ECONOMETRICS FINAL EXAM

PART B

Universidad Carlos III de Madrid

21/05/24

Write your name and group in each answer sheet. Answer the questions in 2:00 hours.

1. (30%) Consider the multiple regression model,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 \ln(X_2) + U,$$

where $E\left(U|\, X_1, X_2\right) = 0$, and the rest of conditions for consistency and asymptotic normality of OLS estimators hold. We have *iid* observations $\{Y_i, X_{1i}, X_{2i}\}_{i=1}^n$ of (Y, X_1, X_2).

a. (1/3) Show that the OLS estimator of β_1 is

$$\hat{\beta}_1 = \frac{\widehat{Cov}\left(Y, \hat{\varepsilon}_1\right)}{\widehat{Var}\left(\hat{\varepsilon}_1\right)},$$

where $\hat{\varepsilon}_{1i} = X_{1i} - \hat{X}_{1i}$, and $\hat{X}_{1i}$ are the OLS fitted values of X_1 given $\ln(X_2)$.

<u>ANSWER</u>:

Let $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ be the OLS estimators of β_0, β_1, β_2, respectively. The corresponding residuals $\hat{U}_i = Y_i - \hat{Y}_i$ with $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 \ln(X_{2i})$, satisfy,

$$\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i = 0 \tag{1}$$

$$\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i X_{1i} = 0 \Leftrightarrow \widehat{Cov}\left(\hat{U}, X_1\right) = 0 \tag{2}$$

$$\frac{1}{n}\sum_{i=1}^{n} \hat{U}_i \ln(X_{2i}) = 0 \Leftrightarrow \widehat{Cov}\left(\hat{U}, \ln(X_2)\right) = 0, \tag{3}$$

Also, we know that

$$\widehat{Cov}\left(\hat{\varepsilon}_1, \ln(X_2)\right) = \overline{\hat{\varepsilon}}_1 = 0, \tag{4}$$

where $\hat{\varepsilon}_{1i} = X_{1i} - \hat{X}_{1i}$ are the OLS residuals of the fit of X_1 on $\ln\left(X_2\right)$, where $\hat{X}_{1i} = \hat{\delta}_0 + \hat{\delta}_1 \ln\left(X_{2i}\right)$

are the X_1 OLS predicted values. Thus,

$$\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}\hat{U}_i\hat{\varepsilon}_{1i} \\
&= \widehat{Cov}\left(Y-\hat{\beta}_0-\hat{\beta}_1X_1-\hat{\beta}_2\ln(X_2),\hat{\varepsilon}_1\right) \\
&= \widehat{Cov}(Y,\hat{\varepsilon}_1)+\underbrace{\widehat{Cov}\left(-\hat{\beta}_0,\hat{\varepsilon}_1\right)}_{=0}+\underbrace{\widehat{Cov}\left(-\hat{\beta}_1X_1,\hat{\varepsilon}_1\right)}_{=-\hat{\beta}_1\widehat{Cov}(X_1,\hat{\varepsilon}_1)}+\underbrace{\widehat{Cov}\left(-\hat{\beta}_2\ln(X_2),\hat{\varepsilon}_1\right)}_{\substack{=-\hat{\beta}_2\widehat{Cov}(\ln(X_2),\hat{\varepsilon}_1)\\ =0\text{ by }(4)}} \\
&= \widehat{Cov}(Y,\hat{\varepsilon}_1)-\hat{\beta}_1\widehat{Cov}(X_1,\hat{\varepsilon}_1) \\
\Longrightarrow&\ \hat{\beta}_1=\frac{\widehat{Cov}(Y,\hat{\varepsilon}_1)}{\widehat{Var}(\hat{\varepsilon}_1)}
\end{aligned}$$

b. (1/3) Suppose that β_1 is estimated by $\hat{\alpha}_1=\widehat{Cov}(Y,X_1)\big/\widehat{Var}(X_1)$. Derive the asymptotic bias of $\hat{\alpha}_1$ as an estimator of β_1.

ANSWER:

$$\begin{aligned}
\hat{\alpha}_1 &= \frac{\widehat{Cov}(Y,X_1)}{\widehat{Var}(X_1)} \\
&= \frac{\widehat{Cov}\left(\hat{Y}+\hat{U},X_1\right)}{\widehat{Var}(X_1)} \\
&= \frac{\widehat{Cov}\left(\hat{\beta}_0+\hat{\beta}_1X_1+\hat{\beta}_2\ln(X_2)+\hat{U},X_1\right)}{\widehat{Var}(X_1)} \\
&= \underbrace{\frac{\widehat{Cov}\left(\hat{\beta}_0,X_1\right)}{\widehat{Var}(X_1)}}_{=0}+\hat{\beta}_1\underbrace{\frac{\widehat{Cov}(X_1,X_1)}{\widehat{Var}(X_1)}}_{=1}+\hat{\beta}_2\frac{\widehat{Cov}(\ln(X_2),X_1)}{\widehat{Var}(X_1)}+\underbrace{\frac{\widehat{Cov}\left(\hat{U},X_1\right)}{\widehat{Var}(X_1)}}_{=0\text{ by (A.2)}} \\
&= \hat{\beta}_1+\hat{\beta}_2\frac{\widehat{Cov}(\ln(X_2),X_1)}{\widehat{Var}(X_1)} \\
&\to \beta_1+\beta_2\frac{Cov(\ln(X_2),X_1)}{Var(X_1)}\ w.p.1.\text{ applying LLN}
\end{aligned}$$

$$Asymptotic\ Bias=\beta_2\frac{Cov(\ln(X_2),X_1)}{Var(X_1)}.$$

c. (1/3) Provide an estimator of the expected elasticity of Y with respect to X_2 evaluated at $X_1=x_1$

and $X_2 = x_2$.

ANSWER:

$$E\left(Y|\, X_1 = x_1, X_2 = x_2\right) = \beta_0 + \beta_1 x_1 + \beta_2 \ln(x_2).$$

The expected elasticity of Y wrt X_2 evaluated in $X_1 = x_1$ and $X_2 = x_2$ is,

$$\begin{aligned}
\xi_{Y,X_1}\left(x_1, x_2\right) &= \frac{dE\left(Y|\, X_1 = x_1, X_2 = x_2\right)}{\frac{dx_2}{x_2}} \frac{1}{E\left(Y|\, X_1 = x_1, X_2 = x_2\right)} \\
&= \frac{dE\left(Y|\, X_1 = x_1, X_2 = x_2\right)}{d\ln\left(x_2\right)} \frac{1}{E\left(Y|\, X_1 = x_1, X_2 = x_2\right)} \\
&= \beta_2 \cdot \frac{1}{\beta_0 + \beta_1 x_1 + \beta_2 \ln(x_2)}.
\end{aligned}$$

The natural estimator is

$$\hat{\xi}_{Y,X_1}\left(x_1, x_2\right) = \hat{\beta}_2 \cdot \frac{1}{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 \ln(x_2)}.$$

2. (30%) Consider the multiple regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U,$$

where $E\left(U|\, X_1, X_2, X_3\right) = 0$, $E\left(U^2|\, X_1, X_2, X_3\right) = \sigma^2\left(X_1, X_2, X_3\right)$ and $\sigma^2\left(\cdot\right)$ is an unknown function, and all the remainder assumptions for the consistency and asymptotic normality of the OLS estimator hold. We have a random sample (iid observations) $\{Y_i, X_{1i}, X_{2i}, X_{3i}\}_{i=1}^n$ of $\left(Y, X_1, X_2, X_3\right)$.

a. (3/5) Explain how to test the following three hypotheses at 5% of significance after suitably reparameterizing the models in such a way that the hypotheses can be expressed as significance tests of the resulting transformed parameters (in each case, you must provide the test statistic, its approximated distribution when the sample size is large, and the rejection rule).

i. (1/3) $H_0 : \beta_1/\,\beta_2 = 1$ & $\beta_2 = -2$ vs $H_1 : \beta_1/\,\beta_2 \neq 1$ and/or $\beta_2 \neq -2$.

ii. (1/3) $H_0 : \beta_1 + \beta_2 + \beta_3 = 0$ vs $H_1 : \beta_1 + \beta_2 + \beta_3 < 0$.

iii. (1/3) $H_0 : \beta_1 + \beta_2 + \beta_3 = 1$ vs $H_1 : \beta_1 + \beta_2 + \beta_3 \neq 1$.

ANSWER:

i. Define $\theta_1 = \beta_1 - \beta_2$ and $\theta_2 = \beta_2 + 2$. So we can write the hypothesis as

$$H_0 : \theta_1 = \theta_2 = 0 \text{ vs } H_1 : \theta_1 \neq 0 \text{ and/or } \theta_2 \neq 0.$$

Then, substitute $\beta_2 = \theta_2 - 2$ and $\beta_1 = \theta_1 + (\theta_2 - 2) = \theta_1 + \theta_2 - 2$ to get

$$\begin{aligned}
Y &= \beta_0 + \underbrace{(\theta_1 + \theta_2 - 2)}_{\beta_1} X_1 + \underbrace{(\theta_2 - 2)}_{\beta_2} X_2 + \beta_3 X_3 + U \\
&\Downarrow \\
\underbrace{Y + 2X_1 + 2X_2}_{Y^*} &= \beta_0 + \theta_1 X_1 + \theta_2 \underbrace{(X_2 + X_1)}_{X^*} + \beta_3 X_3 + U \\
&\Downarrow \\
Y^* &= \beta_0 + \theta_1 X_1 + \theta_2 X^* + \beta_3 X_3 + U \qquad (5)
\end{aligned}$$

We estimate model (5) by OLS and compute the test statistic

$$F = \frac{1}{2} \frac{\left(t_1^2 + t_2^2 - 2t_1 t_2 \hat{\rho}_{t_1,t_2}\right)}{1 - \hat{\rho}^2_{t_1,t_2}} \text{ with } t_j = \frac{\hat{\theta}_j}{SE\left(\hat{\theta}_j\right)},$$

and $\hat{\theta}_j$ is the OLS estimator of θ_j, $j = 1, 2$. Under H_0, F is aproximately distributed as a $\chi^2_2/2$, whose critical value at the 5% of significance is $\chi^2_{2,0.05}/2 = 3$. Thus, we reject H_0 if $F > 3$ for the F computed with the data.

ii. Define $\theta = \beta_1 + \beta_2 + \beta_3$. So we can write the hypothesis as

$$H_0 : \theta = 0 \text{ vs } H_1 : \theta < 0.$$

Then, substitute $\beta_1 = \theta - \beta_2 - \beta_3$ to get

$$\begin{aligned}
Y &= \beta_0 + \underbrace{(\theta - \beta_2 - \beta_3)}_{\beta_1} X_1 + \beta_2 X_2 + \beta_3 X_3 + U \\
&= \beta_0 + \theta X_1 + \beta_2 \underbrace{(X_2 - X_1)}_{=X_3^*} + \beta_3 \underbrace{(X_3 - X_1)}_{=X_3^*} + U \\
&= \beta_0 + \theta X_1 + \beta_2 X_2^* + \beta_3 X_3^* + U. \qquad (6)
\end{aligned}$$

We estimate model (6) by OLS and compute the statistic

$$t = \frac{\hat{\theta}}{SE\left(\hat{\theta}\right)},$$

where $\hat{\theta}$ is the OLS estimator in (6) under H_0, t is aproximately distributed as a standard normal, and H_0 is rejected when $t < -Z_{0.05} = -1.645$.

iii. Define $\theta = \beta_1 + \beta_2 + \beta_3 - 1$. So we can write the hypothesis as

$$H_0 : \theta = 0 \text{ vs } H_1 : \theta \neq 0.$$

Then, substitute $\beta_1 = \theta - \beta_2 - \beta_3 + 1$ to get

$$\begin{aligned} Y &= \beta_0 + \underbrace{(\theta - \beta_2 - \beta_3 + 1)}_{\beta_1} X_1 + \beta_2 X_2 + \beta_3 X_3 + U \\ &\Downarrow \\ \underbrace{Y - X_1}_{=Y^*} &= \beta_0 + \theta X_1 + \beta_2 \underbrace{(X_2 - X_1)}_{=X_2^*} + \beta_3 \underbrace{(X_3 - X_1)}_{=X_3^*} + U \\ &\Downarrow \\ Y^* &= \beta_0 + \theta X_1 + \beta_2 X_2^* + \beta_3 X_3^* + U. \end{aligned} \tag{7}$$

We estimate model (7) by OLS and compute the statistic

$$t = \frac{\hat{\theta}}{SE\left(\hat{\theta}\right)}, \tag{8}$$

where $\hat{\theta}$ is the OLS estimator in (6) under H_0, t is aproximately distributed as a standard normal, and H_0 is rejected when $|t| > Z_{0.025} = 1.96$.

b. (2/5) Explain, assuming that $\sigma^2(X_{1i}, X_{2i}, X_{3i}) = E(U_i^2)$, whether or not you could test the above hypotheses comparing either the coefficients of determination (R^2) or the sum of squared residuals (SSR) in the restricted and unrestricted models. Provide the corresponding test statistics in each case.

ANSWER:

Under homoskedasticity we can compare either the SSR or the R^2 in the transformed models. That is,

i. The unrestricted model is (5) and the restricted is

$$Y^* = \beta_0 + \beta_3 X_3 + U,$$

with corresponding sums of squares residuals SSR_u and SSR_r and coefficients of determination R_u^2, and R_r^2, respectively. We can use the test statistic

$$F = \frac{1}{2}\frac{R_u^2 - R_r^2}{(1 - R_u^2)/n} = \frac{1}{2}\frac{SSR_r - SSR_u}{SSR_u/n}.$$

The test is performed identically as in the heteroskedastic case.

ii. This test cannot be implemented with an F because the alternative is one sided.

iii. The unrestricted model is (7) and its restricted version is

$$Y^* = \beta_0 + \beta_2 X_2^* + \beta_3 X_3^* + U,$$

with corresponding sums of squares residuals SSR_u and SSR_r and coefficients of determination R_u^2, and R_r^2, respectively. We can use the test statistic

$$F = \frac{R_u^2 - R_r^2}{(1 - R_u^2)/n} = \frac{SSR_r - SSR_u}{SSR_u/n}.$$

Since the hypothesis is homogeneous we can use either the R^2 or the SSR. Thus, under H_0, F is approximately distributed as a χ_1^2, whose critical value at the 5% of significance is $\chi_{1,0.05}^2 = Z_{0.025}^2 = 1.96^2 = 3.84$. We reject H_0 when the observed F is bigger than 3.84.

In addition, extra credits will be given (say a 25% more in this part) if it is noticed that $F = t^2$, with t in (8) but with $SE\left(\hat{\theta}\right)$ imposing the homoskedasticity restriction.

3. (40%) In order to estimate the relationship between aggregate wages, W, aggregate profits, P, and aggregate income, Y, consider the following model,

$$\begin{aligned} W &= \beta_0 + \beta_1 Y + u, \qquad (9) \\ P &= \alpha_0 + \alpha_1 Y + \alpha_2 K + v, \\ Y &= W + P, \end{aligned}$$

where K is aggregate stock of capital and u, v are disturbance terms, assumed independent of each other. The third equation is an identity, all forms of income are classified as wages and profits. The coefficients are intended to be estimated using data from a sample of industrialized countries, with the variables measured on a per capita basis in a common currency. K may be assumed to be exogenous.

a. (2/5) Show that Y is endogenous in the structural equation (9).

ANSWER:

$$\begin{array}{l} W = \beta_0 + \beta_1 Y + u \\ + \\ P = \alpha_0 + \alpha_1 Y + \alpha_2 K + v \\ \hline Y = (\beta_0 + \alpha_0) + (\beta_1 + \alpha_1) Y + \alpha_2 K + (u + v) \end{array}$$

Thus,

$$Y = \frac{\beta_0 + \alpha_0}{1 - \beta_1 - \alpha_1} + \frac{\alpha_2}{1 - \beta_1 - \alpha_1} K + \frac{1}{1 - \beta_1 - \alpha_1} (u + v).$$

We must assume that $\beta_1 + \alpha_1 \neq 1$. Thus,

$$\begin{aligned} Cov(Y, u) &= \frac{1}{1 - \beta_1 - \alpha_1} \left[\alpha_2 \underbrace{Cov(K, u)}_{=0 \text{ because } K \text{ exog.}} + \underbrace{Cov(u, u)}_{=Var(u)} + \underbrace{Cov(v, u)}_{=0 \text{ (assumed)}} \right] \\ &= \frac{Var(u)}{1 - \beta_1 - \alpha_1} \neq 0. \end{aligned}$$

Therefore, Y is endogenous.

b. (1/5) Explain how to obtain the two stages least squares (2SLS) estimator of β_1.

ANSWER:

1st step: Estimate the reduced form equation

$$Y = \pi_0 + \pi_1 K + \varepsilon, \tag{10}$$

where ε is the reduced form error, by OLS, and compute the predicted values

$$\hat{Y}_i = \hat{\pi}_0 + \hat{\pi}_1 K_i, \; i = 1, ..., n.$$

2nd step: Estimate the structural form equation (9), where Y_i is substituted by $\hat{Y}_i$, the 2SLS is

$$\hat{\beta}_1 = \frac{\widehat{Cov}\left(W, \hat{Y}\right)}{\widehat{Var}\left(\hat{Y}\right)}.$$

c. (2/5) Show that β_1 is the ratio between the slopes in the reduced form equations of the endogenous variables in the structural equation (1), and provide the resulting estimator based on the OLS estimators of the slope parameters in the reduced forms (indirect least squares estimator). Show that this estimator is identical to the estimator in (b).

ANSWER: Consider the reduced form of W,

$$W = \gamma_0 + \gamma_1 K + w, \tag{11}$$

where w is the corresponding reduced form error. Now, substitute K

$$K = -\frac{\gamma_0}{\gamma_1} + \frac{1}{\gamma_1} W - \frac{1}{\gamma_1} w$$

into (9) to get

$$\begin{aligned} Y &= \pi_0 + \pi_1 \left(-\frac{\gamma_0}{\gamma_1} + \frac{1}{\gamma_1} W - \frac{1}{\gamma_1} w \right) + \varepsilon \\ &= \underbrace{\left(\pi_0 - \pi_1 \frac{\gamma_0}{\gamma_1} \right)}_{=\beta_0} + \underbrace{\frac{\pi_1}{\gamma_1}}_{=\beta_1} W + \underbrace{\left(\varepsilon - \frac{\pi_1}{\gamma_1} w \right)}_{=u} \end{aligned}$$

Now, once π_1 and γ_1 are estimated by OLS in (10) and (11), respectively, β_1 is estimated by

$$\begin{aligned}
\hat{\beta}_1^{2SLS} &= \frac{\widehat{Cov}\left(\hat{Y}, W\right)}{\widehat{Var}\left(\hat{Y}\right)}, \\
&= \frac{\widehat{Cov}\left(\hat{\pi}_0 + \hat{\pi}_1 K, W\right)}{\widehat{Var}\left(\hat{\pi}_0 + \hat{\pi}_1 K\right)} \\
&= \frac{\hat{\pi}_1}{\hat{\pi}_1^2} \frac{\widehat{Cov}\left(W, K\right)}{\widehat{Var}\left(K\right)} \\
&= \frac{1}{\hat{\pi}_1} \frac{\widehat{Cov}\left(W, K\right)}{\widehat{Var}\left(K\right)} \\
&= \frac{\hat{\gamma}_1}{\hat{\pi}_1}
\end{aligned}$$