Instructor: **Anna Houštecká / Yaroslav Korobka** Updated: July 31, 2025
Course: **Econometrics II**

1 Panel data models

Problem 1

Suppose that the random effects model $y_{it} = x_{it}'\beta + \eta_i + v_{it}$ is to be estimated with a panel in which the groups have different numbers of observations. Let T_i be the number of observations in group i. Show that the pooled least squares estimator is unbiased and consistent despite this complication.

Solution 1

The model is equivalent to

$$y_i = X_i\beta + v_i + \eta_i\iota, \quad y_i \in \mathbb{R}^{T_i}, \quad X_i \in \mathbb{R}^{T_i \times K}, \quad v_i \in \mathbb{R}^{T_i}, \quad \iota := (1, \ldots, 1)' \in \mathbb{R}^{T_i}, \quad i = 1, \ldots, n,$$

and given the random effects model assumption, $\mathbb{E}[\eta_i x_{it}] = 0$. The pooled OLS estimator of β is

$$\hat{\beta} = \left(\sum_{i=1}^{n} X_i'X_i\right)^{-1} \sum_{i=1}^{n} X_i'y_i$$

given that $\sum_{i=1}^{n} X_i'X_i$ is invertible. To show the bias, rewrite

$$\begin{aligned}
\hat{\beta} &= \left(\sum_{i=1}^{n} X_i'X_i\right)^{-1} \sum_{i=1}^{n} X_i' \left(X_i\beta + v_i + \eta_i\iota\right) \\
&= \beta + \left(\sum_{i=1}^{n} X_i'X_i\right)^{-1} \sum_{i=1}^{n} X_i' \left(v_i + \eta_i\iota\right) \\
&= \beta + \left(\sum_{i=1}^{n} X_i'X_i\right)^{-1} \sum_{i=1}^{n} X_i'u_i
\end{aligned}$$

with $u_i := v_i + \eta_i\iota$. Hence, the bias $\mathbb{E}[\hat{\beta} - \beta | X_i]$ is zero if $\mathbb{E}[X_i'u_i|X_i] = X_i'\mathbb{E}[u_i|X_i] = X_i'\left(\mathbb{E}[v_i|X_i] + \mathbb{E}[\eta_i|X_i]\right) = 0$. It holds because the first expectation is zero by the i.i.d. independent mean zero errors v_{it}, and the second expectation is zero by the random effects assumption and the law of iterated expectations.

To show consistency, rewrite

$$\hat{\beta} - \beta = \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} X_i'u_i\right).$$

As $n \to \infty$, using the weak law of large numbers and Slutsky's theorem, we have that

$$\hat{\beta} - \beta \xrightarrow{p} Q^{-1}\mathbb{E}[X_i'u_i],$$

where $\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} X_i'X_i = \mathbb{E}[X_i'X_i] := Q$ is a non-deficient matrix with full rank. Under the random effects assumption and arguments as above, we have that $\mathbb{E}[X_i'u_i] = 0$. Hence, the estimator is consistent.

Problem 2

Consider $y_{it} = x_{it}'\beta + \eta_i + v_{it}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$, where $v_{it} \sim \mathcal{N}(0, \sigma^2)$ and $\beta = 0$. Write out the likelihood for estimating η_i and σ^2, and show that the MLE estimator $\hat{\sigma}^2$ is biased when $T < \infty$.

Solution 2

From the setup, it implies that $y_{it} \sim \mathcal{N}(\eta_i, \sigma^2)$. The individual log-likelihood for each i (across T) is then

$$\log \mathcal{L}(y_{it}|\eta_i, \sigma^2) = C_0 - \frac{T}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_{it} - \eta_i)^2$$

where C_0 is some constant independent of η_i and σ^2. The ML estimator of η_i is the solution to the equation

$$\frac{\partial \log \mathcal{L}(y_{it}|\eta_i, \sigma^2)}{\partial \eta_i} = \sum_{t=1}^{T}(y_{it} - \hat{\eta}_i) = 0,$$

which is $\hat{\eta}_i = T^{-1}\sum_{t=1}^{T} y_{it} := \bar{y}_i$.

To estimate σ^2, we use the joint log-likelihood (across i and T),

$$\log \mathcal{L}(y_{it}|\eta_i, \sigma^2) = C_1 - \frac{NT}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \eta_i)^2$$

where C_1 is some constant independent of η_i and σ^2. The ML estimator of σ^2 is the solution to the equation

$$\frac{\partial \log \mathcal{L}(y_{it}|\eta_i, \sigma^2)}{\partial \sigma^2} = -\frac{NT}{2}\cdot\frac{1}{\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \hat{\eta}_i)^2 = 0.$$

Substituting for $\hat{\eta}_i$ and rearranging, we have

$$\hat{\sigma}^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_i)^2.$$

Expectation of the estimator is

$$\begin{aligned}
\mathbb{E}[\hat{\sigma}^2] &= \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}(y_{it} - \bar{y}_i)^2\right] \\
&= \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} y_{it}^2 - \frac{2}{T}\sum_{t=1}^{T} y_{it}\bar{y}_i + \frac{1}{T}\sum_{t=1}^{T}\bar{y}_i^2\right] \\
&= \sigma^2 - \frac{2}{T}\sum_{t=1}^{T}\mathbb{E}[y_{it}\bar{y}_i] + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\bar{y}_i^2] \\
&= \sigma^2 - \frac{2}{T}\sigma^2 + \frac{1}{T}\sigma^2 = \sigma^2 - \frac{\sigma^2}{T},
\end{aligned}$$

which is not equal to σ^2 unless $T \to \infty$.

Problem 3

Consider $y_{it} = \mathbb{1}\{x_{it}\beta + \eta_i + v_{it} \geq 0\}$, where the errors v_{it} have the logisic cdf. Consider $T = 2$, $x_{i1} = 0$ and $x_{i2} = 1$, and show that the sufficient statistic for η_i is $y_{i1} + y_{i2} = 1$, i.e. conditioning on $y_{i1} + y_{i2} = 1$ implies that the MLE does not depend on η_i.

Solution 3

The log-likelihood function for two periods is given by

$$\begin{aligned}\log \mathcal{L}(y_i|x_i,\beta,\eta_i) &= y_{i1}\log\Lambda\left(x_{i1}\beta+\eta_i\right)+(1-y_{i1})\log\left(1-\Lambda(x_{i1}\beta+\eta_i)\right)\\ &\quad + y_{i2}\log\Lambda\left(x_{i2}\beta+\eta_i\right)+(1-y_{i2})\log\left(1-\Lambda(x_{i2}\beta+\eta_i)\right),\end{aligned}$$

where $\Lambda(z)=1/(1+e^{-z})$, and given known values of the covariates,

$$\log \mathcal{L}(y_i|\beta,\eta_i) = y_{i1}\log\Lambda\left(\eta_i\right)+(1-y_{i1})\log\left(1-\Lambda(\eta_i)\right)+y_{i2}\log\Lambda\left(\beta+\eta_i\right)+(1-y_{i2})\log\left(1-\Lambda(\beta+\eta_i)\right).$$

Taking the first derivative w.r.t. η_i, and using $\Lambda(z)'/\Lambda(z)=(1-\Lambda(z))$, we have

$$\frac{\partial \log \mathcal{L}(y_i|\beta,\eta_i)}{\partial \eta_i} = y_{i1}(1-\Lambda(\hat{\eta}_i))-(1-y_{i1})\Lambda(\hat{\eta}_i)+y_{i2}(1-\Lambda(\hat{\eta}_i+\beta))-(1-y_{i2})\Lambda(\hat{\eta}_i+\beta)=0,$$

which implies

$$y_{i1}+y_{i2}=\Lambda(\hat{\eta}_i)+\Lambda(\hat{\eta}_i+\beta).$$

Now, we discuss three cases:

1. if $y_{i1}+y_{i2}=0$, $\hat{\eta}_i=-\infty$,
2. if $y_{i1}+y_{i2}=2$, $\hat{\eta}_i=\infty$,
3. if $y_{i1}+y_{i2}=1$, $-2\hat{\eta}_i=\beta$, and $\hat{\eta}_i=-\beta/2$.

Hence, in the case 3., it is possible to identify η_i from the estimate of β only. It implies that conditional on $\zeta_i := y_{i1}+y_{i2}=1$, the log-likelihood is independent on η_i making ζ_i a sufficient statistic.

Problem 4

Derive the bias of the OLS estimator for α in a dynamic panel of the form $y_{it}=\alpha y_{it-1}+\eta_i+v_{it}$. Are there any conditions on α that should hold for the estimator to be well-defined?

Solution 4

First, rewrite the model in recursive form,

$$\begin{aligned}y_{it} &= \alpha\left(\alpha y_{it-2}+\eta_i+v_{it-1}\right)+\eta_i+v_{it}\\ &= \alpha^2\left(\alpha y_{it-3}+\eta_i+v_{it-2}\right)+\alpha\eta_i+\alpha v_{it-1}+\eta_i+v_{it}\\ &= \dots\\ &= \alpha^t y_0+\left(\sum_{s=0}^{t-1}\alpha^s\right)\eta_i+\sum_{s=0}^{t-1}\alpha^s v_{it-s}.\end{aligned}$$

The bias of the OLS estimator is driven by $\mathbb{E}[y_{it-1}u_{it}]$, where $u_{it}:=\eta_i+v_{it}$. Using the property of the geometric series, $\sum_{k=0}^{n-1}r^k=\frac{1-r^n}{1-r}$ for $|r|<1$, we have

$$\begin{aligned}y_{it-1} &= \alpha^{t-1}y_0+\left(\sum_{s=0}^{t-2}\alpha^s\right)\eta_i+\sum_{s=0}^{t-2}\alpha^s v_{it-s}\\ &= \alpha^{t-1}y_0+\left(\frac{1-\alpha^{t-1}}{1-\alpha}\right)\eta_i+\sum_{s=0}^{t-2}\alpha^s v_{it-s}.\end{aligned}$$

Now taking expectations,

$$\begin{aligned}\mathbb{E}[y_{it-1}\eta_i] &= \mathbb{E}\left[\alpha^{t-1}y_0\eta_i+\left(\frac{1-\alpha^{t-1}}{1-\alpha}\right)\eta_i^2+\sum_{s=0}^{t-2}\alpha^s v_{it-s}\eta_i\right]\\ &= \alpha^{t-1}\mathbb{E}[y_0\eta_i]+\left(\frac{1-\alpha^{t-1}}{1-\alpha}\right)\sigma_\eta^2\neq 0.\end{aligned}$$

Because we have that $\mathbb{E}[y_{it-1}u_{it}]=\mathbb{E}[y_{it-1}\eta_i]+\mathbb{E}[y_{it-1}v_{it}]$, the bias is nonzero even if $\mathbb{E}[y_{it-1}v_{it}]=0$.

Problem 5

Consider the panel AR(1) model with individual effects,

$$y_{it} = \alpha y_{it-1} + \eta_i + v_{it}$$

where $\eta_i \sim$ i.i.d.$(0, \sigma_\eta^2)$ and $v_{it} \sim$ i.i.d.$(0, \sigma_v^2)$ are mutually independent, and for all i we have $y_{i0} = 0$. Derive var $[y_{it}]$ for $t = 1, \ldots, T$.

Solution 5

In the previous exercise we have shown that

$$y_{it} = \alpha^t y_0 + \left(\sum_{s=0}^{t-1} \alpha^s\right) \eta_i + \sum_{s=0}^{t-1} \alpha^s v_{it-s},$$

and given that $y_{i0} = 0$, we have

$$y_{it} = \left(\sum_{s=0}^{t-1} \alpha^s\right) \eta_i + \sum_{s=0}^{t-1} \alpha^s v_{it-s}.$$

Now, the variance is

$$\text{var}\left[\sum_{s=0}^{t-1} \alpha^s v_{it-s}\right] = \sigma_v^2 \sum_{s=0}^{t-1} \alpha^{2s} = \sigma_v^2 \frac{1-\alpha^{2t}}{1-\alpha^2},$$

so that

$$\text{var}\,[y_{it}] = \sigma_\eta^2 \left(\frac{1-\alpha^t}{1-\alpha}\right)^2 + \sigma_v^2 \frac{1-\alpha^{2t}}{1-\alpha^2}.$$

Problem 6

Assume that we are in the AR(1) dynamic model setup such that

$$y_{it} = \alpha y_{it-1} + \eta_i + v_{it}$$

but now our v_{it} follows an MA(1) process such that

$$v_{it} = w_{it} + b w_{it-1},$$

where $w_{it} \sim$ i.i.d.$(0, \sigma_w^2)$ (i.e. v_{it} is serially correlated). Show that in this case the instrument y_{it-2} is not a valid instrument for estimating α with GMM in first differences, while the instruments y_{it-j} for $j \geq 3$ remain valid.

Solution 6

Given that $v_{it} \sim$ MA(1), the autocovariance of order $s = 0$ is

$$\begin{aligned}
\mathbb{E}[v_{it} v_{it}] &= \mathbb{E}[(w_{it} + b w_{it-1})(w_{it} + b w_{it-1})] \\
&= \mathbb{E}[w_{it}^2] + 2b\mathbb{E}[w_{it} w_{it-1}] + b^2 \mathbb{E}[w_{it-1}^2] \\
&= \sigma_w^2 + b^2 \sigma_w^2 = \sigma_w^2 (1 + b^2),
\end{aligned}$$

of order $s = 1$ is

$$\begin{aligned}
\mathbb{E}[v_{it} v_{it-1}] &= \mathbb{E}[(w_{it} + b w_{it-1})(w_{it-1} + b w_{it-2})] \\
&= \mathbb{E}[w_{it} w_{it-1}] + b\mathbb{E}[w_{it-1}^2] + b\mathbb{E}[w_{it} w_{it-2}] + b^2 \mathbb{E}[w_{it-1} w_{it-2}] = b\sigma_w^2,
\end{aligned}$$

and of order $s > 1$ is

$$\mathbb{E}[v_{it} v_{it-s}] = \mathbb{E}[(w_{it} + b w_{it-1})(w_{it-s} + b w_{it-s-1})] = \ldots = 0,$$

where we use the fact that w_{it} and w_{it-s} for $\forall s$ is independent by assumption of i.i.d. data. The first differences model is then

$$y_{it} - y_{it-1} = \alpha(y_{it-1} - y_{it-2}) + v_{it} - v_{it-1},$$

so to be valid, the instrument should not be correlated with the error $u_{it} := v_{it} - v_{it-1}$. In other words, for y_{it-2} to be valid, we require

$$\mathbb{E}[u_{it}y_{it-2}] = 0,$$

which holds if v_{it-2} (that is inside y_{it-2}) is uncorrelated to v_{it} and v_{it-1}. However, we have shown that

$$\mathbb{E}[v_{it}v_{it-1}] = \mathbb{E}[v_{it-1}v_{it-2}] = b\sigma_w^2 \neq 0,$$

which in turns implies that y_{it-2} is not a valid instrument.

On the other hand, any instrument that is of the form y_{t-j} for $j > 2$ *is* valid, because

$$\mathbb{E}[y_{it-j}u_{it}] = 0$$

due to $\mathbb{E}[v_{it}v_{it-s}] = 0$ for $s > 1$ with $s = j - 1$.

Problem 7

We have data for a panel of companies on gross investment expenditures I_{it} and net capital stock K_{it}. We model the investment rate $y_{it} = I_{it}/K_{it}$ as

$$\left(\frac{I_{it}}{K_{it}}\right) = \alpha\left(\frac{I_{it-1}}{K_{it-1}}\right) + \eta_i + v_{it},$$

and Table 2 shows the results of estimating the model in *levels* by OLS and WG, and the model in *first differences* with one instrument, two instruments, and all Arellano-Bond instruments. For the last two estimators, it also shows the Sargan test statistic and the m_2 statistic for second-order serial correlation in the residuals from the estimated model.

Table 1: Estimation results (703 firms, 4966 observations)

	OLS	WG	2SLS DIF	GMM DIF	GMM DIF
	(1)	(2)	(3)	(4)	(5)
$\hat{\alpha}$	0.2669	-0.0094	0.1626	0.1593	0.1560
	(.0185)	(.0181)	(.0362)	(.0327)	(.0318)
m_2				0.52	0.46
Sargan test				0.36	0.43
Instruments			$(I/K)_{t-2}$	$(I/K)_{t-2}$	$(I/K)_{t-2}$
				$(I/K)_{t-3}$	$(I/K)_{t-3}$
					$\vdots$
					$(I/K)_1$

a) For each of the models in columns (2) and (3), write down the estimated equation(s).

b) Comment on the estimates of α in each of the columns. Are the results in line with theory (in terms of possible bias of the different estimators)? Why do we need to use instruments?

c) Comment on the standard errors of the last three estimators. Are the results in line with theory?

d) For the two GMM estimators (column (4) and (5)), what do you conclude from the two specification tests? What are these tests' null hypotheses and why are these useful to run?

Solution 7

a) In column (2), the WG model is estimated, which takes away individual averages over time. Thus, the estimated equation is

$$\left(\frac{I_{it}}{K_{it}} - \frac{\bar{I}_i}{\bar{K}_i}\right) = \alpha\left(\frac{I_{it-1}}{K_{it-1}} - \frac{\bar{I}_i}{\bar{K}_i}\right) + v_{it} - \bar{v}_i,$$

where $\frac{\bar{I}_i}{\bar{K}_i} = T^{-1}\sum_{t=1}^{T}\frac{I_{it}}{K_{it}}$.

In column (3), the model is estimated in first differences,

$$\left(\frac{I_{it}}{K_{it}} - \frac{I_{it-1}}{K_{it-1}}\right) = \alpha\left(\frac{I_{it-1}}{K_{it-1}} - \frac{I_{it-2}}{K_{it-2}}\right) + v_{it} - v_{it-1}.$$

The estimation is done using 2SLS with the instrument $\frac{I_{it-2}}{K_{it-2}}$, so that in the first stage we project the endogenous regressor on the instrument,

$$\left(\frac{I_{it-1}}{K_{it-1}} - \frac{I_{it-2}}{K_{it-2}}\right) = \gamma_0 + \gamma_1\frac{I_{it-2}}{K_{it-2}} + \varepsilon_{it},$$

and in the second stage we project the outcome variable on fitted values from the first stage,

$$\left(\frac{I_{it}}{K_{it}} - \frac{I_{it-1}}{K_{it-1}}\right) = \alpha\widehat{\left(\frac{I_{it-1}}{K_{it-1}} - \frac{I_{it-2}}{K_{it-2}}\right)} + v_{it} - v_{it-1}$$

where $\widehat{\left(\frac{I_{it-1}}{K_{it-1}} - \frac{I_{it-2}}{K_{it-2}}\right)}$ are fitted values, $\hat{\gamma}_1\left(\frac{I_{it-2}}{K_{it-2}}\right)$.

b) If the specified model is correct, and we have

$$\left(\frac{I_{it}}{K_{it}}\right) = \alpha\left(\frac{I_{it-1}}{K_{it-1}}\right) + \eta_i + v_{it},$$

we are in a dynamic panel setting and running OLS and WG will deliver biased estimates. Even if $\mathbb{E}[y_{it-1}v_{it}] = 0$, we have

$$\operatorname*{plim}_{N\to\infty}\hat{\alpha}_{\text{OLS}} > \alpha, \quad \operatorname*{plim}_{N\to\infty}\hat{\alpha}_{\text{WG}} < \alpha,$$

because $\mathbb{E}[y_{it-1}\eta_i] > 0$. Thus, OLS will overestimate α whereas WG will underestimate α. However, if we estimate the model in first differences and use instruments (as in columns (3), (4), and (5)) we will get consistent estimates of α under the assumption that these are valid instruments. The instrument used in (3) implies the Anderson-Hsiao estimator for α, column (4) expands the number of instruments, and column (5) uses all Arellano-Bond instruments. These rely on the moment conditions $\mathbb{E}[y_{t-j}\Delta v_{it}] = 0$ for $j \geq 2$ which are satisfied as long as v_{it} are serially uncorrelated.

We need instruments in the first place, because the regressor $(\frac{I_{it-1}}{K_{it-1}} - \frac{I_{it-2}}{K_{it-2}})$ is by construction correlated with the error term $v_{it} - v_{it-1}$.

c) The standard errors of the three last columns are decreasing which is what we expect from theory when adding valid instruments. The efficiency gains from the extra instruments are very small though.

d) The two specification tests are the Sargan overidentification test and the test for second-order serial correlation in the residuals.

The Sargan overidentification test's null hypothesis is that the moment conditions are satisfied (the instruments are valid). The Sargan test statistic is distributed as a χ^2 with $(L - K)$ degrees of freedom (which are 1 and $T - 3$ in columns (4) and (5)) and in both models we obtain a low statistic of 0.36 and 0.43. Thus, we cannot reject the null hypothesis that the moment conditions are satisfied, and thus the instruments are valid. This test is crucial to determine whether we are using valid instruments and thus have a consistent estimate of α.

The test for second-order serial correlation in the residuals has as the null hypothesis that the residuals from the first-difference model are *not* second-order serially correlated. If this is the case then it implies that there is no correlation between $v_{it} - v_{it-1}$ and $v_{it-2} - v_{it-3}$ which would indicate that v_{it} are serially uncorrelated. This is a crucial assumption on which the validity of the instruments rests. The test statistic m_2 is normally distributed and with the reported m_2 of 0.52 and 0.46 we cannot reject the null hypothesis that the residuals from the first-difference model are *not* second-order serially correlated. Both specification tests go in the same direction and both imply that the instruments used are valid.

Problem 8

Formulate a linear dynamic panel regression with a single weakly exogenous regressor, and AR(2) feedback in place of AR(1) feedback (i.e. when two most recent lags of the left side variable are present at the right side). Describe the algorithm of estimation of this model.

Solution 8

Please, see *Problemnik 13.8*.

2 Causal inference

Problem 1

In Table below, you have the potential outcomes of 4 cancer patients (as in the example in class) with their potential outcomes and treatment assignment for $D_i = 0$ corresponding to undergoing chemotherapy and $D_i = 1$ corresponding to surgery.

Table 2: Surgery vs therapy

patient	Y_i^1	Y_i^0	δ_i	Y_i	D_i
		case 1			
1	4	3	1	4	1
1	6	5	1	6	1
3	1	2	-1	2	0
4	3	6	-3	6	0
		case 2			
1	5	3	2	5	1
1	6	5	1	6	1
3	7	6	1	6	0
4	10	8	2	8	0

Show that in one case there is no selection bias, whereas in the other there is no heterogeneous treatment effects bias.

Solution 1

Case 1. Given that individual treatment effects are available, we can compute the average treatment effect (ATE) directly as

$$\text{ATE} = \frac{1 + 1 - 1 - 3}{4} = -0.5.$$

The naive approach of estimating it using the simple difference in outcomes (SDO) would yield

$$\text{SDO} = \frac{4 + 6}{2} - \frac{2 + 6}{2} = 1.$$

In general, it holds that SDO = ATE + SB + HTB, where SD is the selection bias and HTB is the heterogeneous treatment effect bias. Because we have that ATE $\neq$ SDO, we conclude that one (or both) of the biases is nonzero.

The selection bias is given by

$$\text{SB} = \mathbb{E}[Y_i^0|D_i = 1] - \mathbb{E}[Y_i^0|D_i = 0] = \frac{3+5}{2} - \frac{2+6}{2} = 0.$$

For the heterogeneous treatment effect bias, we need to compute the average treatment effect on treated and untreated (ATT and ATU) respectively,

$$\text{ATT} = \mathbb{E}[\delta_i|D_i = 1] = \frac{1+1}{2} = 1, \quad \text{ATU} = \mathbb{E}[\delta_i|D_i = 0] = \frac{-1-3}{2} = -2.$$

The HTB is then

$$\text{HTB} = (1-\pi)(\text{ATT} - \text{ATU}) = 0.5(1+2) = 1.5$$

with π denoting the proportion of the treated observations.

Case 2. For the second case, we briefly state the results:

$$\begin{aligned}
\text{ATE} &= \frac{2+1+1+2}{4} = 1.5, \\
\text{SDO} &= \frac{5+6}{2} - \frac{6+8}{2} = -1.5, \\
\text{ATT} &= \frac{2+1}{2} = 1.5, \\
\text{ATU} &= \frac{1+2}{2} = 1.5, \\
\text{SB} &= \frac{3+5}{2} - \frac{6+8}{2} = -3, \\
\text{HTB} &= 0.
\end{aligned}$$

Problem 2

Show that running a regression

$$y_i = \beta_0 + \beta_1 D_i + \varepsilon_i, \quad i = 1, \ldots, N, \quad \mathbb{E}[\varepsilon_i|D_i] = 0,$$

leads to the OLS estimate of β_1 given by

$$\hat{\beta}_1 = \frac{1}{N_1}\sum_{i=1}^{N_1}(y_i|D_i = 1) - \frac{1}{N_0}\sum_{i=1}^{N_0}(y_i|D_i = 0),$$

which is a simple difference in means.

Solution 2

For subsequent derivations, we define $\bar{y}_1 := \frac{1}{N_1}\sum_{i=1}^{N_1}(y_i|D_i = 1)$ and $\bar{y}_0 := \frac{1}{N_0}\sum_{i=1}^{N_0}(y_i|D_i = 0)$ so that we have to show $\hat{\beta}_1 = \bar{y}_1 - \bar{y}_0$. The OLS estimator is given by

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(D_i - \bar{D})(y_i - \bar{y})}{(D_i - \bar{D})^2},$$

where $\bar{D} = N^{-1}\sum_{i=1}^{N} D_i = N_1/N$, and $\bar{y} = N^{-1}\sum_{i=1}^{N} y_i$. Expanding the numerator, we have

$$\sum_{i=1}^{N}(D_i - \bar{D})(y_i - \bar{y}) = \sum_{i=1}^{N} D_i y_i - D_i\bar{y} - \bar{D}y_i + \bar{D}\bar{y}.$$

Working with each term separately,

- $\sum_{i=1}^{N} D_i y_i = \frac{N_1}{N_1}\sum_{i=1}^{N} D_i y_i = N_1 \cdot \frac{1}{N_1}\sum_{i=1, D_i=1} y_i = N_1 \cdot \bar{y}_1,$

- $-\sum_{i=1}^{N} \bar{D}y_i = -\sum_{i=1}^{N} \frac{N_1}{N} y_i = -N_1 \cdot \bar{y},$

- $-\sum_{i=1}^N D_i\bar{y} = -\sum_{i=1,D_i=1}\bar{y} = -N_1 \cdot \bar{y}$,
- $\sum_{i=1}^N \bar{D}\bar{y} = \sum_{i=1}^N \frac{N_1}{N}\bar{y} = N_1 \cdot \bar{y}$,

we have that

$$\sum_{i=1}^{N}(D_i - \bar{D})(y_i - \bar{y}) = N_1(\bar{y}_1 - \bar{y}) = \left(N_1 - \frac{N_1^2}{N}\right)(\bar{y}_1 - \bar{y}_0)$$

where we substitute for $\bar{y}$ using the fact that $N = N_0 + N_1$, and $\bar{y} = N^{-1}(N_1\bar{y}_1 + N_0\bar{y}_0)$.

Now, for the denominator, we have

$$\begin{aligned}
\sum_{i=1}^{N}(D_i - \bar{D})^2 &= \sum_{i=1}^{N} D_i^2 - 2D_i\bar{D} + \bar{D}^2 \\
&= \sum_{i=1}^{N} D_i^2 - 2\sum_{i=1}^{N} D_i\bar{D} + \sum_{i=1}^{N}\bar{D}^2 \\
&= N_1 - 2\sum_{i=1}^{N} D_i\frac{N_1}{N} + \sum_{i=1}^{N}\frac{N_1^2}{N^2} \\
&= N_1 - \frac{N_1^2}{N}.
\end{aligned}$$

Putting everything together, we have

$$\hat{\beta}_1 = \frac{\left(N_1 - \frac{N_1^2}{N}\right)(\bar{y}_1 - \bar{y}_0)}{N_1 - \frac{N_1^2}{N}} = \bar{y}_1 - \bar{y}_0.$$

Problem 3

Show that running a regression

$$y_i = \beta_0 + \beta_1 D_i + \varepsilon_i, \quad i = 1, \ldots, N, \quad \mathbb{E}[\varepsilon_i|D_i] = 0,$$

with heteroskedasticity-robust standard errors provides an estimate that is numerically equivalent to the variance of the average treatment effect computed directly.

Solution 3

In the previous exercise we showed that $\hat{\beta}_1 = \bar{y}_1 - \bar{y}_0$. Answering the question in the setup is equivalent to showing that

$$\widehat{\text{var}}\,[\hat{\beta}_1] = \widehat{\text{var}}\,[\bar{y}_1] - \widehat{\text{var}}\,[\bar{y}_0] = \hat{\sigma}_1^2/N_1 + \hat{\sigma}_0^2/N_0 \equiv \widehat{\text{var}}_{\text{EHW}}[\hat{\beta}_1]$$

where $\hat{\sigma}_k^2$, $k \in \{0,1\}$ are estimated variances of the outcome computed on treated and untreated respectively, and $\widehat{\text{var}}_{\text{EHW}}[\hat{\beta}_1]$ are standard Eicker-Huber-White (EHW) standard errors.

Denote $x_i := \begin{pmatrix}1 & D_i\end{pmatrix}'$. Then the asymptotic heteroskedasticity-robust variance of the OLS estimator is

$$\text{var}\left[\begin{pmatrix}\hat{\beta}_0 \\ \hat{\beta}_1\end{pmatrix}\right] = \frac{1}{N}\mathbb{E}[x_i x_i']^{-1}\mathbb{E}[x_i x_i'\varepsilon_i^2]\mathbb{E}[x_i x_i']^{-1}.$$

Denoting residuals as $\hat{\varepsilon}_i := y_i - x_i'\hat{\beta}$, the estimator is then the usual EHW estimator,

$$\begin{aligned}
\widehat{\text{var}}\left[\begin{pmatrix}\hat{\beta}_0 \\ \hat{\beta}_1\end{pmatrix}\right] &= \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\hat{\varepsilon}_i^2\right)\left(\frac{1}{N}\sum_{i=1}^{N} x_i x_i'\right)^{-1} \\
&= \frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}1 & D_i \\ D_i & D_i\end{pmatrix}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}1 & D_i \\ D_i & D_i\end{pmatrix}\hat{\varepsilon}_i^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\begin{pmatrix}1 & D_i \\ D_i & D_i\end{pmatrix}\right)^{-1} \\
&= \frac{1}{N}\begin{pmatrix}1 & N_1/N \\ N_1/N & N_1/N\end{pmatrix}^{-1}\begin{pmatrix}N^{-1}\sum_{i=1}^{N}\hat{\varepsilon}_i^2 & N^{-1}\sum_{i=1,D_i=1}\hat{\varepsilon}_i^2 \\ N^{-1}\sum_{i=1,D_i=1}\hat{\varepsilon}_i^2 & N^{-1}\sum_{i=1,D_i=1}\hat{\varepsilon}_i^2\end{pmatrix}\begin{pmatrix}1 & N_1/N \\ N_1/N & N_1/N\end{pmatrix}^{-1} \\
&= \begin{pmatrix}N & N_1 \\ N_1 & N_1\end{pmatrix}^{-1}\begin{pmatrix}\sum_{i=1}^{N}\hat{\varepsilon}_i^2 & \sum_{i=1,D_i=1}\hat{\varepsilon}_i^2 \\ \sum_{i=1,D_i=1}\hat{\varepsilon}_i^2 & \sum_{i=1,D_i=1}\hat{\varepsilon}_i^2\end{pmatrix}\begin{pmatrix}N & N_1 \\ N_1 & N_1\end{pmatrix}^{-1}.
\end{aligned}$$

Computing the inverses, and using the fact that $N = N_0 + N_1$, and $\sum_{i=1}^{N} \hat{\varepsilon}_i^2 = \sum_{i=1,D_i=0} \hat{\varepsilon}_i^2 + \sum_{i=1,D_i=1} \hat{\varepsilon}_i^2$ we have

$$\begin{aligned}
\widehat{\text{var}}\left[\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix}\right] &= \begin{pmatrix}1/N_0 & -1/N_0\\ -N_0 & 1/N_0+1/N_1\end{pmatrix}^{-1}\begin{pmatrix}\sum_{i=1}^{N}\hat{\varepsilon}_i^2 & \sum_{i=1,D_i=1}\hat{\varepsilon}_i^2\\ \sum_{i=1,D_i=1}\hat{\varepsilon}_i^2 & \sum_{i=1,D_i=1}\hat{\varepsilon}_i^2\end{pmatrix}\begin{pmatrix}1/N_0 & -1/N_0\\ -N_0 & 1/N_0+1/N_1\end{pmatrix}^{-1}\\
&= \begin{pmatrix}N_0^{-1}\sum_{i=1,D_i=0}\hat{\varepsilon}_i^2 & 0\\ N_1^{-1}\sum_{i=1,D_i=1}\hat{\varepsilon}_i^2 - N_0^{-1}\sum_{i=1,D_i=0}\hat{\varepsilon}_i^2 & N_1^{-1}\sum_{i=1,D_i=1}\hat{\varepsilon}_i^2\end{pmatrix}\begin{pmatrix}1/N_0 & -1/N_0\\ -N_0 & 1/N_0+1/N_1\end{pmatrix}^{-1}\\
&= \begin{pmatrix}N_0^{-2}\sum_{i=1,D_i=0}\hat{\varepsilon}_i^2 & -N_0^{-2}\sum_{i=1,D_i=0}\hat{\varepsilon}_i^2\\ -N_0^{-2}\sum_{i=1,D_i=0}\hat{\varepsilon}_i^2 & N_0^{-2}\sum_{i=1,D_i=0}\hat{\varepsilon}_i^2 + N_1^{-2}\sum_{i=1,D_i=1}\hat{\varepsilon}_i^2\end{pmatrix}.
\end{aligned}$$

We note that the low-right element is exactly the estimated variance of $\hat{\beta}_1$,

$$\begin{aligned}
\widehat{\text{var}}\,[\hat{\beta}_1] &= N_0^{-2}\sum_{i=1,D_i=0}\hat{\varepsilon}_i^2 + N_1^{-2}\sum_{i=1,D_i=1}\hat{\varepsilon}_i^2\\
&:= \hat{\sigma}_0^2/N_0 + \hat{\sigma}_1^2/N_1.
\end{aligned}$$

Problem 4

Show that *i)* if $p(X)$ is the propensity score, then $X \perp D \mid p(X)$; *ii)* if $Y^1, Y^0 \perp D \mid X$, then assignment to treatment is unconfounded given the propensity score, i.e. $Y^1, Y^0 \perp D \mid p(X)$.

Solution 4

For *i)*, first, note that

$$\mathbb{P}\{D=1|X,p(X)\} = \mathbb{E}[D|X,p(X)] = \mathbb{E}[D|X] = \mathbb{P}\{D=1|X\} := p(X),$$

and

$$\mathbb{P}\{D=1|p(X)\} = \mathbb{E}[D|p(X)] = \mathbb{E}[D|p(X),X] = \mathbb{E}\left[\mathbb{E}[D|p(X),X]|p(X)\right] = \mathbb{E}[p(X)|p(X)] = p(X).$$

So in total we have that

$$\mathbb{P}\{D=1|X,p(X)\} = \mathbb{P}\{D=1|p(X)\},$$

that is, conditional on $p(X)$, probability of receiving a treatment is the same with or without conditioning on the observed covariates X. In other words, conditional on $p(X)$, the treatment is independent of covariates.

For *ii)*, we have that

$$\begin{aligned}
\mathbb{P}\{D=1|Y^1,Y^0,p(X)\} &= \mathbb{E}[D|Y^1,Y^0,p(X)]\\
&= \mathbb{E}_X\left[\mathbb{E}[D|X,Y^1,Y^0]|Y^1,Y^0,p(X)\right]\\
&= \mathbb{E}_X\left[\mathbb{E}[D|X]|Y^1,Y^0,p(X)\right]\\
&= \mathbb{E}_X\left[p(X)|Y^1,Y^0,p(X)\right] = p(X),
\end{aligned}$$

where the third equality follows from the "if" assumption in the setup. From *i)*, we have that

$$\mathbb{P}\{D=1|p(X)\} = \mathbb{P}\{D=1|Y^1,Y^0,p(X)\} = p(X),$$

that is, conditional on $p(X)$, the treatment and the potential outcomes are independent.

Problem 5

Show that the IV coefficient given by

$$\delta_{\text{IV}} = \frac{\text{cov}\,[Z_i,Y_i]}{\text{cov}\,[Z_i,D_i]}$$

results in

$$\delta_{\text{W}} = \frac{\mathbb{E}[Y_i|Z_i=1]-\mathbb{E}[Y_i|Z_i=0]}{\mathbb{E}[D_i|Z_i=1]-\mathbb{E}[D_i|Z_i=0]}.$$

Solution 5

First, the numerator of δ_{IV} is

$$\begin{aligned}\text{cov}\,[Z_i, Y_i] &= \mathbb{E}[Z_i Y_i] - \mathbb{E}[Z_i]\mathbb{E}[Y_i] \\ &= \mathbb{E}[Y_i|Z_i = 1] \cdot \mathbb{P}\{Z_i = 1\} - \mathbb{P}\{Z_i = 1\} \cdot (\mathbb{E}[Y_i|Z_i = 1] \cdot \mathbb{P}\{Z_i = 1\} + \mathbb{E}[Y_i|Z_i = 0] \cdot \mathbb{P}\{Z_i = 0\}) \\ &= \mathbb{E}[Y_i|Z_i = 1] \cdot \mathbb{P}\{Z_i = 1\} - \mathbb{P}\{Z_i = 1\}^2\mathbb{E}[Y_i|Z_i = 1] - \mathbb{P}\{Z_i = 1\}(1 - \mathbb{P}\{Z_i = 1\})\mathbb{E}[Y_i|Z_i = 0] \\ &= (\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0])\,\mathbb{P}\{Z_i = 1\}(1 - \mathbb{P}\{Z_i = 1\}).\end{aligned}$$

Next, the denominator is by analogy is

$$\text{cov}\,[Z_i, D_i] = (\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0])\,\mathbb{P}\{Z_i = 1\}(1 - \mathbb{P}\{Z_i = 1\}),$$

so that division $\text{cov}\,[Z_i, Y_i]/\text{cov}\,[Z_i, D_i]$ yields δ_{W}.

Problem 6

Suppose that with heterogeneous treatment effects instead of monotonicity, we assume an eligibility rule of the form

$$\mathbb{P}(D_i = 1|Z_i = 0) = 0.$$

Show that in this case, we can identify the average treatment effect on the treated.

Solution 6

First, compute

$$\begin{aligned}\mathbb{E}[Y_i|Z_i = 1] &= \mathbb{E}[Y_i^0] + \mathbb{E}[(Y_i^1 - Y_i^0)D_i|Z_i = 1] \\ &= \mathbb{E}[Y_i^0] + \mathbb{E}\left[\mathbb{E}[(Y_i^1 - Y_i^0)D_i|D_i = 1, Z_i = 1]|Z_i = 1\right] \\ &= \mathbb{E}[Y_i^0] + \mathbb{E}[Y_i^1 - Y_i^0|D_i = 1, Z_i = 1] \cdot \mathbb{P}\{D_i = 1|Z_i = 1\} \\ &= \mathbb{E}[Y_i^0] + \mathbb{E}[Y_i^1 - Y_i^0|D_i = 1] \cdot \mathbb{P}\{D_i = 1|Z_i = 1\},\end{aligned}$$

where the last line follows from the eligibility rule assumption, because the event $D_i = 1$ is sufficient for the event $Z_i = 1$.

Similarly,

$$\begin{aligned}\mathbb{E}[Y_i|Z_i = 0] &= \mathbb{E}[Y_i^0] + \mathbb{E}[(Y_i^1 - Y_i^0)D_i|Z_i = 0] \\ &= \mathbb{E}[Y_i^0] + \mathbb{E}[(Y_i^1 - Y_i^0)|D_i = 1, Z_i = 0] \cdot \mathbb{P}(D_i = 1|Z_i = 1) \\ &= \mathbb{E}[Y_i^0]\end{aligned}$$

due to the eligibility rule assumption. It follows that

$$\frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{P}(D_i = 1|Z_i = 1)} = \mathbb{E}[Y_i^1 - Y_i^0|D_i = 1].$$

By definition, ATT $:= \mathbb{E}[Y_i^1 - Y_i^0|D_i = 1]$ so we have that

$$\text{ATT} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{P}(D_i = 1|Z_i = 1)} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1]},$$

which is the Wald estimand having $\mathbb{E}[D_i|Z_i = 0] = 0$ by the eligibility rule assumption.

Problem 7

The following table contains data on civilian mortality for white men born in 1950 and 1951. For the men born in 1951, interpret the reduced form (intention-to-treat) and the IV estimates.

Year	Draft eligibility[a]	Number of deaths[b]	Number of suicides[c]	Probability of death[d]	Probability of suicide	Probability of military service[e]
1950	Yes	2,601	436	.0204	.0034	.3527
				(.0004)	(.0002)	(.0325)
	No	2,169	352	.0195	.0032	.1934
				(.0004)	(.0002)	(.0233)
Difference (Yes minus No)				.0009	.0002	.1593
				(.0006)	(.0002)	(.0401)
IV estimates[f]				.0056	.0013	
				(.0040)	(.0013)	
1951	Yes	1,494	279	.0170	.0032	.2831
				(.0004)	(.0002)	(.0390)
	No	2,823	480	.0168	.0029	.1468
				(.0003)	(.0001)	(.0180)
Difference (Yes minus No)				.0002	.0003	.1362
				(.0005)	(.0002)	(.0429)
IV estimates				.0015	.0022	
				(.0037)	(.0016)	

Figure 1: Data on civilian mortality (Angrist, Imbens, and Rubin, 1996)

Solution 7

What are the intention-to-treat (ITT) effects here? Remeber that ITT effects come from the reduced form regression, in which we regress the outcome variable (here, probability of death or suicide) on the assignment to treatment (drafting to the army). The coefficient of the regression where there is only the constant and the assignment to treatment,

$$Y_i = \alpha + \gamma Z_i + \varepsilon_i,$$

is equivalent to the mean comparison between the two groups,

$$\gamma = \mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0].$$

Thus, the *Difference (Yes minus No)* line tells us what is the effect of the assignment on the probability of death and suicide. Men born in 1951 who were assigned treatment had a 0.02% higher probability to die compared to those not assigned, and 0.03% higher probability to commit suicide. These are only ITT effects as we know there were men that were treated and not treated in both assignment groups.

The IV results are based on the Wald estimand,

$$\delta_{\text{W}} = \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]}.$$

Note that the Wald estimand is essentially dividing the ITT by the difference in the probability of serving for those assigned and not assigned (which is shown in the last column). In fact, the IV results are

$$\hat{\delta}_{\text{W, death}} = \frac{0.0002}{0.1362} = 0.0015, \quad \hat{\delta}_{\text{W, suicide}} = \frac{0.0003}{0.1362} = 0.0022.$$

We know that the Wald estimand gives the LATE, that is, the average causal effect for the compliers (those who were induced to serve by the draft). Note, however, that most effects are not statistically significant.

Problem 8

You have seen in class that with sharp regression discontinuity design (RDD), it is important to correctly specify the underlying model of how the continuous variable impacts the outcome variable. One possible misspecification shown was when a nonlinear model with no discontinuity is fit by a linear model.

In the paper *"Evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai river policy"* by Chen, Ebenstein, Greenstone, and Li, the authors use China's Huai river policy as a sharp discontinuity to estimate the effect of air pollution on life expectancy. China's Huai river policy provided free winter heating via provision of coal for boilers in the cities north of the Huai river, but denied heat to the south cities. The authors show that pollution (measured as *TSP*: total suspended particulates) was larger in the north versus the south of the river.

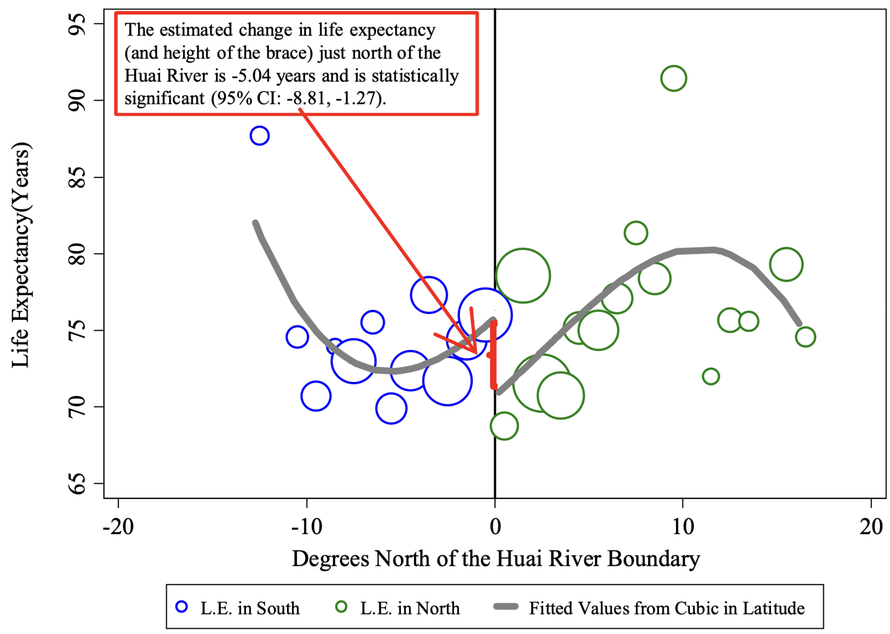

Figure 2: Estimated change in life expectancy (Chen, Ebenstein, Greenstone, and Li, 2013)

	Linear & Controls	Quadratic & Controls	Cubic & Controls	Quartic & Controls	Quintic & Controls
	(1)	(2)	(3)	(4)	(5)
Panel 1: Impact of "North" on the Listed Variable, Ordinary Least Squares					
TSP (100 μg/m^3)	2.89***	2.63***	1.84***	1.95***	1.52**
	(0.56)	(0.49)	(0.63)	(0.59)	(0.72)
	[0.988]	[0.068]	[0.148]	[0.229]	[0.671]
	{492.4}	{489}	{487.2}	{486.3}	{487.5}
ln(All Cause Mortality Rate)	0.12	0.09	0.26*	0.26**	0.37**
	(0.10)	(0.10)	(0.13)	(0.13)	(0.16)
	[0.276]	[0.215]	[0.035]	[0.908]	[0.409]
	{39.88}	{38.8}	{34.11}	{35.92}	{36.13}
ln(Cardiorespiratory Mortality Rate)	0.13	0.09	0.38**	0.39**	0.47**
	(0.13)	(0.13)	(0.16)	(0.16)	(0.19)
	[0.652]	[0.243]	[0.003]	[0.747]	[0.696]
	{102.3}	{101.5}	{91.92}	{93.34}	{94.62}
ln(Non-Cardiorespiratory Mortality Rate)	0.09	0.05	0.08	0.07	0.19
	(0.10)	(0.09)	(0.13)	(0.12)	(0.14)
	[0.135]	[0.151]	[0.933]	[0.973]	[0.35]
	{43.04}	{41.27}	{43.13}	{45.07}	{44.97}
Life Expectancy (years)	-1.62	-1.29	-5.52**	-5.67**	-5.43*
	(1.66)	(1.68)	(2.39)	(2.36)	(2.94)
	[0.101]	[0.6]	[0.001]	[0.737]	[0.984]
	{757.1}	{758}	{746.8}	{748.2}	{750.2}

Figure 3: Robustness checks of choice of functional form (Chen, Ebenstein, Greenstone, and Li, 2013)

The authors fit a third-degree polynomial to the data on degrees north of the river (the running variable where the effect is discontinuous at 0 due to the policy), and show Figure 2. Also, they provide details on different functional approximations, varying the polynomial degree from 1 to 5, see Table 3.

Comment on what you think may be the issues with this kind of analysis with respect to the model specification.

Solution 8

The first thing to notice is that in the graph, if we forget about the fitted third-degree polynomial, it is not very easy to see any large shift in life expectancy. This means that we may have the opposite problem than the one shown in the lecture. Here, fitting a nonlinear model actually shows an effect on life expectancy at the cut-off (zero) but if we had fit a linear model, this effect would not be there. This is confirmed by the results in the table (which authors present as a robustness check). In fact, for a linear or quadratic model, there is no significant effect on life expectancy of being left or right to the cut-off. Only at higher-order polynomials we start seeing an effect.

Problem 9

Show that from the following specification

$$Y_{it} = \beta_0 + \beta_D D_i + \beta_T T_t + \beta D_i T_t + U_{it}, \quad \mathbb{E}[U_{it}|D_i, T_t] = 0,$$

we have that

$$\begin{aligned}\beta = &\left(\mathbb{E}[Y_{it}|D_i = 1, T_t = 1] - \mathbb{E}[Y_{it}|D_i = 1, T_t = 0]\right) \\ &- \left(\mathbb{E}[Y_{it}|D_i = 0, T_t = 1] - \mathbb{E}[Y_{it}|D_i = 0, T_t = 0]\right).\end{aligned}$$

Solution 9

First, compute conditional expectations exactly,

$$\begin{aligned}\mathbb{E}[Y_{it}|D_i = 1, T_t = 1] &= \beta_0 + \beta_D + \beta_T, \\ \mathbb{E}[Y_{it}|D_i = 1, T_t = 0] &= \beta_0 + \beta_D, \\ \mathbb{E}[Y_{it}|D_i = 0, T_t = 1] &= \beta_0 + \beta_T, \\ \mathbb{E}[Y_{it}|D_i = 0, T_t = 0] &= \beta_0.\end{aligned}$$

From the third and the forth equation we have

$$\beta_T = \mathbb{E}[Y_{it}|D_i = 0, T_t = 1] - \mathbb{E}[Y_{it}|D_i = 0, T_t = 0].$$

Then, from the first and the second equation we have

$$\begin{aligned}\beta &= \mathbb{E}[Y_{it}|D_i = 1, T_t = 1] - (\beta_0 + \beta_D) - \beta_T \\ &= \mathbb{E}[Y_{it}|D_i = 1, T_t = 1] - \mathbb{E}[Y_{it}|D_i = 1, T_t = 0] - \beta_T \\ &= \left(\mathbb{E}[Y_{it}|D_i = 1, T_t = 1] - \mathbb{E}[Y_{it}|D_i = 1, T_t = 0]\right) \\ &\quad - \left(\mathbb{E}[Y_{it}|D_i = 0, T_t = 1] - \mathbb{E}[Y_{it}|D_i = 0, T_t = 0]\right).\end{aligned}$$

Problem 10

Show how the coefficients from the interaction model as in Card and Krueger (1993),

$$Y_{ist} = \alpha + \gamma \mathrm{NJ}_s + \lambda_t d_t + \beta_{(1)} \left(\mathrm{NJ}_s \cdot d_t\right) + \varepsilon_{ist} \tag{1}$$

correspond to the coefficients in the two-way fixed effects model,

$$Y_{ist} = \gamma_s + \theta_t + \beta_{(2)} D_{st} + \xi_{ist}, \tag{2}$$

and show that $\beta_{(1)} = \beta_{(2)}$ is exactly the same in each specification.

Solution 10

In (1), we have

$$\mathrm{NJ}_s := \begin{cases} 1, & \text{if } i \text{ is in New Jersey,} \\ 0, & \text{if } i \text{ is in Pennsylvania,} \end{cases} \qquad d_t := \begin{cases} 1, & \text{if } i \text{ is in November,} \\ 0, & \text{if } i \text{ is in February.} \end{cases}$$

In (2), we have

$$\gamma_s := \begin{cases} \gamma_{\mathrm{NJ}}, & \text{if } i \text{ is in New Jersey,} \\ \gamma_{\mathrm{PA}}, & \text{if } i \text{ is in Pennsylvania,} \end{cases} \qquad \theta_t := \begin{cases} \theta_{\mathrm{Nov}}, & \text{if } i \text{ is in November,} \\ \theta_{\mathrm{Feb}}, & \text{if } i \text{ is in February,} \end{cases},$$

and

$$D_{st} := \begin{cases} 1, & \text{if } i \text{ is in New Jersey in November,} \\ 0, & \text{otherwise.} \end{cases}$$

Computing conditional expectations from (1), we have

$$\begin{aligned} \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Feb}] &= \alpha, \\ \mathbb{E}[Y_{ist}|s = \mathrm{NJ}, t = \mathrm{Feb}] &= \alpha + \gamma, \\ \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Nov}] &= \alpha + \lambda_t, \\ \mathbb{E}[Y_{ist}|s = \mathrm{NJ}, t = \mathrm{Nov}] &= \alpha + \lambda + \lambda_t + \beta_{(1)}, \end{aligned}$$

from which follows that

$$\begin{aligned} \lambda_t &= \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Nov}] - \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Feb}], \\ \gamma &= \mathbb{E}[Y_{ist}|s = \mathrm{NJ}, t = \mathrm{Feb}] - \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Feb}], \\ \beta_{(1)} &= (\mathbb{E}[Y_{ist}|s = \mathrm{NJ}, t = \mathrm{Nov}] - \mathbb{E}[Y_{ist}|s = \mathrm{NJ}, t = \mathrm{Feb}]) \\ &\quad - (\mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Nov}] - \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Feb}) \,. \end{aligned}$$

From (2), we have that

$$\begin{aligned} \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Feb}] &= \gamma_{\mathrm{PA}} + \theta_{\mathrm{Feb}}, \\ \mathbb{E}[Y_{ist}|s = \mathrm{NJ}, t = \mathrm{Feb}] &= \gamma_{\mathrm{NJ}} + \theta_{\mathrm{Feb}}, \\ \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Nov}] &= \gamma_{\mathrm{PA}} + \theta_{\mathrm{Nov}}, \\ \mathbb{E}[Y_{ist}|s = \mathrm{NJ}, t = \mathrm{Nov}] &= \gamma_{\mathrm{NJ}} + \theta_{\mathrm{Nov}} + \beta_{(2)}, \end{aligned}$$

Combining (1) with (2), we have

$$\begin{aligned} \alpha &= \gamma_{\mathrm{PA}} + \theta_{\mathrm{Feb}}, \\ \gamma &= \gamma_{\mathrm{NJ}} + \theta_{\mathrm{Feb}} - \gamma_{\mathrm{PA}} - \theta_{\mathrm{Feb}} = \gamma_{\mathrm{NJ}} - \gamma_{\mathrm{PA}}, \\ \lambda_t &= \gamma_{\mathrm{PA}} + \theta_{\mathrm{Nov}} - \gamma_{\mathrm{PA}} - \theta_{\mathrm{Feb}} = \theta_{\mathrm{Nov}} - \theta_{\mathrm{Feb}}, \end{aligned}$$

and from the expression for $\beta_{(2)}$ from (2),

$$\begin{aligned} \beta_{(2)} &= \mathbb{E}[Y_{ist}|s = \mathrm{NJ}, t = \mathrm{Nov}] - \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Nov}] - \underbrace{(\gamma_{\mathrm{NJ}} - \gamma_{\mathrm{PA}})}_{=\gamma} \\ &= \mathbb{E}[Y_{ist}|s = \mathrm{NJ}, t = \mathrm{Nov}] - \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Nov}] \\ &\quad - (\mathbb{E}[Y_{ist}|s = \mathrm{NJ}, t = \mathrm{Feb}] - \mathbb{E}[Y_{ist}|s = \mathrm{PA}, t = \mathrm{Feb}]) \,, \end{aligned}$$

and rearranging we see that $\beta_{(1)} = \beta_{(2)}$.