Econ 2120 problem set 2: solutions

October 9, 2022

Part 1

Ex 1.

(a) Results:

	constant	educ	exp	exp2
$\hat{\alpha}$	401.9767	8.7061	9.4386	-0.2495

(b) If we add IQ we get the $\hat{\beta}$ coefficients. But we know, by our residual regression discussions, that

$$\hat{\alpha}_{\text{educ}} = \hat{\beta}_{\text{educ}} + \hat{\beta}_{\text{IQ}} \hat{\gamma}_{\text{educ}}$$

where $\hat{\gamma}_{\text{educ}}$ is the schooling coefficient on a regression of IQ on a constant, schooling, experience, and experience squared. So if we solve this equation for $\hat{\gamma}_{\text{educ}}$ we get

$$\frac{8.7061 - 7.3645}{0.4653} = 2.8834$$

And we see that it is the same coefficient from the auxiliary regression.

	constant	educ	exper	exper2	iq
$\hat{\beta}$	372.2434	7.3645	9.2691	-0.2387	0.4653
$\hat{\gamma}$	63.9041	2.8834	0.3644	-0.0232	—

Ex 2. We can construct w as the residual of the auxiliary regression, that is

$$w = \text{IQ} - \hat{\text{IQ}}$$

The idea here is that we can write

$$\begin{aligned} \hat{y} &= \hat{\beta}_0 + \hat{\beta}_{\text{educ}}\text{educ} + \hat{\beta}_{\text{exper}}\text{exper} + \hat{\beta}_{\text{exper}^2}\text{exper}^2 + \hat{\beta}_{\text{IQ}}\text{IQ} = \\ &= \underbrace{\hat{\beta}_0 + \hat{\beta}_{\text{educ}}\text{educ} + \hat{\beta}_{\text{exper}}\text{exper} + \hat{\beta}_{\text{exper}^2}\text{exper}^2 + \hat{\beta}_{\text{IQ}}\hat{\text{IQ}}}_{\text{linear combinations of constant, educ, exper, and exper2}} + \hat{\beta}_{\text{IQ}} w \end{aligned}$$

And, because w is orthogonal to linear functions of the other regressors, we can regress y on just w and still get the same coefficient $\hat{\beta}_{\text{IQ}}$.

If we actually run the regression we get

	:
$\hat{\beta}$	0.4653

	constant	educ	exp	exp2	iq	fed	med
$\hat{\beta}$	370.1164	6.8720	8.8621	-0.2168	0.4421	0.2967	0.9710

Ex 3. (a) Results:

The coefficients should be interpreted in terms of log-point changes on earnings, or changes on the log of earnings. We could approximate that as percentage changes.

In terms of magnitude, they look reasonable: an extra year of education would predict an increase of approximately 7% of earnings; experience has a similar magnitude, but has decreasing returns; an extra standard deviation of IQ would predict[1] roughly the same increase in earnings; an extra year of mother's education would predict an effect of around 1% in earnings, which looks small, but you need to consider that we are already controlling by the person education, probably one of the main channel through which mother's education would influence the earnings of a child; the effect of father's education is one third of that.

(b) Controlling for IQ and family background reduces the coefficient on education by almost 2 percentage points, which is almost a quarter of the original coefficient. But this change is expected as schooling is positively correlated with IQ, fed and med.

Ex 4. This just follows the derivation of the asymptotic distribution derived in Lecture Note 3.

without assuming homoskedasticity	5.0842	8.6598
assuming homoskedasticity	5.2073	8.5368

Ex 5. `bblp` takes in an n-vector `y` of the variable to be predicted; an $n \times K$-matrix `x` of the predictors, where the ith row is the vector of predictors for the ith observation; and `ndraw`, the number of draws to take from the posterior.

The program makes `ndraw` draws from the posterior distribution of β, the best linear predictor coefficients of Y_i given X_i, assuming a multinomial model for (X_i, Y_i) and the (improper) limiting Dirichlet prior.

The output is a $K\times$ `ndraw` matrix, where each column is a draw from the posterior distribution of β.

Ex 6. I used 2000 draws and got

	constant	educ	exp	exp2	iq	fed	med
means	369.8211	6.8868	8.8687	-0.2169	0.4428	0.3015	0.9692
std	23.5133	0.9129	2.4783	0.0874	0.1167	0.3082	0.4084

Ex 7. The tables below gives the quantiles, the normal approximation, and the asymptotic frequentist CIs in that order.

const	324.7688	415.7490
β_{educ}	5.1585	8.6549
β_{exp}	3.7736	13.6184
β_{exp2}	-0.3870	-0.0406
β_{iq}	0.2054	0.6677
β_{fed}	-0.3008	0.9036
β_{med}	0.1821	1.8002

const	323.7351	415.9072
β_{educ}	5.0974	8.6761
β_{exp}	4.0113	13.7260
β_{exp2}	-0.3881	-0.0456
β_{iq}	0.2141	0.6714
β_{fed}	-0.3026	0.9056
β_{med}	0.1687	1.7696

const	324.4120	415.8207
β_{educ}	5.0842	8.6598
β_{exp}	3.9637	13.7605
β_{exp2}	-0.3892	-0.0443
β_{iq}	0.2117	0.6726
β_{fed}	-0.3262	0.9195
β_{med}	0.1517	1.7903

The quantiles and normal approximation look really similar, which hints that our normal approximation is fine for $\hat{\beta}_j$, but it may not be fine for other functions of β.

These freqentist intervals are very close to the 95% credible intervals from the Bayesian bootstrap. This occurs because the frequentist sampling distribution of $\hat{\beta}$ is close to the Bayesian posterior distribution of β.

Ex 8. If we look at another group of people with the same regressors but one extra year of education we expect the log earnings to be higher by β_{educ}, if we increase the IQ points by γ we expect the log earnings to go up by $\beta_{\text{IQ}}\gamma$, so if we want them to be equal we must have

$$\gamma = \frac{\beta_{\text{educ}}}{\beta_{\text{IQ}}}$$

And a natural estimator for γ is to replace the β's by their estimators $\hat{\beta}$.

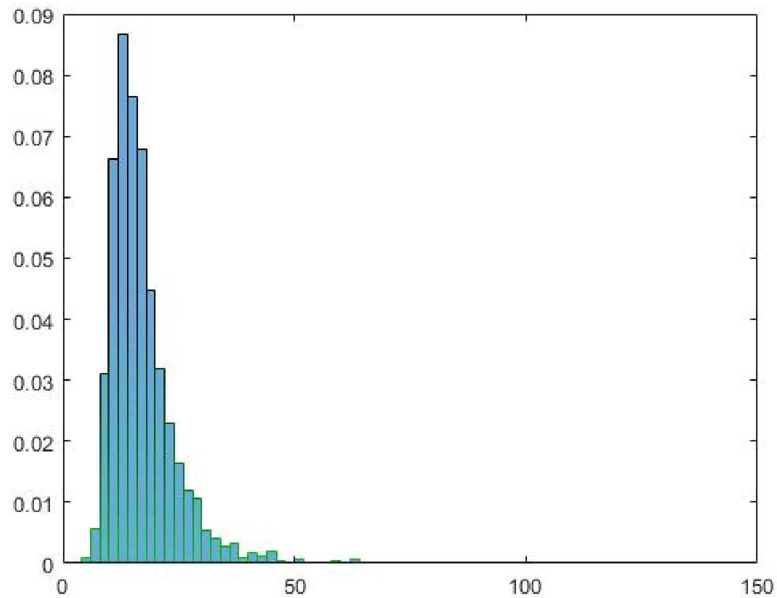
If we do that and use the bootstrap and the normal approximation we get 95% intervals as the following

quantile	8.5553	36.7298
normal	1.0582	33.5895

They disagree. The problem here is that, even though our sample size is large, the normal approximation is not very good. One diagnostic here is to look at the histogram from the bootstrap (below),

[1]This is easier to interpret than an increase of 1 IQ point, as the latter would change depending on which scale we are measuring IQ.

we see that the distribution is asymmetrical, which is why the normal approximation gives a different 95% probability interval.



Part 2

a)

Yes. As (Z_1, Z_2) are both binary, there are four possible combinations: $(0,0),(1,0),(0,1),(1,1)$. The conditional expectation of Y on (Z_1, Z_2) also have four values: $E[Y|(0,0)]$.... We have a CEF linear in $(1, Z_1, Z_2, Z_1 \times Z_2)$ that spans the four combinations. Please note this result can be generalized to any space spanned by discrete variables. We can always partition the data by a finite number of combinations of Z's.

b)

$$\begin{aligned}
E[Y|(0,0)] &= \beta_0 \\
E[Y|(1,0)] &= \beta_0 + \beta_1 \\
E[Y|(0,1)] &= \beta_0 + \beta_2 \\
E[Y|(1,1)] &= \beta_0 + \beta_1 + \beta_2 + \beta_3 \\
&\rightarrow \beta_3 = (E[Y|(1,1)] - E[Y|(0,1)]) - (E[Y|(1,0)] - E[Y|(0,0)])
\end{aligned}$$

where β_3 is often referred to as a difference in difference. We form an estimate:

$$\begin{aligned}
\hat{\beta}_3 &= \bar{Y}_{11} - \bar{Y}_{01} - \bar{Y}_{10} + \bar{Y}_{00}) \\
&\rightarrow p\beta_3
\end{aligned}$$

as the sample means are consistent estimators for conditional expectations, by the Law of Large Numbers.

Part 3

Ex 1.

(a) First notice that U_i is orthogonal to any linear function of 1, $\tilde{Z}_i$, $Z_{1,i}$, and $Z_{2,i}$, that $V_{1,i}$ is orthogonal to any linear function of 1, $\tilde{Z}_i$, and $Z_{2,i}$, and that $V_{2,i}$ is orthogonal to any linear function of 1, $\tilde{Z}_i$, and $Z_{1,i}$.

Using this we have that

$$\mathrm{Cov}(Y_i, Z_{1,i}) = \mathrm{Cov}(\beta_0 + \beta_1 \tilde{Z}_i + U_i, \tilde{Z}_i + V_{1,i}) = \beta_1 \mathrm{Var}[\tilde{Z}_i]$$

and

$$\mathrm{Cov}(Y_i, Z_{2,i}) = \mathrm{Cov}(\beta_0 + \beta_1 \tilde{Z}_i + U_i, \tilde{Z}_i + V_{2,i}) = \beta_1 \mathrm{Var}[\tilde{Z}_i]$$

and

$$\mathrm{Cov}(Z_{1,i}, Z_{2,i}) = \mathrm{Cov}(\tilde{Z}_i + V_{1,i}, \tilde{Z}_i + V_{2,i}) = \mathrm{Var}[\tilde{Z}_i]$$

We also have the variances

$$\mathrm{Var}[Y_i] = \mathrm{Var}[\beta_0 + \beta_1 \tilde{Z}_i + U_i] = \beta_1^2 \mathrm{Var}[\tilde{Z}_i] + \mathrm{Var}[U_i]$$

and

$$\mathrm{Var}[Z_{1,i}] = \mathrm{Var}[\tilde{Z}_i + V_{1,i}] = \mathrm{Var}[\tilde{Z}_i] + \mathrm{Var}[V_{1,i}]$$

and

$$\mathrm{Var}[Z_{2,i}] = \mathrm{Var}[\tilde{Z}_i + V_{2,i}] = \mathrm{Var}[\tilde{Z}_i] + \mathrm{Var}[V_{2,i}]$$

Putting everything together we get

$$\begin{aligned}\Sigma_{Y,Z_1,Z_2} &= \begin{bmatrix} \mathrm{Var}[Y_i] & \mathrm{Cov}(Y_i, Z_{1,i}) & \mathrm{Cov}(Y_i, Z_{2,i}) \\ \cdot & \mathrm{Var}[Z_{1,i}] & \mathrm{Cov}(Z_{1,i}, Z_{2,i}) \\ \cdot & \cdot & \mathrm{Var}[Z_{2,i}] \end{bmatrix} = \\ &= \begin{bmatrix} \beta_1^2 \mathrm{Var}[\tilde{Z}_i] + \mathrm{Var}[U_i] & \beta_1 \mathrm{Var}[\tilde{Z}_i] & \beta_1 \mathrm{Var}[\tilde{Z}_i] \\ \cdot & \mathrm{Var}[\tilde{Z}_i] + \mathrm{Var}[V_{1,i}] & \mathrm{Var}[\tilde{Z}_i] \\ \cdot & \cdot & \mathrm{Var}[\tilde{Z}_i] + \mathrm{Var}[V_{2,i}] \end{bmatrix}\end{aligned}$$

(b) Now it is easy to see that we can have β_1 in two ways

$$\beta_1 = \frac{\mathrm{Cov}(Y_i, Z_{1,i})}{\mathrm{Cov}(Z_{1,i}, Z_{2,i})} = \frac{\mathrm{Cov}(Y_i, Z_{2,i})}{\mathrm{Cov}(Z_{1,i}, Z_{2,i})}$$

and so β_1 is identified (because we can write it as a function of the distribution of observables).

(c) Two natural choices here are either

$$\hat{\beta}_1 = \frac{\widehat{\text{Cov}}(Y_i, Z_{1,i})}{\widehat{\text{Cov}}(Z_{1,i}, Z_{2,i})} = \frac{\frac{1}{n}\sum_{i=1}^n Y_i Z_{1,i} - \left(\frac{1}{n}\sum_{i=1}^n Y_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_{1,i}\right)}{\frac{1}{n}\sum_{i=1}^n Z_{2,i} Z_{1,i} - \left(\frac{1}{n}\sum_{i=1}^n Z_{2,i}\right)\left(\frac{1}{n}\sum_{i=1}^n Z_{1,i}\right)}$$

Or the version swapping $Z_{1,i}$ and $Z_{2,i}$.

The general approach for consistency (at least in the linear setting) will be to write our estimator as a function of sample means (as we did for $\hat{\beta}_1$), then use the Law of Large Numbers to show that each sample mean will converge in probability to its expectation, and, finally, invoke the continuous mapping theorem (aka Slutsky part 1) to conclude that our function of the sample means is converging to the function applied to the expectations.

We have

$$\frac{1}{n}\sum_{i=1}^n Y_i Z_{1,i} \xrightarrow{p} E\left[Y Z_1\right]$$

by the Law of Large Numbers. The same for all the other sample means in $\hat{\beta}_1$. Therefore, by the Slustky Theorem part 1, we have that

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^n Y_i Z_{1,i} - \left(\frac{1}{n}\sum_{i=1}^n Y_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_{1,i}\right)}{\frac{1}{n}\sum_{i=1}^n Z_{2,i} Z_{1,i} - \left(\frac{1}{n}\sum_{i=1}^n Z_{2,i}\right)\left(\frac{1}{n}\sum_{i=1}^n Z_{1,i}\right)} \xrightarrow{p} \frac{E\left[Y Z_1\right] - \left(E\left[Y\right]\right)\left(E\left[Z_1\right]\right)}{E\left[Z_2 Z_1\right] - \left(E\left[Z_2\right]\right)\left(E\left[Z_1\right]\right)}$$

And that is just

$$\frac{E\left[Y Z_1\right] - \left(E\left[Y\right]\right)\left(E\left[Z_1\right]\right)}{E\left[Z_2 Z_1\right] - \left(E\left[Z_2\right]\right)\left(E\left[Z_1\right]\right)} = \frac{\text{Cov}(Y, Z_1)}{\text{Cov}(Z_2, Z_1)} = \beta_1$$

Important point, to apply our Law of Large Number we need to work with sums of iid terms, so the version we are using is not strong enough to handle, for instance

$$\frac{1}{n}\sum_{i=1}^n \left(Y_i - \frac{1}{n}\sum_{i'=1}^n Y_{i'}\right)\left(Z_{1,i} - \frac{1}{n}\sum_{i'=1}^n Z_{1,i'}\right)$$

because the terms inside the outer summation are not iid.

Part 4

We want to show that $\hat{\alpha}\hat{\Sigma}\hat{\alpha}' \xrightarrow{p} \alpha\Sigma\alpha$.

First, $\hat{\alpha}^{-1} \xrightarrow{p} \alpha^{-1}$. By LLN:

$$\hat{\alpha}^{-1} = \frac{1}{n}\sum_{i=1}^n X_i X_i' \xrightarrow{p} E(X_1 X_1')$$

Next, $\hat{\Sigma} \xrightarrow{p} \Sigma$. Since

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^n U_i^2 X_i X_i' + \underbrace{\frac{1}{n}\sum_{i=1}^n (\hat{U}_i^2 - U_i^2) X_i X_i'}_{\equiv Q_n}$$

and by LLN,

$$\frac{1}{n}\sum_{i=1}^n U_i^2 X_i X_i' \xrightarrow{p} \Sigma,$$

it suffices to show (by Slutsky) that Q_n converges in probability to a matrix of zeros.

Notice that

$$\hat{U}_i = Y_i - X_i'b = U_i - X_i'(b - \beta).$$

Then

$$\hat{U}_i^2 - U_i^2 = -2U_i X_i'(b - \beta) + (b - \beta)' X_i X_i'(b - \beta)$$

so taking the (j, k) entry of the matrix Q_n:

$$\begin{aligned}
(Q_n)_{jk} &= \frac{1}{n}\sum_{i=1}^{n}(-2U_i X_i'(b - \beta) + (b - \beta)' X_i X_i'(b - \beta)) X_{ij} X_{ik} \\
&= -2\left(\frac{1}{n}\sum_{i=1}^{n} U_i X_i' X_{ij} X_{ik}\right)(b - \beta) + (b - \beta)'\left(\frac{1}{n}\sum_{i=1}^{n} X_{ij} X_{ik} X_i X_i'\right)(b - \beta).
\end{aligned}$$

Assume each of the inner terms converges in probability to something finite. We know that $b - \beta$ converges in probability to the zero vector as $n \to \infty$. Applying Slutsky, it follows that $(Q_n)_{jk} \xrightarrow{p} 0$ as $n \to \infty$.

To finish, see that $\hat{\Lambda} = (\hat{\alpha}^{-1})^{-1}\hat{\Sigma}((\hat{\alpha}^{-1})^{-1})'$ is a continuous function of $\hat{\alpha}^{-1}$ and $\hat{\Sigma}$. Applying Slutsky, it follows that $\hat{\Lambda} \xrightarrow{p} (\alpha^{-1})^{-1}\Sigma((\alpha^{-1})^{-1})' = \Lambda$.