\left(\hat{u}_t, \hat{u}_{t-1}\right) = 0$ and use the formula for a quadratic equation to solve for phi as $\hat{\phi} = \left(-b \pm \sqrt{b^2 - 4ac}\right)/2a$.