ECMA 31000: Problem Set 6

Joe Hardwick

Due Nov 30 by 11:59PM

Question 1 Let $x = (1, x_1, x_2')'$, where x_1 is a scalar random variable and x_2 is a random vector. Suppose you observe an iid sample of $\{y_i, x_i\}_{i=1}^n$. Consider the model

$$y = \beta_0 + \beta_1 x_1 + x_2'\beta_2 + u; \qquad \mathrm{E}\left(u|x\right) = 0, Var\left(u|x\right) = \sigma^2.$$

a) Use the Frisch-Waugh-Lovell decomposition to find a formula for the OLS estimator of β_1. Call it $\hat{\beta}_1^{OLS}$.

b) Find the mean and variance of $\hat{\beta}_1^{OLS}$, conditional on $\{x_i\}_{i=1}^n$.

c) Suppose I omit the entire vector x_2 from the regression, and estimate

$$y = b_0 + b_1 x_1 + \epsilon,$$

using OLS. Suppose $\hat{b}_1^{OLS}$ is treated as an estimator of β_1. Derive the omitted variables bias of $\hat{b}_1^{OLS}$ conditional on $\{x_i\}_{i=1}^n$. Show that $\hat{b}_1^{OLS} \overset{p}{\nrightarrow} \beta_1$ in general.

d) (Inclusion of irrelevant variables). Now suppose $\beta_2 = 0$. In this case we say x_2 is "irrelevant", because $\mathrm{E}\left(y|x\right)$ does not depend on x_2. What is the bias of $\hat{b}_1^{OLS}$? Compare the variances of $\hat{b}_1^{OLS}$ and $\hat{\beta}_1^{OLS}$ conditional on $\{x_i\}_{i=1}^n$.

e) Construct the regression

$$x_1 = \gamma_0 + x_2'\gamma_1 + v; \qquad \mathrm{E}\left(x_2 v\right) = 0, \mathrm{E}\left(v\right) = 0.$$

Show that

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{Var\left(u\right)}{Var\left(v\right)}\right).$$

Also derive the asymptotic distribution of $\hat{b}_1$ under the assumption that $\beta_2 = 0$. Hint: If you use part a), show first that $M_{X_2}X_1 = M_{X_2}V$, where X_2 is a matrix containing the observations of $(1, x_2')'$, and X_1 is a column vector containing the observations of x_1.

f) Use your results in part (e) to explain why, if an irrelevant variable is included that is highly correlated with a relevant variable, it can increase the finite sample and asymptotic variance significantly. Use your answers to (c),(d) and (e) to argue that the choice to include regressors presents a tradeoff between bias and variance.

Question 2 Your goal is to estimate the average treatment effect of offering job training grants on firm productivity. You observe an iid sample of $\{y_i, d_i, x_{i2}\}_{i=1}^n$, where y_i is the productivity of firm i, x_{i2} is a vector of observable characteristics of firm i, and

$$d_i = \begin{cases} 1 & \text{if firm } i \text{ receives a job training grant,} \\ 0 & \text{otherwise.} \end{cases}$$

Define the potential outcomes y_{i0}, y_{i1} as

$$\begin{aligned} y_{i0} &= \text{scrap rate of firm } i \text{ without grant,} \\ y_{i1} &= \text{scrap rate of firm } i \text{ with grant.} \end{aligned}$$

You are able to completely randomize the allocation of grants to firms, so y_{i0}, y_{i1} are independent of d_i. Suppose $\mathrm{P}(d = 1) = p \in (0, 1)$. Suppose $\mathrm{E}(y_i^2) < \infty$.

a) Argue that without loss of generality, you can write

$$y_i = \beta_0 + \beta_1 d_i + u_i, \qquad \mathrm{E}(u_i | d_i) = 0,$$

where $\beta_1 = \mathrm{E}(Y_{i1} - Y_{i0})$ equals the average treatment effect in the population of firms.

Let $x = (1, d, x_2)$ and assume $\mathrm{E}(xx')$ exists and invertible. Suppose you add in the observable characteristics to your equation, and now you assume

$$y_i = b_0 + b_1 d_i + x_{i2}' b_2 + v_i; \qquad \mathrm{E}(x_i v_i) = 0. \tag{1}$$

Suppose also that d_i is independent of x_{i2} (this would not be the case if x_{i2} contained y_i, for example). Assume $\mathrm{E}(v^2 xx')$ exists.

b) Is it necessarily the case that $\mathrm{E}(u_i | d_i, x_{i2})$=0 or $\mathrm{E}(v_i | d_i, x_{i2}) = 0$?

c) Derive the bias of $\hat{b}_1^{OLS}$ conditional on $\{d_i, x_{i2}\}_{i=1}^n$, when considered as an estimator of β_1. Show that leaving out x_{i2} would yield an unconditionally unbiased estimator.

d) Despite your answer to c), argue that $b_1 = \beta_1$. Show explicitly where you use the independence assumption.

e) Argue that $\hat{b}_1^{OLS} \xrightarrow{p} \beta_1$.

f) Show that

$$\sqrt{n}\left(\hat{b}_1 - \beta_1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathrm{E}\left[(d_i - p)^2 (y_i - b_0 - b_1 d_i - x_{i2}' b_2)^2\right]}{p^2 (1-p)^2}\right).$$

g) Specialize your result in f) by assuming $\mathrm{E}(v | d, x) = 0$ and $Var(v | d, x) = Var(v)$. Argue that adding in covariates (as long as they are independent of d) weakly lowers the asymptotic variance. Does it lower the finite sample variance conditional on $\{d_i, x_i\}_{i=1}^n$?

Question 3 (Efficient estimation when the form of heteroskedasticity is known) Let $\{y_i, x_i\}_{i=1}^n$ be an iid sample of observations of y, x, where y is a scalar random variable and $x \in \mathbb{R}^k$, and suppose

$$y_i = x_i'\beta + u; \qquad \mathrm{E}(u_i|x_i) = 0, Var(u_i|x_i) = \sigma(x_i)^2.$$

In class we proved the Gauss-Markov theorem, in the case that $\sigma(x_i)^2 = \sigma^2$. In this question we find the best linear unbiased estimator when the function $\sigma(\cdot)$ is known and $\sigma^2(x) > 0$ for all values of x. Stack the observations to form a matrix representation of the model $Y = X\beta + U$. Let

$$Var(U|X) = \begin{pmatrix} \sigma(x_1)^2 & 0 & 0 & 0 \\ 0 & \sigma(x_2)^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma(x_n)^2 \end{pmatrix} := \Omega.$$

a) Pre-multiply this model by $\Omega^{-1/2}$ to obtain a transformed model

$$Y^* = X^*\bar{\beta} + U^*.$$

What are Y^*, X^*, U^* and $\bar{\beta}$? What is the variance covariance matrix of the error term in your newly transformed model?

b) What is the Best Linear Unbiased Estimator of $\bar{\beta}$ in your newly transformed model, conditional on X^*?

c) What is the Best Linear Unbiased Estimator of β in your original model, conditional on X? Hint: Show that the estimator from part b) is a linear and unbiased estimator of β.

Question 4 (Ridge regression) Let $\{y_i, x_i\}_{i=1}^n$ be an iid sample of observations of y, x, where y is a scalar random variable and $x \in \mathbb{R}^k$, and suppose

$$y_i = x_i'\beta + u; \qquad \mathrm{E}(u_i|x_i) = 0, Var(u_i|x_i) = \sigma^2.$$

Suppose $\mathrm{E}(xx') < \infty$ and is invertible. Suppose you use the "ridge regression" estimator of β:

$$\hat{\beta} = \left(\sum_{i=1}^n x_i x_i' + \lambda I_k\right)^{-1} \sum_{i=1}^n x_i y_i,$$

where $\lambda > 0$ is a constant and I_k is the $(k \times k)$ identity matrix.

a) Show that

$$\sum_{i=1}^n x_i x_i' + \lambda I_k$$

is always invertible.

b) Is $\hat{\beta}$ a consistent estimator of β? Is it unbiased?

c) Suppose further that $Var(u|x) = \sigma^2$. Derive the asymptotic distribution of $\hat{\beta}$.

Question 5 (Computational Question) Download the data PS6.csv from Canvas. The dataset contains data from a simulated experiment. Do not use packages designed for regression except to check your answers. You observe outcomes Y_i, treatment status D_i, and characteristics X_i for each of $n = 1643$ individuals. You are interested in studying the average treatment effect $\mathrm{E}(Y_{i1} - Y_{i0})$. You may assume that D_i is randomly assigned and independent of X_i. Consider the following regressions:

$$Y_i = \beta_0 + \beta_1 D_i + U_i; \quad \mathrm{E}(U_i) = \mathrm{E}(D_i U_i) = 0;$$
$$Y_i = \gamma_0 + \gamma_1 D_i + X_i' \gamma_2 + V_i; \quad \mathrm{E}(V_i) = \mathrm{E}(D_i V_i) = 0, \mathrm{E}(X_i V_i) = 0.$$

a) Compute the OLS estimators of (β_0, β_1) and $(\gamma_0, \gamma_1, \gamma_2)$. Are both estimators consistent for the average treatment effect? Are they unbiased?

b) Compute a consistent estimate of the asymptotic covariance matrix of the OLS estimates of (β_0, β_1) assuming homoskedasticity. Do the same for $(\gamma_0, \gamma_1, \gamma_2)$. What do you notice about the estimated asymptotic variance of $\hat{\beta}_1$ versus that of $\hat{\gamma}_1$? Explain in the context of Questions 1 and 2.

c) Drop the assumption of homoskedasticity. Now repeat part b) with an estimate of the asymptotic covariance matrix that is robust to heteroskedasticity. Do the estimates seem to be much different?