Queen's University
School of Graduate Studies and Research
Department of Economics

Economics 850 **Econometrics I** **Fall, 2018**

Professor James MacKinnon

Answers to Final Examination

December 13, 2018. Time Limit: 3 hours

Please answer any **four (4)** of the following **six (6)** questions. Each question has **four parts** and is worth 25% of the final mark. Before deciding which questions to answer, it would be a very good idea to read every question carefully.

Note: A table with some critical values of the χ^2 distribution appears at the end of the examination.

1. Consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}, \tag{1}$$

where there are n observations and k exogenous regressors, with k_1 regressors in the matrix $\boldsymbol{X}_1$ and k_2 regressors in the matrix $\boldsymbol{X}_2$. Assume initially that the elements of $\boldsymbol{u}$ are uncorrelated and have the same variance.

a) Write down the F statistic for the hypothesis that $\boldsymbol{\beta}_2 = \boldsymbol{0}$ as a function of $\boldsymbol{y}$, $\boldsymbol{X}_1$, and $\boldsymbol{X}_2$, using projection matrices to simplify the algebra. How is this statistic distributed asymptotically under the null hypothesis? Would it follow a known distribution in finite samples? Would it follow a known distribution in finite samples if you made additional assumptions? Explain briefly.

b) Write down the Wald statistic for the hypothesis that $\boldsymbol{\beta}_2 = \boldsymbol{0}$ as a function of $\boldsymbol{y}$, $\boldsymbol{X}_1$, and $\boldsymbol{X}_2$. How is this statistic distributed asymptotically under the null hypothesis? Explain how it is related to the F statistic you derived in part a).

c) Now suppose that the elements of $\boldsymbol{u}$ are independent but not identically distributed, with $\mathrm{E}(u_i^2) = \sigma_i^2$ for observation i, where the σ_i^2 are unknown. If $\boldsymbol{\beta}_2^0$ denotes the true value of $\boldsymbol{\beta}_2$, how is the vector $n^{1/2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0)$ distributed asymptotically?

d) Using the results of part c), write down the Wald statistic for the hypothesis that $\boldsymbol{\beta}_2 = \boldsymbol{0}$. Explain how you would perform a bootstrap test of this hypothesis based on your Wald statistic, a suitable bootstrap generating process, and the appropriate method for computing a bootstrap P value.

ANSWER [6 marks for parts a) though c), 7 marks for part d)]

a) By the FWL Theorem, (1) is equivalent to

$$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{M}_1\boldsymbol{u}.$$

The RSSR here is just $\boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{y}$, and the USSR is $\boldsymbol{y}^\top\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{y} = \boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{y} - \boldsymbol{y}^\top\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}\boldsymbol{y}$. Under the null, $\boldsymbol{M}_1$ annihilates $\boldsymbol{X}_1\boldsymbol{\beta}_1$, so that we can replace $\boldsymbol{y}$ by $\boldsymbol{u}$. Thus, under the null,

$$F = \frac{\boldsymbol{u}^\top\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}\boldsymbol{u}/k_2}{\boldsymbol{u}^\top\boldsymbol{M}_{\boldsymbol{X}}\boldsymbol{u}/(n-k)} = \frac{1}{s^2}\boldsymbol{u}^\top\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}\boldsymbol{u}/k_2.$$

Since the denominator tends to σ^2 as $n \to \infty$, and the numerator is $1/k_2$ times a quadratic form involving a projection matrix of rank k_2, k_2 times this statistic must be asymptotically distributed as $\chi^2(k_2)$.

The F statistic would not follow a known distribution in finite samples without the additional assumption that the vector $\boldsymbol{u}$ is $\mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$. Under that assumption, it would be distributed as $F(k_2, n-k)$.

b) The Wald statistic is

$$W = \hat{\boldsymbol{\beta}}_2^\top\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2)^{-1}\hat{\boldsymbol{\beta}}_2.$$

Using the facts that

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y}$$

and

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2) = s^2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1},$$

we find that

$$W = \frac{1}{s^2}\boldsymbol{y}^\top\boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{y} = \frac{1}{s^2}\boldsymbol{y}^\top\boldsymbol{P}_{\boldsymbol{M}_1\boldsymbol{X}_2}\boldsymbol{y},$$

which is precisely k_2 times the F statistic. By the arguments already given for F, it must be asymptotically distributed as $\chi^2(k_2)$.

c) When the elements of $\boldsymbol{u}$ are independent (a weaker assumption with the same implications would have been that they are uncorrelated), the covariance matrix of $\boldsymbol{u}$ can be written as $\boldsymbol{\Omega}$, a diagonal matrix with σ_i^2 as a typical diagonal element. Then the covariance matrix of $\hat{\boldsymbol{\beta}}_2$ is

$$(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2\boldsymbol{M}_1\boldsymbol{\Omega}\boldsymbol{M}_1\boldsymbol{X}_2(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}.$$

This implies that

$$n^{1/2}(\hat{\boldsymbol{\beta}}_2 - \boldsymbol{\beta}_2^0) \overset{a}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{V}_2),$$

where

$$\boldsymbol{V}_2 \equiv \Big(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2\Big)^{-1} \Big(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{\Omega} \boldsymbol{M}_1 \boldsymbol{X}_2\Big) \Big(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2\Big)^{-1}.$$

d) For the Wald statistic, we need to replace $\boldsymbol{\Omega}$ by $\hat{\boldsymbol{\Omega}}$, where $\hat{\boldsymbol{\Omega}}$ is a diagonal matrix with squared residuals on the diagonal, to obtain the covariance matrix estimate

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2) = (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \hat{\boldsymbol{\Omega}} \boldsymbol{M}_1 \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1}.$$

The Wald statistic can still be written as

$$W = \hat{\boldsymbol{\beta}}_2^\top \widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2)^{-1} \hat{\boldsymbol{\beta}}_2.$$

with this new definition of $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_2)$.

For the bootstrap test, we need to use the wild bootstrap. The procedure is the usual one:

Actual test statistic is W.

Bootstrap DGP is $y_i^* = \boldsymbol{x}_{1i}\tilde{\boldsymbol{\beta}}_1 + \tilde{u}_i v_i^*$. This requires estimating the restricted model to obtain $\tilde{\boldsymbol{\beta}}_1$ and $\tilde{\boldsymbol{u}}$. v_i^* is a realization of the Rademacher distribution.

For each bootstrap sample, estimate the model without restrictions and compute a bootstrap Wald statistic W_b^*.

Compute upper-tail bootstrap P value:

$$P^* = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}(W_b^* > W).$$

Reject if $P^* < \alpha$. Somewhere, they need to say that $\alpha(B+1)$ is an integer. And they must *not* use an equal-tail P value.

2. Consider the linear regression model

$$y_i = \beta_1 + \beta_2 t_i + u_i, \quad i = 1, \ldots, n, \tag{2}$$

where t_i is a dummy variable that equals 1 for treated observations and 0 for untreated ones. Assume that n is fairly large, say, $n \geq 2000$.

a) Explain how, without using bootstrap methods, you could test the hypothesis that $\beta_2 = 0$ if the u_i were assumed to be uncorrelated but possibly heteroskedastic, with unknown variances σ_i^2.

b) Now suppose that the n observations fall into G groups, where $\mathrm{E}(u_i u_j) \neq 0$ whenever i and j belong to the same group, but $\mathrm{E}(u_i u_j) = 0$ whenever i and j belong to different groups. Suppose further that treatment occurs at the group level, so that, for all observations within each group, either $t_i = 0$ or $t_i = 1$. Explain how, without using bootstrap methods, you could test the hypothesis that $\beta_2 = 0$ under these assumptions.

c) The procedure you proposed in part b) should yield a test statistic that follows the standard normal distribution when G is infinite. Suppose the value of this statistic is 4.37. What would you conclude if there were 17 groups, but only one of them was treated? What would you conclude if there were 17 groups and 6 of them were treated? Explain briefly.

d) Suppose you averaged the data within each group and ran the regression

$$\bar{y}_g = \beta_1 + \beta_2 t_g + u_g, \quad g = 1, \ldots, G, \tag{3}$$

where $\bar{y}_g$ is the average value of y_i for group g, and t_g is a dummy variable that equals 1 for treated groups and 0 for untreated ones. You would like to test the hypothesis $\beta_2 = 0$ using the *ordinary* t statistic for $\beta_2 = 0$ in equation (3). Are there any assumptions you can make about G, the u_i in equation (2), and the numbers of observations in each group under which this would be a sensible thing to do? Explain briefly.

ANSWER [4 marks for part a), 5 marks for part b), 8 marks for parts c) and d)]

a) Simply run a regression of y_i on t_i, using an HCCME. The hetero-robust t statistic on $\hat{\beta}_2$ provides an asymptotically valid test statistic. It should be compared to the $\mathrm{N}(0,1)$ distribution using a two-tailed test, or maybe $t(n-2)$, but since n is large this does not matter.

b) Run the same regression, but this time use a CRVE. In general, it is the matrix

$$C(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \left(\sum_{g=1}^{G} \boldsymbol{X}_g^\top \hat{\boldsymbol{u}}_g \hat{\boldsymbol{u}}_g^\top \boldsymbol{X}_g \right) (\boldsymbol{X}^\top \boldsymbol{X})^{-1},$$

where C is a degrees-of-freedom factor. The cluster-robust test statistic is just $\hat{\beta}_2$ divided by the square root of the second diagonal element of this matrix. It should be compared to the $t(G-1)$ distribution using a two-tailed test.

c) With one treated group ($G_1 = 1$) out of $G = 17$ groups, the CRVE is totally unreliable. The t statistic of part b) cannot be trusted at all, because the standard error of $\hat{\beta}_2$ is probably much too small.

If $G_1 = 6$ groups were treated, the t statistic still cannot be trusted, but the standard error is probably not underestimated nearly as severely as when $G_1 = 1$. I would be willing to reject at the .05 level, and probably even at the

.01 level, in this case. For comparison, the .01 critical value of $t(16)$ is 2.92, so there is a quite a gap between it and 4.37.

d) This regression would have G observations, or 17 in the case of part c). Thus we would use the $t(G-2)$ distribution. The test statistic would evidently have exactly this distribution if the original disturbances were normally, independently, and identically distributed, and all clusters were the same size. This would be true even if G were very small. Even without normality, the approximation should be very good if n/G is large, because we are averaging over observations within each cluster.

When we go from (2) to (3), we have

$$\bar{y}_g = \frac{1}{N_g} \sum_{j=1}^{N_g} y_{gj},$$

where the y_{gj} denote the y_i that belong to group g. Therefore, the u_g must be averages of the corresponding u_{gj}. If the latter have constant variance, the u_g must be extremely close to being normally and identically distributed. This will still be true if they have nonconstant variances, provided those variances do not vary systematically across groups.

But the approximation will fail if the N_g vary at all, because that would introduce heteroskedasticity at the cluster level. Minor variations across clusters probably won't matter much, but substantial ones certainly will.

We could replace the ordinary t statistic by a hetero-robust one, but this will fail drastically when $G_1 = 1$ and will probably not work very well when $G_1 = 6$. It would surely be better to use GLS, assuming that the variance of the $\bar{y}_g$ are inversely proportional to N_g.

3. You are interested in the coefficient on $\boldsymbol{y}_2$ in the linear equation

$$\boldsymbol{y}_1 = \beta \boldsymbol{y}_2 + \boldsymbol{Z\gamma} + \boldsymbol{u}, \tag{4}$$

where there are 1244 observations. As the notation implies, $\boldsymbol{y}_2$ is endogenous. The matrix $\boldsymbol{Z}$ is 1244×5 and contains observations on exogenous variables, including a vector of ones. The vector $\boldsymbol{y}_2$ is determined by the equation

$$\boldsymbol{y}_2 = \boldsymbol{W\pi} + \boldsymbol{v} = \boldsymbol{W}_1\boldsymbol{\pi}_1 + \boldsymbol{Z\pi}_2 + \boldsymbol{v}, \tag{5}$$

where $\boldsymbol{W}$ is a 1244×9 matrix of instruments, some of which also belong to $\boldsymbol{Z}$. In both equations (4) and (5), the disturbances are assumed to be IID. The correlation between a typical element of $\boldsymbol{u}$ and the corresponding element of $\boldsymbol{v}$ is ρ.

a) Will the OLS estimate $\hat{\beta}$ from (4) be unbiased? Will it be consistent? Does your answer depend on the value of ρ? Explain briefly.

b) How would you obtain a consistent estimate of β? Would this estimate, say $\check{\beta}$, be unbiased? Explain briefly.

c) For the estimation method you used in part b), the regression package reports that the SSR is 827.76 and that

$$(\boldsymbol{y} - \check{\beta}\boldsymbol{y}_2 + \boldsymbol{Z}\check{\boldsymbol{\gamma}})^\top \boldsymbol{P_W}(\boldsymbol{y} - \check{\beta}\boldsymbol{y}_2 + \boldsymbol{Z}\check{\boldsymbol{\gamma}}) = 7.34, \tag{6}$$

where $\boldsymbol{P_W}$ projects orthogonally onto the space spanned by the columns of $\boldsymbol{W}$. How many overidentifying restrictions are there? Calculate a test statistic for the hypothesis that these overidentifying restrictions are valid. Based on the asymptotic distribution of this test statistic, would you reject these restrictions at the .01 level?

d) Suppose that the R^2 from regression (5) is 0.3254 and the R^2 from regressing $\boldsymbol{y}_2$ on $\boldsymbol{Z}$ is 0.3142. Based on these numbers, would you expect $\check{\beta}$ to be well approximated by its asymptotic distribution? Explain briefly.

ANSWER [6 marks for parts a), b), and c); 7 marks for part d)]

a) In general, $\hat{\beta}$ will be biased and inconsistent. However, in the special case in which $\rho = 0$, it will actually be unbiased (since all regressors and instruments are assumed to be exogenous).

b) We can obtain a consistent (but biased) estimate of β by using generalized IV (2SLS). If we replace $\boldsymbol{y}_2$ by its projection onto $\boldsymbol{W}$, equation (4) becomes

$$\boldsymbol{y}_1 = \beta\boldsymbol{P_W}\boldsymbol{y}_2 + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{u}.$$

Thus the estimate of β is the one from the FWL regression

$$\boldsymbol{M_Z}\boldsymbol{y}_1 = \beta\boldsymbol{M_Z}\boldsymbol{P_W}\boldsymbol{y}_2 + \text{resids},$$

which is

$$\check{\beta} = (\boldsymbol{y}_2^\top\boldsymbol{P_W}\boldsymbol{M_Z}\boldsymbol{P_W}\boldsymbol{y}_2)^{-1}\boldsymbol{y}_2^\top\boldsymbol{P_W}\boldsymbol{M_Z}\boldsymbol{y}_1.$$

This is the IV estimator. Unless $\rho = 0$, it is biased, but it is consistent.

c) In this case, since there are 9 instruments and 6 coefficients in the structural equation, there must be 3 overidentifying restrictions. The test statistic is

$$7.34/\hat{\sigma}^2 = 7.34/0.66863 = 10.978,$$

using $\hat{\sigma}^2 = 827.76/1238 = 0.66863$. We could also have used $827.76/1244 = 0.6654$, yielding a test statistic of 11.031. Under the null hypothesis, this test statistic is asymptotically distributed as $\chi^2(3)$. From the table, the critical values at the .05, .025, and .01 levels are 7.815, 9.348, and 11.345. So we cannot reject at the .01 level.

d) The two values of R^2 seem very close, but, with 1244 observations, the F statistic for $\boldsymbol{\pi}_1 = \mathbf{0}$ is

$$F = \frac{(0.3254 - 0.3142)/4}{(1 - 0.3254)/1235} = 5.126.$$

Here 1235 is the denominator of s^2 for the unrestricted model. In a conventional setting, this would be quite significant. However, it falls well short of the levels suggested in Staiger and Stock and Stock and Yogo. This suggests that $\breve{\beta}$ may not be well approximated by its asymptotic distribution.

4. Consider the linear regression model

$$y_t = \beta_1 + \beta_2 x_{1t} + \beta_3 x_{2t} + u_t, \tag{7}$$

which is to be estimated using a sample of 83 observations. The regressors x_{1t} and x_{2t} are assumed to be exogenous, and the disturbances u_t are assumed to be independent but possibly heteroskedastic. You are interested in the parameter $\gamma \equiv \beta_2/\beta_3$.

a) Explain how you would obtain a consistent estimate $\hat{\gamma}$. Would $\hat{\gamma}$ be biased or unbiased? Explain briefly.

b) Explain how you would obtain both a standard error $s(\hat{\gamma})$ based on asymptotic theory and a bootstrap standard error $s^*(\hat{\gamma})$.

c) Explain how you would perform a bootstrap test of the hypothesis that $\gamma = 2$ at the .01 level.

d) Briefly discuss two ways to obtain a 95% bootstrap confidence interval for γ. Only one of them should make use of $s^*(\hat{\gamma})$.

ANSWER [4 marks for part a), 7 marks for parts b), c) and d)]

a) $\hat{\gamma} = \hat{\beta}_2/\hat{\beta}_3$ is evidently consistent, since both $\hat{\beta}_2$ and $\hat{\beta}_3$ are consistent. We need to assume that $\beta_3 \neq 0$. However, $\hat{\gamma}$ will in general be biased, because it is a nonlinear function of two unbiased estimators.

b) For the asymptotic standard error, we need to use the delta method. OLS estimation of (7) yields a hetero-robust covariance matrix estimator with typical element $\hat{\sigma}_{jk}$. The elements that we need are $\hat{\sigma}_{22}$, $\hat{\sigma}_{23}$, and $\hat{\sigma}_{33}$. The derivatives of γ with respect to β_2 and β_3 are $1/\beta_3$ and $-\beta_2/\beta_3^2$, respectively. Thus we estimate the variance of $\hat{\gamma}$ by

$$[1/\hat{\beta}_3 \quad -\hat{\beta}_2/\hat{\beta}_3^2] \begin{bmatrix} \hat{\sigma}_{22} & \hat{\sigma}_{23} \\ \hat{\sigma}_{23} & \hat{\sigma}_{33} \end{bmatrix} \begin{bmatrix} 1/\hat{\beta}_3 \\ -\hat{\beta}_2/\hat{\beta}_3^2 \end{bmatrix}.$$

The asymptotic standard error $s(\hat{\gamma})$ is the square root of this variance.

Another approach would be to run the GNR for the nonlinear model

$$y_t = \beta_1 + \gamma\beta_3 x_{1t} + \beta_3 x_{2t} + u_t,$$

plugging in $\hat{\beta}$, $\hat{\beta}_3$, and $\hat{\gamma}$. Then use the hetero-robust standard error on the coefficient of the regressor that corresponds to γ as the standard error of $\hat{\gamma}$.

For the bootstrap standard error, we use the unrestricted wild bootstrap to generate B samples. For each of them, we calculate $\hat{\gamma}_b^*$. Then the bootstrap standard error $s^*(\hat{\gamma})$ is the square root of

$$\frac{1}{B}\sum_{b=1}^{B}(\hat{\gamma}_b^* - \bar{\gamma}^*)^2,$$

where $\bar{\gamma}^*$ is the arithmetic mean of the $\hat{\gamma}_b^*$.

c) The *wrong* answer is to divide $\hat{\gamma} - 2$ by $s^*(\hat{\gamma})$ and perform a t test using the $t(80)$ distribution. The right answer is to use the restricted wild bootstrap:

Calculate the test statistic

$$t(\gamma = 2) = \frac{\hat{\gamma} - 2}{s(\hat{\gamma})},$$

where $s(\hat{\gamma})$ is the delta-method standard error from part b).

Estimate the model under the restriction that $\gamma = 2$. This may be done by running the regression

$$y_t = \beta_1 + \beta_3(2x_{1t} + x_{2t}) + u_t.$$

Doing this yields estimates $\tilde{\beta}_1$, $\tilde{\beta}_3$, and $\tilde{\beta}_2 = 2\tilde{\beta}_3$, plus residuals $\tilde{u}_t$.

Generate B bootstrap samples using the wild bootstrap DGP

$$y_t^{*b} = \tilde{\beta}_1 + \tilde{\beta}_2 x_{1t} + \tilde{\beta}_3 x_{2t} + v_t^{*b}\tilde{u}_t.$$

For each bootstrap sample, calculate $\hat{\gamma}_b^*$ and $s(\hat{\gamma}_b^*)$, the delta-method standard error. Use these to calculate the bootstrap analog of $t(\gamma = 2)$:

$$t_b^*(\gamma = 2) = \frac{\hat{\gamma}_b^* - 2}{s(\hat{\gamma}_b^*)}.$$

Then calculate an equal-tailed bootstrap P value in the usual way. Reject if it is less than 0.01. Because $\hat{\gamma}$ is almost certainly biased, it would be wrong to use a symmetric P value.

Another approach would be to run the (ordinary) GNR to get a hetero-robust t statistic, and run it again B times for samples generated by the same wild bootstrap DGP as above to obtain B bootstrap analogs of this test statistic.

d) The easiest way to construct a bootstrap confidence interval is to use the bootstrap standard error $s^*(\hat{\gamma})$:

$$\left[\hat{\gamma} - C_{1-\alpha/2}s^*(\hat{\gamma}),\ \ \hat{\gamma} + C_{1-\alpha/2}s^*(\hat{\gamma})\right],$$

where $C_{1-\alpha/2}$ is the $1-\alpha/2$ (in this case, the .975) quantile of the $t(80)$ distribution.

A better approach (in many cases) is to use a studentized bootstrap interval. It differs from the one above in two ways. First, it uses $s(\hat{\gamma})$ instead of $s^*(\hat{\gamma})$. Second, instead of $C_{1-\alpha/2}$, it uses two different critical values. These are the .975 and .025 quantiles of the $t^*_b(\gamma = \hat{\gamma})$, which come from B bootstrap samples based on unrestricted estimates. The lower limit of the interval uses the .975 quantile, and the upper tail uses the .025 quantile.

5. Suppose you set out to gather samples from two different populations, with the samples having n_1 and n_2 observations, respectively. Observations from the first population have mean μ_1, and observations from the second population have mean μ_2. Your objective is to test the hypothesis that $\mu_1 = \mu_2$.

a) Explain how you could test the hypothesis that $\mu_1 = \mu_2$ by running a single linear regression. Assume that all observations have the same variance.

b) If you can only afford to gather 250 observations, how would you choose n_1 and n_2 to maximize the power of the test under the assumptions of part a)?

c) Suppose you notice that observations in the first sample are less variable than observations in the second sample. You believe it is reasonable to assume that $\sigma_1/\sigma_2 = \gamma$, where γ is an unknown parameter. Explain how you could obtain more efficient estimates of μ_1 and μ_2 than the ones you obtained in part a).

d) Explain how you could test the hypothesis that $\mu_1 = \mu_2$ using the estimates from part c). The values of n_1 and n_2 chosen in part b) evidently do not maximize the power of this test. Explain why not. Should the ratio of n_1 to n_2 be larger or smaller than the one you chose in part b)?

ANSWER [4 marks for part a), 7 marks for parts b), c) and d)]

a) We simply need to create a sample with $n = n_1 + n_2$ observations and define a regressor d that equals 1 when an observation comes from sample 2 and 0 otherwise. Then we run the regression

$$y_i = \mu_1 + \delta d_i + u_i,$$

where $\delta \equiv \mu_2 - \mu_1$. An ordinary t-test for $\delta = 0$ provides an asymptotically valid test for the hypothesis that $\mu_1 = \mu_2$. Since we did not assume normality, the test is not exact.

b) To maximize the power of the test, we need to minimize the standard error of $\hat{\delta}$ from the equation of part a). The (true) standard error is

$$\sigma(\boldsymbol{d}^\top \boldsymbol{M}_{\iota} \boldsymbol{d})^{-1/2},$$

where $\boldsymbol{M}_{\iota}$ takes deviations from the mean. Thus we want to choose n_1 and n_2 to minimize $\boldsymbol{d}\boldsymbol{M}_{\iota}\boldsymbol{d}$. This quantity is the square of $\boldsymbol{M}_{\iota}\boldsymbol{d}$, each element of which is either $-\bar{d}$ (for the first n_1 observations) or $1-\bar{d}$ (for the last n_2 observations). Note that $\bar{d} = n_2/n$. Thus

$$\begin{aligned} \boldsymbol{d}^\top \boldsymbol{M}_{\iota} \boldsymbol{d} &= n_1 \frac{n_2^2}{n^2} + n_2 \frac{n_1^2}{n^2} \\ &= (n - n_2)\frac{n_2^2}{n^2} + n_2 \frac{(n-n_2)^2}{n^2} \\ &= nn_2^2 - n_2^3 + n_2 n^2 - 2nn_2^2 + n_2^3 \\ &= n_2 n^2 - nn_2^2 = nn_2(n - n_2) = n_1 n_2 n. \end{aligned}$$

It is obvious, and could be proved by calculus if n_2 were a real number instead of an integer, that, conditional on n, this quantity is maximized when $n_1 = n_2$. Thus we want each sample to have 125 observations.

c) First, we have to estimate γ. The obvious way to do this is to estimate σ_1 and σ_2 as the standard errors of the residuals from the two subsamples for the unrestricted regression (which is equivalent to estimating two subsample means). Then $\hat{\gamma}$ is just $\hat{\sigma}_1/\hat{\sigma}_2$.

Next, we use feasible GLS. This just means using weighted least squares to estimate our test regression. The weights are 1 for the first subsample and $\hat{\gamma}$ for the second. Thus we decrease the weights for the second sample because the disturbances for the first sample have a smaller variance.

d) We can still use the t statistic for $\delta = 0$, but this time it comes from the weighted regression. This will not be an exact test, and it will probably be less reliable than the original one, but it should be asymptotically valid.

It is no longer optimal to set $n_1 = n_2$. We now want to have more observations in the second subsample, because it has the larger variance. Students were not asked to figure out the optimal values.

6. This question deals with the nonlinear regression model

$$y_t = \alpha + \gamma \sum_{j=0}^{5} \big(\beta^j x_{t-j}\big) + u_t, \tag{8}$$

where it is assumed that $\mathrm{E}(u_t \,|\, x_t, x_{t-1}, \ldots) = 0$, that $\mathrm{E}(u_t u_s) = 0$ for $s \neq t$, and that $\mathrm{E}(u_t^2) = \sigma_t^2$, with the σ_t^2 unknown. You wish to estimate the parameters α, β, and γ using 324 monthly observations. Note that β^j means β raised to the power j.

a) Would nonlinear least squares estimates of (8) be unbiased? Would they be consistent? Explain briefly.

b) The model (8) involves nonlinear restrictions on a linear regression model. What is the unrestricted model, and how many restrictions are there? How would you test these restrictions in the context of NLS estimation under the stated assumptions?

c) Explain how you could obtain estimates of the three parameters in (8) that are consistent and asymptotically more efficient than the NLS estimates.

d) How you would test the restrictions of part b) when equation (8) is estimated by the method you proposed in part c)? Explain briefly.

ANSWER [5 marks for parts a) and d), 7 for b), 8 for c)]

a) (5 points) In general, NLS is biased, and that will certainly be the case here, because of the nonlinearity. However, the estimates will be consistent, because the disturbances have expectation zero conditional on the regressors. The conditional heteroskedasticity does not affect consistency.

b) The unrestricted model is

$$y_t = \alpha + \sum_{j=0}^{5} \delta_j x_{t-j} + u_t.$$

This model has 7 parameters. The model (8) has just 3. Thus there are 4 restrictions. In order to test the restrictions, we cannot use a nonlinear F test because of the heteroskedasticity. We can use an LM test, probably based on a GNR, or a Wald test.

For the Wald test, we need to estimate the unrestricted model, obtain an HCCME for the 7 parameters, write down the 4 restrictions in terms of those parameters, and find the derivatives of the restrictions in terms of them. This eventually gets us to the Wald statistic

$$W(\hat{\boldsymbol{\delta}}) = (\hat{\boldsymbol{R}}\hat{\boldsymbol{\delta}})^{\top}\big(\hat{\boldsymbol{R}}\widehat{\mathrm{Var}}(\hat{\boldsymbol{\delta}})\hat{\boldsymbol{R}}^{\top}\big)^{-1}(\hat{\boldsymbol{R}}\hat{\boldsymbol{\delta}}),$$

where $\hat{\boldsymbol{R}}$ is the matrix of derivatives of the restrictions evaluated at the unrestricted estimates. There is no vector $\boldsymbol{r}$ in this case because the restrictions do not involve any constants.

The restrictions are

$$\gamma = \delta_0,\;\; \gamma\beta = \delta_1,\;\; \gamma\beta^2 = \delta_2, \ldots,\;\; \gamma\beta^3 = \delta_4,\;\; \gamma\beta^4 = \delta_5$$

We have to rewrite these purely as functions of the δ_j. Evidently,

$$\delta_0/\delta_2 - \delta_1/\delta_2 = 0,\;\; \delta_1/\delta_2 - \delta_2/\delta_3 = 0,\;\; \delta_2/\delta_3 - \delta_3/\delta_4 = 0,\;\; \delta_3/\delta_4 - \delta_4/\delta_5 = 0.$$

These impose four restrictions on the six δ_j.

Alternatively, we could run the GNR for the unrestricted model, which is basically the one for the restricted model plus four lags of x_t. Then test whether the coefficients on those four lags equal zero. With homoskedasticity, we could use an F test. With heteroskedasticity, we need to use a Wald test based on the GNR, or perhaps use an HRGNR.

c) Use GMM. First, estimate the unrestricted model by OLS to obtain a vector of residuals. Then minimize the GMM criterion function based on seven moment conditions:

$$\boldsymbol{\iota}^\top\big(\boldsymbol{y}-\boldsymbol{f}(\alpha,\beta,\gamma)\big),\;\; \boldsymbol{x}^\top\big(\boldsymbol{y}-\boldsymbol{f}(\alpha,\beta,\gamma)\big),$$
$$\boldsymbol{x}_1{}^\top\big(\boldsymbol{y}-\boldsymbol{f}(\alpha,\beta,\gamma)\big),\ldots,\;\; \boldsymbol{x}_5{}^\top\big(\boldsymbol{y}-\boldsymbol{f}(\alpha,\beta,\gamma)\big),$$

where $\boldsymbol{x}_j$ means a vector with typical element x_{t-j}. The criterion function then has the form

$$\big(\boldsymbol{y}-\boldsymbol{f}(\alpha,\beta,\gamma)\big)^\top\boldsymbol{W}(\boldsymbol{W}^\top\hat{\boldsymbol{\Omega}}\boldsymbol{W})^{-1}\boldsymbol{W}^\top\big(\boldsymbol{y}-\boldsymbol{f}(\alpha,\beta,\gamma)\big),$$

where $\hat{\boldsymbol{\Omega}}$ is a diagonal matrix with squared residuals on the diagonal, and $\boldsymbol{W}$ denotes the matrix $[\boldsymbol{\iota},\boldsymbol{x},\boldsymbol{x}_1,\ldots,\boldsymbol{x}_5]$.

d) Use a test based on the criterion function. This is really easy because the unrestricted value is 0, since $\boldsymbol{W}$ is orthogonal to the unrestricted residuals. The test statistic is simply the criterion function from part c), which is asymptotically distributed as $\chi^2(4)$.

Table 1. Some Critical Values of the χ^2 Distribution

D.F. / Level	.10	.05	.025	.01
1	2.706	3.841	5.024	6.635
2	4.605	5.991	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086
6	10.645	12.592	14.449	16.812