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

Economics 850 | **Econometrics I** | **Fall, 2019**

Professor James MacKinnon

Answers to Final Examination

December 9, 2019. 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 $\boldsymbol{X}_2$. The elements u_i of $\boldsymbol{u}$ are assumed to be uncorrelated but to have different variances σ_i^2, which are unknown.

a) Explain how you would test the hypothesis that $\boldsymbol{\beta}_2 = \boldsymbol{0}$ using an asymptotic test. Write down the test statistic explicitly as a function of $\boldsymbol{y}$, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and functions of those things, using projection matrices to simplify the algebra. How is this statistic distributed asymptotically under the null hypothesis?

b) Explain how you would generate B bootstrap samples for a bootstrap test of the hypothesis $\boldsymbol{\beta}_2 = \boldsymbol{0}$ based on the test statistic of part a).

c) Let τ denote the actual test statistic from part a) and τ_b denote the bootstrap statistic for the b^{th} bootstrap sample from part b). Precisely how would you compute the bootstrap P value for the test of $\boldsymbol{\beta}_2 = \boldsymbol{0}$.

d) Suppose a student used $B = 100$ and obtained a bootstrap P value of 0.07 using the procedure of parts b) and c). The student concludes that the hypothesis that $\boldsymbol{\beta}_2 = \boldsymbol{0}$ cannot be rejected at the .05 level. Would you congratulate the student on a job well done? Why or why not?

Continued on next page ...

ANSWER

a) [7] Use a Wald test based on an HCCME. The FWL regression is

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

Thus

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

The HCCME for $\hat{\boldsymbol{\beta}}_2$ is

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

Thus the Wald statistic is

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

because the product of the factors $\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2$ and $(\boldsymbol{X}_2^\top\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}$ is an identity matrix. This test statistic is asymptotically distributed as $\chi^2(k_2)$.

b) [6] Use the restricted ordinary wild bootstrap. Estimate the model under the null to obtain estimates $\tilde{\boldsymbol{\beta}}_1$ and residuals $\tilde{\boldsymbol{u}}$. Rescale the residuals by dividing each of them by the square root of the corresponding diagonal of the hat matrix. (I did not deduct marks if they did not do this.) Then generate each bootstrap sample as

$$\boldsymbol{y}_b^* = \boldsymbol{X}_1\tilde{\boldsymbol{\beta}}_1 + \boldsymbol{v}_b * \tilde{\boldsymbol{u}},$$

where $\boldsymbol{v}_b$ is a vector of realizations from the Rademacher distribution.

c) [5] The important thing here is to perform a test that rejects in the upper tail only. The bootstrap P value is

$$\frac{1}{B}\sum_{b=1}^{B}\mathbb{I}(\tau_b > \tau).$$

d) [7] $B = 100$ is a bad choice for two reasons. It is too small, and it does not satisfy the condition that $\alpha(B+1)$ be an integer for any sensible value of α. With $B = 99$, there are $5/100$ cases that lead to rejection. But with $B = 100$, there are $6/101$ cases (assuming that $P = 0.05$ counts as a rejection).

Since the poor choice of B leads to overrejection, one might think that it is safe to not reject here with $P = 0.07$. But that number could easily have occurred by chance. We need a much larger value of B before we can be confident that P is greater (or less) than .05.

Continued on next page ...

2. Consider the nonlinear regression model

$$y_i = \beta_1 + \beta_2(x_i^\alpha + z_i^\alpha) + u_i, \quad \mathrm{E}(u_i \,|\, x_i, z_i) = 0, \tag{2}$$

where the disturbances u_i are assumed to be independent, but it is not assumed that they are homoskedastic.

a) Explain how you could you test the hypothesis that $\alpha = 0.5$ without doing any nonlinear estimation at all.

b) Suppose that you estimate equation (2) by nonlinear least squares. If the program you are using is not capable of computing heteroskedasticity-robust standard errors for NLS estimates, but can do so for OLS estimates, explain how you would compute an asymptotically valid standard error for $\hat{\alpha}$ and how you would use it to construct a 95% confidence interval.

c) Suppose equation (2) is to be estimated using a sample of 2150 observations, of which 1110 come from Ontario and 1040 come from Quebec. You wish to test the null hypothesis that all the parameters are the same for the two provinces against the alternative that they are different. Explain how to do so using NLS just once (for the restricted model) along with one OLS regression. How is your test statistic distributed asymptotically under the null?

d) Now explain how to test the hypothesis of part c) using NLS just once (for an unrestricted model). Will your test statistic be numerically equal to the one from part c)? Will it have the same asymptotic distribution under the null? Explain briefly.

ANSWER

a) [7] First, estimate the linear regression model

$$y_i = \beta_1 + \beta_2(x_i^{0.5} + z_i^{0.5}) + u_i$$

by OLS. Then obtain the GNR for the unrestricted model and run it, with the regressand and regressors evaluated at the restricted estimates. In general, the GNR is

$$y_i - \beta_1 - \beta_2(x_i^\alpha + z_i^\alpha) = b_1 + b_2(x_i^\alpha + z_i^\alpha) + b_3\big(x_i^\alpha \log(x_i) + z_i^\alpha \log(z_i)\big) + \text{resids.}$$

Thus the test regression is

$$\tilde{u}_i = b_1 + b_2(x_i^{0.5} + z_i^{0.5}) + b_3\big(x_i^{0.5} \log(x_i) + z_i^{0.5} \log(z_i)\big) + \text{resids.}$$

Use a heteroskedasticity-robust t statistic for $b_3 = 0$ as the test statistic.

b) [6] Again, use a GNR, but this time an unrestricted one. The GNR is

$$\hat{u}_i = b_1 + b_2(x_i^{\hat{\alpha}} + z_i^{\hat{\alpha}}) + b_3\big(x_i^{\hat{\alpha}} \log(x_i) + z_i^{\hat{\alpha}} \log(z_i)\big) + \text{resids.}$$

Continued on next page . . .

All three estimated coefficients here should equal 0, but the hetero-robust standard error for $\hat{b}_3$ provides an asymptotically valid standard error. Call it $\text{se}(\hat{\alpha})$. Then the confidence interval is

$$\left[\hat{\alpha} - 1.96\text{se}(\hat{\alpha}),\;\; \hat{\alpha} + 1.96\text{se}(\hat{\alpha})\right].$$

c) [6] The restricted model here is the unrestricted model for part b), but estimated over the data for both provinces. Now the GNR has six regressors instead of three. The first three regressors have the same form as the ones in the GNR of part b), but they depend on different estimates The other three are the same regressors multiplied by a dummy variable for either Ontario or Quebec. If the coefficients on the latter three regressors are c_1 through c_3, the test statistic has the form of a hetero-robust Wald test for those three coefficients to be zero. It must be asymptotically distributed as $\chi^2(3)$.

d) [6] Now we need to estimate the unrestricted model

$$y_i = \beta_1 + \beta_2(x_i^\alpha + z_i^\alpha) + \gamma_1\text{DQ}_i + \gamma_2\text{DQ}_i(x_i^\delta + z_i^\delta) + u_i,$$

where DQ_i is the Quebec dummy. Estimate all six parameters by NLS. Then compute a hetero-robust Wald test for $\gamma_1 = \gamma_2 = \delta = 0$. It will not be numerically equal to the test statistic of part c), but it will have the same asymptotic distribution under the null, that is, $\chi^2(3)$.

3. Consider the linear regression model

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \tag{3}$$

which is to be estimated using a sample of 187 observations. The regressors x_{2i} and x_{3i} are assumed to be exogenous. You are interested in the parameter $\gamma \equiv \beta_2/\beta_3$.

a) Explain how you would obtain an estimate $\hat{\gamma}$ and an asymptotically valid standard error $s(\hat{\gamma})$ analytically (i.e., without doing any simulations) under the assumption that the disturbances in (3) are independent but may display heteroskedasticity of unknown form.

b) Explain how you would obtain a bootstrap standard error $s^*(\hat{\gamma})$ under the assumptions of part a). How could you use that standard error to construct a confidence interval for γ? Would your interval be symmetric around $\hat{\gamma}$?

c) Explain how you would perform a bootstrap test of the hypothesis that $\gamma = 1.25$ under the assumptions of part a).

d) Explain how you would construct a studentized bootstrap confidence interval for γ under the assumptions of part a). Would your interval be symmetric around $\hat{\gamma}$? Why or why not?

Continued on next page ...

ANSWER

a) [7] Estimate (3) by OLS and obtain $\hat\gamma \equiv \hat\beta_2/\hat\beta_3$. Use a hetero-robust covariance matrix estimator to obtain $\mathrm{Var}(\hat\beta_2)$, $\mathrm{Var}(\hat\beta_3)$, and $\mathrm{Cov}(\hat\beta_2, \hat\beta_3)$. These are estimates, but I am not bothering with the wide hats.

Now use the delta-method formula that

$$\mathrm{Var}(\hat\gamma) = \hat{\boldsymbol{G}}\hat{\boldsymbol{V}}\hat{\boldsymbol{G}}^\top.$$

Here $\hat{\boldsymbol{G}}$ is a vector of the derivatives of γ with respect to β_2 and β_3. These are

$$[\,1/\hat\beta_3 \quad -\hat\beta_2/\hat\beta_3^2\,].$$

Thus we need to evaluate the quadratic form

$$[\,1/\hat\beta_3 \quad -\hat\beta_2/\hat\beta_3^2\,]\begin{bmatrix}\mathrm{Var}(\hat\beta_2) & \mathrm{Cov}(\hat\beta_2,\hat\beta_3)\\ \mathrm{Cov}(\hat\beta_2,\hat\beta_3) & \mathrm{Var}(\hat\beta_3)\end{bmatrix}\begin{bmatrix}1/\hat\beta_3\\ -\hat\beta_2/\hat\beta_3^2\end{bmatrix}.$$

The result is

$$\mathrm{Var}(\hat\beta_2)/\hat\beta_3 + \frac{\hat\beta_2^2}{\hat\beta_3^4}\,\mathrm{Var}(\hat\beta_3) - 2\frac{\hat\beta_2}{\hat\beta_3^3}\,\mathrm{Cov}(\hat\beta_2,\hat\beta_3).$$

The estimated standard error of $\hat\gamma$, say s_γ, is the square root of this expression (if I have done the algebra correctly!). Students do not have to work out what the quadratic form is equal to.

b) [6] Use the ordinary wild bootstrap based on unrestricted estimates. The bootstrap DGP is

$$y_i^{*b} = \hat\beta_1 + \hat\beta_2 x_{2i} + \hat\beta_3 x_{3i} + \hat u_i v_i^b,$$

where as usual v_i^b is a realization of the Rademacher distribution. It would be better to rescale the $\hat u_i$ first by dividing them by the square roots of the diagonals of the hat matrix, but I won't deduct marks. Generate B bootstrap samples, estimate the model and thus obtain $\hat\gamma_b^*$ for each of them. Then use the standard deviation of the $\hat\gamma_b^*$ as $s^*(\hat\gamma)$. Finally, construct the confidence interval

$$\big[\hat\gamma - c_{1-\alpha/2}s^*(\hat\gamma),\ \ \hat\gamma + c_{1-\alpha/2}s^*(\hat\gamma)\big].$$

Here $c_{1-\alpha/2}$ is the appropriate quantile of the normal distribution if we want an interval at level $1-\alpha$. This interval will evidently be symmetric.

c) [6] One possibility is to calculate the test statistic $\tau = (\hat\gamma - 1.25)/s_\gamma$, where s_γ came from part a).

Estimate the model under the restriction that $\gamma = 1.25$. This can be done by running the regression

$$y_i = \beta_1 + 1.25\beta_3(x_{2i} + x_{3i}) + u_i,$$

Continued on next page ...

from which we obtain estimates $\tilde\beta_1$ and $\tilde\beta_3$, along with restricted residuals $\tilde u_i$. Then generate B bootstrap samples from the wild bootstrap DGP

$$y_i^{*b} = \tilde\beta_1 + 1.25\tilde\beta_3(x_{2i} + x_{3i}) + \tilde u_i v_i^b.$$

For each bootstrap sample, estimate the unrestricted model and calculate $\hat\gamma_b^*$ and its standard error s_γ^{*b} using the formula from part a). Use these to calculate

$$\tau_b^* = \frac{\hat\gamma_b^* - 1.25}{s_\gamma^{*b}}.$$

Then calculate an equal-tailed P value as

$$P^*(\tau) = \frac{2}{B}\min\Bigg(\sum_{b=1}^{B}\mathbb{I}(\tau_b^* \le \hat\tau),\ \sum_{b=1}^{B}\mathbb{I}(\tau_b^* > \hat\tau)\Bigg).$$

Alternatively, and this is easier, construct a t statistic for the hypothesis that $\beta_2 = 1.25\beta_3$. This has the form

$$\tau = \frac{\hat\beta_2 - 1.25\hat\beta_3}{\big(\widehat{\mathrm{Var}}(\hat\beta_2 - 1.25\hat\beta_3)\big)^{1/2}}.$$

Here the variance is simply

$$\mathrm{Var}(\hat\beta_2) + 1.25^2\,\mathrm{Var}(\hat\beta_3) - 2.5\,\mathrm{Cov}(\hat\beta_2, \hat\beta_3).$$

For the bootstrap, proceed as before. It is better to use an equal-tail P value than a symmetric one. If the Wald statistic τ^2 is used, we necessarily use the latter [1 mark off].

d) [6] For the studentized bootstrap confidence interval, we need to use an unrestricted wild bootstrap DGP, as in part b). Each bootstrap sample yields $\hat\gamma_b^*$ and s_γ^{*b}, from which we can compute the bootstrap test statistic

$$\hat\tau_b^* = \frac{\hat\gamma_b^* - \hat\gamma}{s_\gamma^{*b}}.$$

Sort these and obtain their $\alpha/2$ and $1-\alpha/2$ quantiles, $c_{\alpha/2}$ and $c_{1-\alpha/2}$. If B satisfies the condition that $\alpha(B+1)/2$ is an integer, these are just numbers $(B+1)\alpha/2$ and $(B+1)(1-\alpha/2)$ in the sorted list. Then the studentized bootstrap confidence interval is

$$\big[\hat\gamma - c_{1-\alpha/2}s^*(\hat\gamma),\ \ \hat\gamma - c_{\alpha/2}s^*(\hat\gamma)\big].$$

Continued on next page ...

It will evidently not be symmetric around $\hat{\gamma}$, because, with probability essentially 1, $c_{1-\alpha/2} \neq -c_{\alpha/2}$.

4. Suppose that you have samples of households drawn from ten different cities, where, for the i^{th} city, the sample is of size n_i. You believe that an appropriate model is

$$\boldsymbol{y}_i = \alpha_i \boldsymbol{\iota}_i + \boldsymbol{X}_i \boldsymbol{\beta} + \boldsymbol{u}_i, \quad \mathrm{E}(\boldsymbol{u}_i \boldsymbol{u}_i^\top) = \boldsymbol{\Omega}_i. \tag{4}$$

Here the $\boldsymbol{y}_i$ are $n_i \times 1$ vectors of observations on the dependent variables, the $\boldsymbol{\iota}_i$ are $n_i \times 1$ vectors of 1s, the $\boldsymbol{u}_i$ are $n_i \times 1$ vectors of disturbances, the $\boldsymbol{X}_i$ are $n_i \times k$ matrices of observations on exogenous variables, and the $\boldsymbol{\Omega}_i$ are positive definite covariance matrices of dimension $n_i \times n_i$.

a) Explain how you would estimate the parameters α_i, $i = 1, \ldots, 10$, and $\boldsymbol{\beta}$ jointly by ordinary least squares. Then explain how you would estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$, the OLS estimate of $\boldsymbol{\beta}$.

b) Regression (4) involves $k+10$ regressors. Explain how you could obtain exactly the same estimates $\hat{\boldsymbol{\beta}}$ by running a regression with only k regressors. How would you estimate the covariance matrix of $\hat{\boldsymbol{\beta}}$ now?

c) Suppose you wish to test the restrictions that $\beta_1 = \beta_2 = \beta_3$, where these parameters are the first three elements of $\boldsymbol{\beta}$. Explain how you would compute a test statistic that asymptotically follows a $\chi^2(r)$ distribution. What is the value of r? If the test statistic were 6.253, would the null hypothesis be rejected at the .05 level using an asymptotic test? Would it be wise to rely on such a test in this case?

d) Briefly explain how you could instead perform a bootstrap test of the same null hypothesis. If $\hat{\tau} = 6.253$ is the actual test statistic and τ_b denotes the b^{th} bootstrap statistic out of 999 such statistics, precisely how would you calculate the bootstrap P value?

ANSWER

a) [6] Just stack all the data for the $\boldsymbol{y}_i$ and $\boldsymbol{X}_i$ into a vector $\boldsymbol{y}$ and a matrix $\boldsymbol{X}$ with $n = \sum n_i$ observations. Then regress $\boldsymbol{y}$ on $\boldsymbol{X}$ and ten dummy variables, one for each city. For the covariance matrix, just use a CRVE with ten clusters, one for each city. It would be good, but is not essential, to write down the usual formula

$$(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \left(\sum_{i=1}^{10} \boldsymbol{Z}_i^\top \hat{\boldsymbol{u}}_i \hat{\boldsymbol{u}}_i^\top \boldsymbol{Z}_i \right) (\boldsymbol{Z}^\top \boldsymbol{Z})^{-1},$$

multiplied by the scalar factor $(10/9)(n-1)/(n-k-10)$. Here $\boldsymbol{Z}$ and the $\boldsymbol{Z}_i$ contain $\boldsymbol{X}$ or $\boldsymbol{X}_i$ plus all ten dummies.

b) [6] Just project $\boldsymbol{y}$ and $\boldsymbol{X}$ off the dummies, to obtain $\ddot{\boldsymbol{y}}$ and $\ddot{\boldsymbol{X}}$. Regressing $\ddot{\boldsymbol{y}}$ on $\ddot{\boldsymbol{X}}$ should yield the same estimates. Then the covariance matrix has the same form as before, but with $\ddot{\boldsymbol{y}}$ and the $\ddot{\boldsymbol{y}}_i$ replacing $\boldsymbol{y}$ and the $\boldsymbol{y}_i$, and with $\ddot{\boldsymbol{X}}$ and the $\ddot{\boldsymbol{X}}_i$ replacing $\boldsymbol{Z}$ and the $\boldsymbol{Z}_i$.

c) [7] It is natural to use a Wald statistic. There are two restrictions, not three. They can be written as $\beta_1 - \beta_2 = 0$ and $\beta_2 - \beta_3 = 0$. Find the covariance matrix of $\hat{\boldsymbol{\beta}}$ using the CRVE in part a) or part b). In either case, extract the 3×3 matrix $\hat{\boldsymbol{V}}$ that corresponds to the first three coefficients. Then the Wald statistic is

$$W = \begin{bmatrix} \hat{\beta}_1 & -\hat{\beta}_2 & 0 \\ 0 & \hat{\beta}_2 & -\hat{\beta}_3 \end{bmatrix} \hat{\boldsymbol{V}}^{-1} \begin{bmatrix} \hat{\beta}_1 & 0 \\ -\hat{\beta}_2 & \hat{\beta}_2 \\ 0 & -\hat{\beta}_3 \end{bmatrix}.$$

This should be asymptotically distributed as $\chi^2(2)$.

The value of 6.253 falls between the .05 and .025 critical values of the $\chi^2(2)$ distribution. It is fairly close to the former. Since the test is probably not very reliable, because there are only ten clusters, it would be unwise to reject the null hypothesis based on this result.

d) [6] Here the bootstrap P value is

$$P^*(\hat{\tau}) = \frac{1}{999}\sum_{b=1}^{999} \mathbb{I}(\tau_b^* > \hat{\tau}) = \frac{1}{999}\sum_{b=1}^{999} \mathbb{I}(\tau_b^* > 6.253).$$

Note that this is a one-tail test in the upper tail.

5. Consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} = \boldsymbol{Z}\boldsymbol{\beta}_1 + \boldsymbol{Y}\boldsymbol{\beta}_2 + \boldsymbol{u}. \tag{5}$$

Here $\boldsymbol{y}$ and $\boldsymbol{u}$ are $n \times 1$, $\boldsymbol{X} \equiv [\boldsymbol{Z}\ \boldsymbol{Y}]$ is $n \times k$, $\boldsymbol{Z}$ is $n \times k_1$, and $\boldsymbol{Y}$ is $n \times k_2$. The columns of $\boldsymbol{Z}$ are exogenous regressors, while the columns of $\boldsymbol{Y}$ are possibly endogenous. The k–vector of coefficients $\boldsymbol{\beta}$ is divided into two subvectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ of dimensions k_1 and k_2, respectively.

a) Suppose that, in addition to $\boldsymbol{Z}$, you have available an $n \times \ell$ matrix of instruments $\boldsymbol{W}$, with ℓ large enough so that $\boldsymbol{\beta}$ is either exactly identified or overidentified. What is the IV estimator $\hat{\boldsymbol{\beta}}$? This estimator can be obtained by minimizing a certain criterion function. What is that function?

b) How many overidentifying restrictions are there? Explain how you can test these overidentifying restrictions under the assumption that the elements of $\boldsymbol{u}$ are homoskedastic. What would you conclude if $n = 17,243$, $k_1 = 12$, $k_2 = 2$, $\ell = 5$, and the value of your test statistic were 15.24?

c) Asymptotically, it must be true that

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \boldsymbol{A}\boldsymbol{d}, \tag{6}$$

Continued on next page . . .

where $\boldsymbol{A}$ is a $k\times k$ matrix and $\boldsymbol{d}$ is a $k\times 1$ vector. Rewrite equation (6) using the actual expressions for $\boldsymbol{A}$ and $\boldsymbol{d}$, making sure that they include the appropriate factors of n,

d) Suppose that the i^{th} element u_i of $\boldsymbol{u}$ has mean 0 and variance σ_i^2. Based on your answer to part c), you should be able to derive the asymptotic distribution of $n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$. What is this distribution? Just give the result. Do not attempt to prove anything.

ANSWER

a) [5] The estimator is just the usual GIVE estimator:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{y},$$

where $\boldsymbol{W}^* = [\boldsymbol{W} \;\; \boldsymbol{Z}]$. It can be obtained by minimizing the criterion function

$$(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}})^\top \boldsymbol{P}_{\boldsymbol{W}^*}(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}).$$

b) [6] In general, there are $\ell - k_2$ overidentifying restrictions. Just divide the minimized value of the criterion function by the IV estimate of σ^2. The resulting test statistic will be asymptotically distributed as $\chi^2(\ell - k_2)$. In this case, $\ell - k_2 = 3$. The value 15.24 is quite a bit larger than even the .01 critical value for $\chi^2(3)$, so you should reject the null. Either at least one column of $\boldsymbol{W}$ should have been included in the regression, or at least one column of $\boldsymbol{W}$ is not a valid instrument.

c) [7] The actual form of (6) is

$$n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) = \left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{X}\right)^{-1} n^{-1/2}\boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{u}.$$

The plim is important.

d) [7] The asymptotic distribution of $n^{-1/2}\boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{u}$ is multivariate normal with mean vector zero and covariance matrix

$$\operatorname*{plim}_{n\to\infty}\left(\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{\Omega}\boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{X}\right),$$

where $\boldsymbol{\Omega}$ is a diagonal matrix with typical diagonal element σ_i^2. Therefore, $n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)$ is asymptotically multivariate normal with mean vector zero and covariance matrix

$$\left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{X}\right)^{-1}\left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{\Omega}\boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{X}\right)\left(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{X}^\top \boldsymbol{P}_{\boldsymbol{W}^*}\boldsymbol{X}\right)^{-1}.$$

Continued on next page . . .

6. Consider the linear regression model

$$\boldsymbol{y} = \beta_1 \boldsymbol{\iota} + \beta_2 \boldsymbol{x} + \boldsymbol{u}, \quad \boldsymbol{u} \sim \mathrm{NID}(\mathbf{0}, \sigma^2 \mathbf{I}), \tag{2}$$

where all vectors have n elements. Here $\boldsymbol{\iota}$ denotes a vector of 1s, and $\boldsymbol{x}$ denotes a vector of observations on an exogenous regressor.

a) Write down the t statistic for $\beta_2 = 1$ explicitly and explain how to make a regression package display it as part of the output from a linear regression. Precisely how is this statistic distributed under the null hypothesis?

b) How is the t statistic of part a) distributed under the fixed alternative that $\beta_2 = 1 + \delta$, for $\delta \neq 0$?

c) Using the result of part b), explain how to calculate the power of a one-tailed test that $\beta_2 \leq 1$ against the alternative that $\beta_2 > 1$ as a function of δ. How would power for $\delta = 0.3$ change if the sample size n were doubled? How would it change if σ were doubled?

d) Again using the result of part b), explain how to calculate the power of a two-tailed test that $\beta_2 = 1$ against the alternative that $\beta_2 \neq 1$ as a function of δ. When $\delta > 0$, will this test be more powerful than the test of part c), equally powerful, or less powerful? Explain carefully.

ANSWER

a) [6] The t statistic is

$$\frac{(\boldsymbol{x}^\top \boldsymbol{M}_\iota \boldsymbol{x})^{-1} \boldsymbol{y}^\top \boldsymbol{M}_\iota \boldsymbol{x} - 1}{s^2 (\boldsymbol{x}^\top \boldsymbol{M}_\iota \boldsymbol{x})^{1/2}}.$$

To make a regression package display it, run the regression

$$\boldsymbol{y} - \boldsymbol{x} = \beta_1 \boldsymbol{\iota} + (\beta_2 - 1)\boldsymbol{x} + \boldsymbol{u}.$$

The t statistic on $\boldsymbol{x}$ is what we want. Under the null, it follows the Student's t distribution with $n - 2$ degrees of freedom.

b) [7] When $\beta_2 = 1 + \delta$, the t statistic will follow a noncentral t distribution with $n - 2$ degrees of freedom. Instead of having mean 0, it will have mean

$$\lambda = \frac{(\boldsymbol{x}^\top \boldsymbol{M}_\iota \boldsymbol{x})^{-1} \boldsymbol{x}^\top \boldsymbol{M}_\iota \boldsymbol{x}(1 + \delta) - 1}{\sigma^2 (\boldsymbol{x}^\top \boldsymbol{M}_\iota \boldsymbol{x})^{1/2}} = \frac{\delta}{\sigma^2 (\boldsymbol{x}^\top \boldsymbol{M}_\iota \boldsymbol{x})^{1/2}}.$$

This is the NCP.

c) [6] To calculate the power of the test, we need to find the probability that a t statistic with mean λ given above will be greater than an appropriate critical value as a function of δ. The probability that $t(\lambda) > C$ is the same as the

Continued on next page ...

probability that $t(0) > C - \lambda$. For example, if $C = 1.96$, the probability is 0.025 when $\lambda = 0$ and 0.5 when $\lambda = 1.96$.

For given δ, λ is inversely proportional to the square root of the sample size, because $\boldsymbol{x}^\top \boldsymbol{M}_{\boldsymbol{\iota}} \boldsymbol{x} = O(n)$. It is also inversely proportional to σ. Thus power would go up more when σ is doubled than when n is doubled. Just how much depends on the value of the (central) Student's t distribution for the original value of λ and two values associated with 2σ or $2n$.

d) [6] Figuring out the power of a two-tail test is harder, because we reject in both tails. Now we need

$$\Pr\bigl(t(0) > C - \lambda\bigr) + \Pr\bigl(t(0) \le -C - \lambda\bigr),$$

for a different value of C. The new value of C corresponds to the $1 - \alpha/2$ quantile instead of the $1 - \alpha$ quantile, so it is larger. This implies that power will be lower than for the one-tail test. When $\lambda = 0$, both tests have the same power. But as δ, and hence λ, increases, the gain in power in the upper tail is partly offset by a loss of power in the lower tail.

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