Economics 675: Applied Microeconometrics – Fall 2018

Assignment 1 – Due date: Mon 24-Sep

Last updated: September 6, 2018

Contents

Guidelines:

- You may work in (small) groups while solving this assignment.
- Submit individual solutions via `http://canvas.umich.edu` in one PDF file collecting everything (e.g., derivations, figures, tables, computer code).
- Computer code should be done in both `Stata` and `R`. If the numerical results do not agree across the two statistical software platforms, you must explain why that is the case.
- Start each question on a separate page. Always add a reference section if you cite other sources.
- Clearly label all tables and figures, and always include a brief footnote with useful information.
- Always attach your computer code as an appendix, with annotations/comments as appropriate.
- Please provide as much detail as possible in your answers, both analytical and empirical.

1 Question 1: Simple Linear Regression with Measurement Error

Consider first cross-section data. Let $\{(x_i, \varepsilon_i, u_i) : i = 1, 2, \cdots, n\}$ be i.i.d., satisfying

$$
\begin{aligned}
y_i &= x_i\beta + \varepsilon_i, \qquad \mathbb{E}[x_i] = 0 = \mathbb{E}[\varepsilon_i], \quad \mathbb{V}[x_i] = \sigma_x^2, \quad \mathbb{V}[\varepsilon_i] = \sigma_\varepsilon^2, \quad \mathbb{E}[x_i\varepsilon_i] = 0, \\
\tilde{x}_i &= x_i + u_i, \qquad \mathbb{E}[u_i] = 0, \quad \mathbb{V}[u_i] = \sigma_u^2, \quad \mathbb{E}[x_i u_i] = 0, \quad \mathbb{E}[\varepsilon_i u_i] = 0,
\end{aligned}
$$

where $(\beta, \sigma_x^2, \sigma_\varepsilon^2, \sigma_u^2)' \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$. We assume that *only* $\{(y_i, \tilde{x}_i) : i = 1, 2, \cdots, n\}$ is observed; that is, x_i is *not* observed. Let $\mathbf{y} = (y_1, \cdots, y_n)'$, $\mathbf{x} = (x_1, \cdots, x_n)'$, $\tilde{\mathbf{x}} = (\tilde{x}_1, \cdots, \tilde{x}_n)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)'$ and $\mathbf{u} = (u_1, \cdots, u_n)'$.

Please answer the following questions, providing regularity conditions as necessary.

1. (**OLS estimator**) Let $\hat{\beta}_{\texttt{LS}} = (\tilde{\mathbf{x}}'\tilde{\mathbf{x}})^{-1}\tilde{\mathbf{x}}'\mathbf{y}$. Show that $\hat{\beta}_{\texttt{LS}} \to_p \lambda\, \beta$, and characterize the (attenuation) factor $\lambda \in \mathbb{R}$. Is $\hat{\beta}_{\texttt{LS}}$ biased upwards or downwards (for β)?

2. (**Standard Errors**) Let $\hat{\sigma}_{\boldsymbol{\varepsilon}}^2 = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}/n$ with $\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \tilde{\mathbf{x}}\hat{\beta}_{\texttt{LS}}$. Show that

$$\hat{\sigma}_\varepsilon^2 \to_p \sigma_\varepsilon^2 + (1-\lambda)^2\beta^2\sigma_x^2 + \lambda^2\beta^2\sigma_u^2.$$

   Is $\hat{\sigma}_\varepsilon^2$ biased upwards or downwards (for σ_ε^2)?

   Compute the probability limit of $\hat{\sigma}_\varepsilon^2(\tilde{\mathbf{x}}'\tilde{\mathbf{x}}/n)^{-1}$. Is it upward or downward biased (for $\sigma_\varepsilon^2/\sigma_x^2$)?

3. (**t-test**) Show that

$$\frac{\hat{\beta}_{\texttt{LS}}}{\sqrt{\hat{\sigma}_\varepsilon^2(\tilde{\mathbf{x}}'\tilde{\mathbf{x}}/n)^{-1}}} \to_p \frac{\sqrt{\lambda}\beta}{\sqrt{\sigma_\varepsilon^2/\sigma_x^2 + (1-\lambda)\beta^2}}.$$

   Is the usual t-test for $H_0 : \beta = \beta_0$ biased upwards or downwards?

4. (**Second measurement, Consistency**) Suppose there exists a second, independent (possibly noisy) i.i.d. observed measurement $\{\check{x}_i : i = 1, 2, \cdots, n\}$ for each x_i satisfying:

$$\mathbb{E}[\check{x}_i x_i] \neq 0, \qquad \mathbb{E}[\check{x}_i\varepsilon_i] = 0, \qquad \mathbb{E}[\check{x}_i u_i] = 0.$$

   Construct a consistent estimator of β using observed data, denoted $\hat{\beta}_{\texttt{IV}}$.

5. (**Second measurement, Distribution**) Suppose the conditions of part 4 hold. Imposing appropriate assumptions (e.g., moment conditions), show that $\sqrt{n}(\hat{\beta}_{\texttt{IV}} - \beta) \to_d \mathcal{N}(0, V_{\texttt{IV}})$ and characterize the asymptotic variance $V_{\texttt{IV}}$.

6. (**Second measurement, Inference**) Suppose the conditions of parts 4 and 5 hold. Construct a 95% asymptotically valid confidence interval for β robust to conditional heteroskedasticity.

7. (**Validation sample, Consistency**) Suppose there exists another estimator $\check{\sigma}_x^2$ satisfying $\check{\sigma}_x^2 \to_p \sigma_x^2$. (For example, this may be obtained using a validation output/study.) Construct a consistent estimator of β, denoted $\hat{\beta}_{\texttt{VS}}$.

8. (**Validation sample, Distribution**) Suppose the conditions of part 7 hold, and assume that

$$\sqrt{n}\left(\begin{bmatrix} \hat{\beta}_{\mathtt{LS}} \\ \tilde{\mathbf{x}}'\tilde{\mathbf{x}}/n \\ \check{\sigma}_x^2 \end{bmatrix} - \begin{bmatrix} \lambda\beta \\ \sigma_x^2 + \sigma_u^2 \\ \sigma_x^2 \end{bmatrix}\right) \to_d \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right), \qquad \boldsymbol{\Sigma} \in \mathbb{R}^{3\times 3}$$

Imposing appropriate assumptions, show that $\sqrt{n}(\hat{\beta}_{\mathtt{VS}} - \beta) \to_d \mathcal{N}(0, V_{\mathtt{VS}})$ and characterize the asymptotic variance $V_{\mathtt{VS}}$.

9. (**Validation sample, Inference**) Suppose the conditions of parts 7 and 8 hold, and assume there exists an estimator $\hat{\boldsymbol{\Sigma}}$ satisfying $\hat{\boldsymbol{\Sigma}} \to_p \boldsymbol{\Sigma}$. Construct a 95% asymptotically valid confidence interval for β.

Consider now panel data. Let $\{(x_{i1}, x_{i2}, \varepsilon_{i1}, \varepsilon_{i2}, u_{i1}, u_{i2}) : i = 1, 2, \cdots, n\}$ be i.i.d., time dependence is allowed, satisfying

$$\begin{aligned} y_{it} &= x_{it}\beta + \mu_i + \varepsilon_{it}, \qquad \mathbb{E}[x_{it}] = 0 = \mathbb{E}[\varepsilon_{it}], \quad \mathbb{V}[x_{it}] = \sigma_{x_t}^2, \quad \mathbb{V}[\varepsilon_{it}] = \sigma_{\varepsilon_t}^2, \quad \mathbb{E}[x_{it}\varepsilon_{is}] = 0, \\ \tilde{x}_{it} &= x_{it} + u_{it}, \qquad \mathbb{E}[u_{it}] = 0, \quad \mathbb{V}[u_{it}] = \sigma_{u_t}^2, \quad \mathbb{E}[x_{it}u_{is}] = 0, \quad \mathbb{E}[\varepsilon_{it}u_{is}] = 0, \end{aligned}$$

for all $t, s = 1, 2$ and where $(\beta, \sigma_{x_t}^2, \sigma_{\varepsilon_t}^2, \sigma_{u_t}^2)' \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$, $t = 1, 2$. We assume that *only* $\{(y_{i1}, \tilde{x}_{i1}, y_{i2}, \tilde{x}_{i2}) : i = 1, 2, \cdots, n\}$ is observed; that is, x_{i1} and x_{i2} are *not* observed. Let $\mathbf{y}_i = (y_{i1}, y_{i2})'$, $\mathbf{x}_i = (x_{i1}, x_{i2})'$, $\tilde{\mathbf{x}}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2})'$, $\mathbf{u}_i = (u_{i1}, u_{i2})'$, and

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}, \qquad \tilde{\mathbf{X}} = \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \vdots \\ \tilde{\mathbf{x}}_n \end{bmatrix}, \qquad \boldsymbol{\iota} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Please answer the following questions, providing regularity conditions as necessary.

10. (**FE estimator, Consistency**) Let $\hat{\beta}_{\mathtt{FE}} = (\tilde{\mathbf{X}}(\mathbf{I}_{2n} - \mathbf{P_D})\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}(\mathbf{I}_{2n} - \mathbf{P_D})\mathbf{y}$, with $\mathbf{P_D} = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ and $\mathbf{D} = \mathbf{I}_n \otimes \boldsymbol{\iota}$. Show that

$$\hat{\beta}_{\mathtt{FE}} \to_p \gamma\, \beta, \qquad \gamma = \frac{\sigma_{\Delta x}^2}{\sigma_{\Delta x}^2 + \sigma_{\Delta u}^2}, \qquad \sigma_{\Delta x}^2 = \mathbb{V}[x_{i2} - x_{i1}], \qquad \sigma_{\Delta u}^2 = \mathbb{V}[u_{i2} - u_{i1}].$$

Is $\hat{\beta}_{\mathtt{FE}}$ biased upwards or downwards (for β)?

11. (**FE estimator, Time dependence**) Suppose that $\{x_{it} : t \geq 1\}$ and $\{u_{it} : t \geq 1\}$ are covariance stationary. Show that

$$\gamma = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2\frac{1-\rho_u}{1-\rho_x}}, \qquad \rho_u = \frac{\mathbb{C}[u_{i2}, u_{i1}]}{\sigma_u^2}, \quad \sigma_u^2 = \sigma_{u_t}^2, \qquad \rho_x = \frac{\mathbb{C}[x_{i2}, x_{i1}]}{\sigma_x^2}, \quad \sigma_x^2 = \sigma_{x_t}^2.$$

12. (**FE estimator, Implications**) Using the results above, consider the implications of measurement error for panel data. In particular, what happens to γ when $\rho_x \approx 1$ and $\rho_u \approx 0$? Can you provide intuition for this result? What are its implications for empirical work?

2 Question 2: Implementing Least-Squares Estimators

Let $(y_i, \mathbf{x}_i')'$, $i = 1, 2, \cdots, n$, be a random sample from $(y, \mathbf{x}')'$ with $y \in \mathbb{R}$ and $\mathbf{x} \in \mathbb{R}^d$. The classical linear model is $y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ or, in matrix form, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\mathbf{y} = [y_1, y_2, \cdots, y_n]'$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]'$, and $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n]'$. Let $\mathbf{P}_\mathbf{A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ be the usual projection matrix (onto the column space of the matrix $\mathbf{A}$). You may impose additional assumptions (e.g., moment conditions) so that the usual Law of Large Numbers (LLNs) and/or Central Limit Theorems (CLTs) hold.

Consider the family of estimators:

$$\hat{\boldsymbol{\beta}}(\mathbf{W}) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^d} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^d} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2_\mathbf{W},$$

with $\mathbf{W}$ being positive definite and $\|\mathbf{W} - \bar{\mathbf{W}}\| \to_p 0$ for some non-random positive definite matrix $\bar{\mathbf{W}}$. For example, $\hat{\boldsymbol{\beta}}_{\texttt{OLS}} = \hat{\boldsymbol{\beta}}(\mathbf{I}_n)$, $\hat{\boldsymbol{\beta}}_{\texttt{GLS}} = \hat{\boldsymbol{\beta}}(\boldsymbol{\Omega}^{-1})$ for $\mathbb{V}[\mathbf{y}|\mathbf{X}] = \boldsymbol{\Omega}$, and $\hat{\boldsymbol{\beta}}_{\texttt{2SLS}} = \hat{\boldsymbol{\beta}}(\mathbf{P}_\mathbf{Z})$ for $\mathbf{Z}$ a conformable matrix of "instruments", among other possibilities. Let $\mathbf{e}_j \in \mathbb{R}^d$ denote the j-th unit vector.

Please answer the following questions, providing regularity conditions as necessary.

1. Show that $\hat{\boldsymbol{\beta}}(\mathbf{W}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$.

2. Show that $\sqrt{n}(\hat{\boldsymbol{\beta}}(\mathbf{W}) - \boldsymbol{\beta}) \to_d \mathcal{N}(\mathbf{0}, \mathbf{V}(\mathbf{W}))$, and give the exact form of the asymptotic variance-covariance matrix $\mathbf{V}(\mathbf{W})$. What happens if $\mathbb{V}[\mathbf{y}|\mathbf{X}, \mathbf{W}] = \sigma^2\mathbf{I}_n$?

3. Propose a consistent variance-covariance estimator (VCE) of $\mathbf{V}(\mathbf{W})$, denoted $\hat{\mathbf{V}}(\mathbf{W})$, and assuming conditional heteroskedasticity of unknown form.

4. Set $\mathbf{W} = \mathbf{I}_d$. Implement the following inference procedures in your preferred matrix-oriented language. Use both (i) a symmetric inverse and (ii) a Cholesky inverse; do the results change?

   (a) Point estimators: $\hat{\boldsymbol{\beta}}(\mathbf{W}) \in \mathbb{R}^d$ and $\hat{\mathbf{V}}(\mathbf{W}) \in \mathbb{R}^{d\times d}$.

   (b) t-test for $\mathsf{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$:

   $$t_j = \frac{\mathbf{e}_j'\hat{\boldsymbol{\beta}}(\mathbf{W}) - \mathbf{e}_j'\boldsymbol{\beta}_0}{\sqrt{\mathbf{e}_j'\hat{\mathbf{V}}(\mathbf{W})\mathbf{e}_j/n}}, \qquad j = 1, 2, \cdots, d.$$

   (c) p-value (or significance level) under $\mathsf{H}_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ using the statistic t_j ($j = 1, 2, \cdots, d$).

   (d) Symmetric, Wald-type α-percent Confidence interval (for $\boldsymbol{\beta}_0$):

   $$I_j = \left[\mathbf{e}_j'\hat{\boldsymbol{\beta}}(\mathbf{W}) - \Phi^{-1}_{\alpha/2}\sqrt{\mathbf{e}_j'\hat{\mathbf{V}}(\mathbf{W})\mathbf{e}_j/n}\ ,\ \mathbf{e}_1'\hat{\boldsymbol{\beta}}(\mathbf{W}) + \Phi^{-1}_{\alpha/2}\sqrt{\mathbf{e}_j'\hat{\mathbf{V}}(\mathbf{W})\mathbf{e}_j/n}\right], \qquad j = 1, 2, \cdots, d.$$

5. Consider the `LaLonde_1986.csv` data described in 4. For $i = 1, 2, \cdots, n$, with $n = 445$, let $y_i = \texttt{earn78}_i$ and $\mathbf{x}_i' = (1, \texttt{treat}_i, \texttt{black}_i, \texttt{age}_i, \texttt{educ}_i, \texttt{educ}_i^2, \texttt{earn74}_i, \texttt{black}_i \cdot \texttt{earn74}_i, \texttt{u74}, \texttt{u75})$; $d = 10$.

   (a) Report in one table the OLS point estimator, standard error, t-statistic, p-value and 95% confidence interval for each element of $\boldsymbol{\beta}_0$ using your matrix-based implementations from part 4.

   (b) Repeat part (a) but now using a function provided by the statistical software you used (e.g., `regress` in `Stata` and `lm(·)` in `R`). Do the number coincide? If not, explain why.

3 Question 3: Analysis of Experiments

Consider again the `LaLonde_1986.csv` data described in 4. For $i = 1, 2, \cdots, n$, with $n = 445$, let $Y_i = \texttt{earn78}_i \in \mathbb{R}_+$ and $T_i = \texttt{treat}_i \in \{0, 1\}$. Employ a causal inference framework where potential outcomes are non-random, and the only source of randomness is the treatment status for each unit i (i.e., the probability law for the random variable T_i). Assume throughout that $\mathbf{T} = (T_1, T_2, \cdots, T_n)'$ follows a fixed margins (or complete randomization) distribution, with N_1 treated units and $n - N_1$ control units.

1. **Neyman's approach**. Consider the difference-in-means statistic:

$$T_{\texttt{DM}} = \bar{Y}_1 - \bar{Y}_0, \qquad \bar{Y}_t = \frac{1}{N_t}\sum_{i=1}^{n} D_i(t) Y_i, \qquad N_t = \sum_{i=1}^{n} D_i(t), \qquad D_i(t) = \mathbf{1}(T_i = t), \qquad t = 0, 1.$$

   (a) Show that

$$\mathbb{E}[T_{\texttt{DM}}] = \frac{1}{n}\sum_{i=1}^{n} Y_i(1) - \frac{1}{n}\sum_{i=1}^{n} Y_i(0) =: \tau_{\texttt{ATE}}.$$

   Compute an unbiased estimator of the average treatment effect using the `LaLonde_1986.csv` data.

   (b) It can be shown that, under appropriate regularity conditions,

$$\frac{T_{\texttt{DM}} - \tau_{\texttt{ATE}}}{\sqrt{\mathbb{V}[T_{\texttt{DM}}]}} \to_d \mathcal{N}(0,1) \quad \text{and} \quad \mathbb{V}[T_{\texttt{DM}}] \le \frac{\bar{S}_0^2}{N_0} + \frac{\bar{S}_1^2}{N_1}, \qquad \bar{S}_t^2 = \frac{1}{N_t - 1}\sum_{i=1}^{n} D_i(t)(Y_i - \bar{Y}_t)^2, \qquad t = 0, 1.$$

   Construct an asymptotically conservative 95% confidence interval for the average treatment effect using the `LaLonde_1986.csv` data.

2. **Fisher's approach**. Consider both the difference-in-means test statistic ($T_{\texttt{DM}}$) and the Kolgomorov-Smirnov test statistic (denoted by $T_{\texttt{KS}}$).

   (a) Report (an approximation to) the significant level, or p-value, for the sharp null hypothesis of no treatment effect using the `LaLonde_1986.csv` data.

   (b) Using a constant treatment effect model (and SUTVA) construct a finite-sample valid 95% confidence interval for the treatment effect using the `LaLonde_1986.csv` data.

3. **Power calculations**. In lecture, we discussed power calculations using large sample approximations in a super population model. Here we consider the large sample approximations in part 1 (Neyman's approach) instead.

   (a) Using the results in part 1, and the `LaLonde_1986.csv` data, derive the (conservative) power function for the testing problem: $\mathsf{H}_0 : \tau_{\texttt{ATE}} = \tau_0$ vs. $\mathsf{H}_1 : \tau_{\texttt{ATE}} \neq \tau_0$.
   Plot the power function for $\tau_0 = 0$.

   (b) Suppose you would like to conduct a follow-up experiment. Using the `LaLonde_1986.csv` data, calculate the minimum sample size needed to detect a $1,000$ increase in earnings with a power of 0.80, assuming participants will be assigned to the treatment group with probability 2/3.

4 Appendix: `LaLonde_1986.csv` Data Description

In the famous study, LaLonde (1986) examined the impact of the National Supported Work (NSW) Demonstration on the post-training income. The data was from a randomized control experiment, hence treatment effect could be easily estimated.

The program was a federally and privately funded program implemented in the mid-1970s to provide work experience for a period of 6-18 months to individuals who had faced economic and social problems prior to enrollment in the program. Candidates eligible for the NSW were randomized into the program between March 1975 and July 1977.

The file `LaLonde_1986.csv` contains the data used in Dehejia and Wahba (1999), which is a subset of the initial LaLonde (1986), where earnings in 1974 is available (number of observations: 445).

Variable	Mean	St.Dev	Min	Max	Description
age	25.37	7.10	17	55	age, measured in years
educ	10.20	1.79	3	16	educational attainment, measured in years
black	.833	.373	0	1	dummy, 1 =black
hisp	.088	.283	0	1	dummy, 1 =hispanic
married	.169	.375	0	1	dummy, 1 =married
nodegr	.782	.413	0	1	dummy, 1 =high school dropouts (i.e. $educ >= 12$)
u74	.733	.443	0	1	unemployed in 1974 (i.e. $earn74 = 0$)
u75	.649	.478	0	1	unemployed in 1975 (i.e. $earn75 = 0$)
treat	.416	.493	0	1	treatment indicator, 1 =participated in the training
earn74	2102.3	5363.6	0	39570.7	earnings 13-24 months prior to the training, measured in $
earn75	1377.1	3151.0	0	25142.2	earnings in 1975, measured in $
earn78	5300.8	6631.5	0	60307.9	earnings in 1978 (post-program outcome), measured in $

References

Dehejia, R. H., and S. Wahba (1999): "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94(448), 1053–1062.

LaLonde, R. J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76(4), 604–620.