Econ 6200: Econometrics II
Prelim, March 20^{th}, 2025

This exam consists of ten questions, not of equal length or difficulty, grouped into two exercises. Keep in mind that you can skip questions. Each question is worth 10 points. Remember to always explain your answer.

Good luck!

1. (Not the same question as last year!) In rank-rank regression, covariate and outcome are expressed in terms of their rank in their respective marginal distribution. Rank-rank regression is popularly applied in analysis of intergenerational mobility, e.g. by regressing income percentiles of children (observed at adult age) on those of their parents.

Question: In an imagined future, I collected observations on $n = 999$ children and their mothers. Each mother was assigned an outcome $X_i \in \{1, \ldots, 999\}$ based on household income, where $X_i = 1$ corresponded to the poorest mother and so on. There was no overlap in parents and no tie in parental household income, so that X_i took each value from 1 to 999 exactly once. Similarly, each child was assigned a rank position Y_i. However, by the time the kids were adults, perfect equality had been achieved, and therefore every Y_i took the same value, normalized to equal 500.

1.1 In the regression of Y on X, what values did the estimated intercept and slope take?

1.2 What value did R^2 take? (Trick question!)

1.3 I then also attempted to run the reverse regression of X on Y. Explain why the coefficient estimators were not well-defined.

1.4 Explain why the fitted values $\hat{X}$ *are* defined. What value did they take? What value did R^2 take?

2 Consider the simple linear regression model

$$Y = \alpha + \beta X + \varepsilon,$$

where all variables are scalars and where we assume that $\mathbb{E}(\varepsilon \mid X) = 0$. Under that condition, one not only has the moment condition

$$\mathbb{E}\big(X(Y - \alpha - \beta X)\big) = 0,$$

but also

$$\mathbb{E}\big(X^2(Y-\alpha-\beta X)\big)=0.$$

What happens if one uses it? To investigate this, consider using $Z=(X,X^2)^\top$ as instruments. (Assume throughout that data are i.i.d.)

To keep matrix algebra tractable, suppose that we do regression without a constant, i.e. we know $\alpha=0$ and use

$$\boldsymbol{X}=\begin{pmatrix}X_1\\ \vdots\\ X_n\end{pmatrix},\quad \boldsymbol{Z}=\begin{pmatrix}X_1 & X_1^2\\ \vdots & \vdots\\ X_n & X_n^2\end{pmatrix}.$$

2.1 Verify matrix algebraically that the Two-Stage Least Squares (TSLS) estimator simplifies to $\hat\beta=\mathbb{E}_n(XY)/\mathbb{E}_n(X^2)$. (If you cannot show the result, assume it henceforth.)

Reminders:

1. $\hat\beta=\big(\boldsymbol{X}^\top\boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top\boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{Y}$,
2. For matrix $\boldsymbol{A}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$, we have $\boldsymbol{A}^{-1}=\frac{1}{\det\boldsymbol{A}}\begin{pmatrix}d & -b\\ -c & a\end{pmatrix}$.

2.2 Given this simplification, the TSLS estimator has a simple interpretation. Explain.

2.3 Without algebra, this finding is obvious from the interpretation that gave TSLS its name. What do I mean here? Use this insight to state what the TSLS estimator will be for $Z=(X,X^2,X^3,X^4)^\top$.

2.4 Returning to $Z=(X,X^2)^\top$, consider now the linear GMM estimator with general weighting matrix

$$\boldsymbol{W}=\begin{pmatrix}w_{11} & w_{12}\\ w_{12} & w_{22}\end{pmatrix}.$$

If I assume homoskedasticity (i.e., $\mathrm{var}(\varepsilon\mid X=x)=\sigma^2$ for all x) and use a weighting matrix that is efficient given this constraint, do we recover the previous simplification? If yes, please prove it. If no, can you give an intuition for why the simple estimator might now be asymptotically inefficient?

2.5 Do we also recover this simplification *without* homoskedasticity? If yes, please prove it. If no, can you give an intuition for why the simple estimator might now be asymptotically inefficient?

2.6 You should find that, under certain assumptions, theorems from class assert efficiency advantages of the estimator discussed here. Why, then, is the estimator not commonly used?

Brief Answers

1.1 The estimated equation was $\hat{y} = 500 + 0 \cdot x$.

1.2 $R^2 = 1 - SSR/SST = 1 - 0/0 = ?$, i.e. not defined.

1.3 The projection coefficients are not unique because, with the data at hand, the problem $\min_{b_0,b_1} \sum_i (x_i - b_0 - b_1 y_i)^2$ is solved by any (b_0, b_1) s.t. $b_0 + 500 \cdot b_1 = 500$.

1.4 The fitted values are unique and all equal 500.

The question about R^2 was not entirely clear. In my book, the official definition of R^2 is $1 - SSR/SST$, i.e. proportion of variation "explained" by fitted values, and is then defined as long as fitted values are. It here takes value 0. If you think of R^2 for simple regression as squared correlation coefficient, then that's not defined here (it involves dividing zero by zero). I advise to think of the latter as a characterization (and should maybe add a remark to that effect on slides), but I accept answers that coherently argue R^2 is not defined.

2.1 The following algebra is on the chatty side for clarity. I accepted all sorts of notational shortcuts.

Because elements of the TSLS expression repeat, you want to divide and conquer. Start with

$$\begin{aligned}
(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} &= \left(\begin{pmatrix} X_1 & \cdots & X_n \\ X_1^2 & \cdots & X_n^2 \end{pmatrix} \begin{pmatrix} X_1 & X_1^2 \\ \vdots & \vdots \\ X_n & X_n^2 \end{pmatrix} \right)^{-1} \\
&= \begin{pmatrix} \sum_i X_i^2 & \sum_i X_i^3 \\ \sum_i X_i^3 & \sum_i X_i^4 \end{pmatrix}^{-1} \\
&= \frac{1}{\det(\cdot)} \begin{pmatrix} \sum_i X_i^4 & -\sum_i X_i^3 \\ -\sum_i X_i^3 & \sum_i X_i^2 \end{pmatrix},
\end{aligned}$$

where we do not evaluate the determinant because we spotted that it will cancel out (me not spelling it out in the question was meant as a hint). Also,

$$\begin{aligned}
\boldsymbol{X}^\top \boldsymbol{Z} &= \begin{pmatrix} X_1 & \cdots & X_n \end{pmatrix} \begin{pmatrix} X_1 & X_1^2 \\ \vdots & \vdots \\ X_n & X_n^2 \end{pmatrix} = \left(\textstyle\sum_i X_i^2, \sum_i X_i^3\right) \\
\implies \boldsymbol{X}^\top \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} &= \frac{1}{\det(\cdot)} \left(\textstyle\sum_i X_i^2, \sum_i X_i^3\right) \begin{pmatrix} \sum_i X_i^4 & -\sum_i X_i^3 \\ -\sum_i X_i^3 & \sum_i X_i^2 \end{pmatrix} \\
&= \frac{1}{\det(\cdot)} \left(\textstyle\sum_i X_i^2 \sum_i X_i^4 - \sum_i X_i^3 \sum_i X_i^3, 0\right) \\
\boldsymbol{Z}^\top \boldsymbol{X} &= \begin{pmatrix} \sum_i X_i^2 \\ \sum_i X_i^3 \end{pmatrix} \\
\boldsymbol{Z}^\top \boldsymbol{Y} &= \begin{pmatrix} \sum_i X_i Y_i \\ \sum_i X_i^2 Y_i \end{pmatrix}
\end{aligned}$$

Next, anticipating cancellation of determinants,

$$\begin{aligned}
\hat{\beta} &= \left(\boldsymbol{X}^\top \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{Z}(\boldsymbol{Z}^\top \boldsymbol{Z})^{-1} \boldsymbol{Z}^\top \boldsymbol{Y} \\
&= \left(\left(\textstyle\sum_i X_i^2 \sum_i X_i^4 - \sum_i X_i^3 \sum_i X_i^3, 0\right) \begin{pmatrix} \sum_i X_i^2 \\ \sum_i X_i^3 \end{pmatrix}\right)^{-1} \\
&\qquad \left(\textstyle\sum_i X_i^2 \sum_i X_i^4 - \sum_i X_i^3 \sum_i X_i^3, 0\right) \begin{pmatrix} \sum_i X_i Y_i \\ \sum_i X_i^2 Y_i \end{pmatrix} \\
&= \left(\left(\textstyle\sum_i X_i^2 \sum_i X_i^4 - \sum_i X_i^3 \sum_i X_i^3\right) \cdot \sum_i X_i^2\right)^{-1} \\
&\qquad \left(\textstyle\sum_i X_i^2 \sum_i X_i^4 - \sum_i X_i^3 \sum_i X_i^3\right) \cdot \sum_i X_i Y_i \\
&= \frac{\sum_i X_i Y_i}{\sum_i X_i^2}
\end{aligned}$$

as desired. To see the last step, notice that all the sums are scalars and so inversion is just "one over" and terms cancel.

Admittedly, getting this right in finite time crucially depends on catching the simplification to 0. I myself did get that at first try, but given my own predilection for linear algebra typos, there was an element of luck involved.

2.2 It's the OLS estimator without a constant (as it should be since we dropped the constant). So TSLS simplifies to OLS here.

2.3 If you think of TSLS as literally a two-stage regression, it becomes intuitively clear. The fitted values from regressing X on (X, X^2) recover X (with projection coefficients $(0, 1, 0)$) and so the regression on fitted values is just a regression on X. That also clarifies that the result will be the same with more powers of X added.

2.4 With homoskedasticity, we have factorization of the weighting matrix and a return to previous algebra.

2.5 Without homoskedasticity, this is *not* true and the algebra becomes very complicated. To the best of my knowledge, there is no neat simplification. In particular, there is no reason to expect simplification to OLS.

With heteroskedasticity, by using a different weighting matrix, we can in principle adjust the weight of observations in which X^2 is large. If X^2 is highly correlated with $\mathrm{var}(\varepsilon \mid X)$, then such a weighting could *in principle* (i.e., on a very literal reading of theorems; see below) be efficient. Think of it as an ad hoc way of doing FGLS.

2.6 Many reasons, here are some:

- While the mean independence assumption in principle generates many moment inequalities, consistency of OLS only crucially depends on uncorrelatedness. Any efficiency gain from using these funky instruments therefore comes at an expense in terms of robustness to slight misspecification.
- The asymptotic efficiency gain comes with extra finite sample noise, and yet the gain will be nonnegligible only if X^2 strongly predicts the error variance. That does not sound like a good deal. And if we believed in this particular heteroskedasticity, there would be more direct implementations of FGLS anyway.
- Realistically, while not perfectly collinear, X and X^2 may be correlated enough (e.g., if $X \geq 0$) that the finite sample performance would be dubious. This may only get worse with higher powers.