ARE212 FINAL EXAM

ETHAN LIGON

This is the final exam for ARE212, covering material from the second half of the course taught in Spring 2022. The exam is "take-at-home"; you may consult any resources you wish in completing it (notes, textbooks, lecture videos, etc.) except for other people. This last restriction isn't easily enforceable; I rely on you to approach this as principled adults who adhere to the Berkeley Honor Code.

More guidance:

- The exam is due at 10am on Tuesday May 10.
- In completing the exam you should develop written arguments (e.g., expressed using LaTeX or pencil and paper). In some cases you may wish to supplement these written arguments with computation, such as Monte Carlo experiments. Should you do so, please provide me with your working, open source, well-documented code. (This last could be links to a github repo, a Jupyter notebook attached to an email, or similar). In any case please be sure that materials you submit are well-organized and clearly documented—if I overlook some file you've sent or can't run it that's on you.
- You are welcome (and even encouraged) to use arguments developed in our `piazza` discussions, but in this case please clearly cite the person and discussion (e.g., "As argued by Michelle in a discussion 'Tests of Normality' (`@53_f11`) the optimal weighting matrix can be written as a function of the unknown parameters.")
- Please email files or links to `ligon@berkeley.edu`.

I. Some short questions

(1) The model of Wright (1934) is "normal" in the sense that all of the underlying random variation comes from two normally distributed random variables. Why are there "only five useful moments"? Why aren't higher order moments involved? If there were three normally distributed random variables how many "useful moments" would there be?

(2) From a sample of N original observations (y_i, X_i) we estimate the vector of parameters β in $\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{u}$, using ordinary least squares, obtaining $\beta^{(OLS)}$.

It's often good practice to use some sort of resampling procedure to estimate the finite sample distribution of our estimators.

Consider two different approaches to this:

Bootstrap: From a sample of N original observations (y_i, X_i) we randomly select N observations (with replacement), and for selection s we estimate the vector of parameters β in $\boldsymbol{y}^{(s)} = \boldsymbol{X}^{(s)}\beta + \boldsymbol{u}^{(s)}$, obtaining $\beta^{(s)}$. We do this N times (for some not very well-motivated reason), so that we wind up with N different estimates of β.

. *Date*: May 9, 2022.

Cross-Validation: From the same sample of N original observations (y_i, X_i) we compute N different "leave-out-one" estimates of β from $\boldsymbol{y}^{(-i)} = \boldsymbol{X}^{(-i)}\beta + \boldsymbol{u}^{(-i)}$, each time obtaining $\beta^{(-i)}$, based on $N-1$ observations. This again gives us N different estimates of β.

How could you use the Bootstrap or Cross-Validation sets of estimates to construct estimates of the standard errors of the OLS estimator $\beta^{(OLS)}$? What relationship might you expect between the sample variance of the Bootstrap estimates and the cross-validation estimates?

(3) We wish to estimate the model $y = X\beta + u$ using a sample of N original, *independent* observations (y_i, X_i). The economic model we're interested in implies $\mathbb{E}(u|X) = 0$ so that we have $\mathbb{E}(\boldsymbol{X}^\top u | X = \boldsymbol{X}) = 0$, but says nothing about the variance of u across observations.
   (a) Use the identifying assumption to derive the GMM estimator of β.
   (b) Using the fact that the observations are independent, what can you say about the structure of the covariance matrix $\mathbb{E}\boldsymbol{X}^\top uu^\top \boldsymbol{X}$?
   (c) Use your result in (b) to construct an estimator of the covariance matrix of your estimator β. How does this compare with the estimator in the homoskedastic case?

II. General Weighted Regressions

Consider the regression $y = X\beta + u$, where $\mathbb{E}uu^\top = \Omega$. We've discussed a variety of estimators of β under different conditions, and asserted that many of these can be expressed as the solution to a set of linear equations written in matrix form as

$$\boldsymbol{T}^\top \boldsymbol{y} = \boldsymbol{T}^\top \boldsymbol{X}\beta, \tag{1}$$

where $(\boldsymbol{X}, \boldsymbol{y})$ are observed data. For each of the following estimators, what is $\boldsymbol{T}$, and under what conditions will the estimator be (i) consistent; and (ii) asymptotically efficient?

(1) OLS
(2) Two-stage least squares
(3) Generalized least squares
(4) Generalized Method of Moments
(5) If you replace the matrix $\boldsymbol{T}$ with the matrix *function* $\boldsymbol{T}(x)$, and replace the linear regression with the non-parametric form $y = m(x) + u$, then the kernel regression (Nadaraya-Watson) estimator can be similarly expressed in the form $\boldsymbol{T}(x)\boldsymbol{y} = \boldsymbol{T}(x)m(x)$. What is $\boldsymbol{T}(x)$ in this case?

III. Clustered sample

We wish to estimate a relationship $y = X\beta + u$, and adopt the identifying assumption that $\mathbb{E}(Z^\top u) = 0$, with $\mathbb{E}Z^\top X = \boldsymbol{Q}$ having full column rank (and more rows than columns).

(1) With an sample of N observations (X_i, y_i, Z_i), use the identifying assumption and the assumption that $\mathbb{E}uu^\top = \Omega$ (with Ω unknown to the econometrician) and construct an asymptotically efficient GMM estimator of β. Describe its limiting distribution.
(2) As in the previous question, but assume $\Omega = \sigma^2 \boldsymbol{I}$ (still unobserved). Compare with the Two-stage least-squares estimator. Compare the efficiency of this estimator with the estimator in (1). Explain.

(3) Now suppose that the sample is based on a clustered-design, so that each of the N observations is drawn from one of $K \leq N/2$ clusters. We assume $\mathbb{E}u_i = \sigma^2$ for all observations. But for any two distinct observations within a cluster we assume $\mathbb{E}u_i u_j = \gamma$, while for two observations *across* clusters we assume $\mathbb{E}u_i u_j = 0$. How could you exploit this structure in a GMM estimator? Compare the asymptotic variance of this estimator with that in (1) and (2).

IV. Uniform Random Variables

Suppose a scalar random variable x is uniformly distributed on the interval $[a, b]$. Then the nth moment of x is given

$$\mathbb{E}(x^n) = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}.$$

(1) Suppose we have a sample of N realizations of x, $\{x_i\}_{i=1}^N$. Suggest a just identified estimator of the unknown parameters a and b based on the expression for the moments of x.
(2) What can you say about the relationship of the smallest value of x_i in your sample relative to your estimate of a?
(3) Now suppose that we are interested in testing the *hypothesis* that x is uniformly distributed. Suggest a practical test based again on the moments of x under the null hypothesis of uniformity.

V. Omitted Variables

You are asked to serve as a referee for a paper submitted to a top field journal. In the submitted paper the researcher uses a sample of size N to estimate a model

$$y = \alpha + \beta x + u.$$

The coefficient β seems to be significantly different from zero, but the researcher is concerned about omitted variable bias, so they also estimate a variety of alternative specifications of the form

$$y = \alpha + \beta x + \gamma w + u,$$

where w is one of a number of other variables that the researcher hypothesizes might have some effect on y as a way of testing the first model.

The researcher finds a particular variable w which enters the regression significantly, and so (i) rejects the first model, concluding that the first estimate of β was in fact affected by omitted variable bias; (ii) declares the augmented regression to be their "preferred specification;" and (iii) proceeds to construct standard t-statistics for β and γ as a way of proceeding with inference.

Peer reviews in economics usually include some "notes for the author." What might your notes say about the paper's approach to omitted variable bias? Comment specifically on each of (i), (ii), and (iii). Try to make your remarks critical yet constructive—what shortcomings do you see, and how might the author address these?

VI. Nested Samples

Consider the linear model $y = X\beta + u$, where X is thought to depend on u, but where we have a set of instruments Z such that $\mathbb{E}Z^\top u = 0$. In this case our observations on y are limited, in that we don't always observe y even when we do observe (X, Z). We can think of this as having two samples, nested in the following way. We have N_1 iid observations on the triple $(y, X, Z)_1$ but $N_2 > N_1$ iid observations on $(X, Z)_2$, with $(X, Z)_1$ (i.e., the observations on X and Z in the first dataset) a subset of $(X, Z)_2$. How can we best make use of all these data?

(1) One econometrician suggests an augmented sort of two-stage-least squares approach, using the richer dataset to estimate a linear relationship $X_2 = Z_2\pi + v$, and thus constructing a "first-stage" prediction equation $\hat{X} = Z\hat{\pi}_2$ which is more precisely estimated that it would be in the usual case in which only data in $(y, X, Z)_1$ was exploited.
   - (a) Continue the argument by substituting into the second stage. What can you say about the properties of the augmented estimator compared to the properties of the usual two-stage least squares estimator?
   - (b) Under what conditions would the augmented estimator be preferred to two-stage least squares on just the sample of N_1 observations?

(2) A second econometrician suggests using the smaller sample to construct a sample moment condition $(Z_1^\top y_1) = (Z_1^\top X_1)b$, and argues that if b in this condition identifies β, then it should be possible to construct $\hat{u}_2 = \hat{y}_2 - X_2 b$, and that for this larger set of observations we must have $\mathbb{E}Z_2^\top \hat{u}_2 = 0$. She argues that these two sets of moment conditions could then be combined into an over-identified optimally-weighted GMM estimator.
   - (a) How would you construct the optimal GMM weighting matrix for this approach? Derive an expression for the asymptotic variance matrix for the estimator b. How does it depend on the larger sample?
   - (b) Comment on this approach. Does the second set of moment conditions add useful information?
   - (c) If you also knew that u was homoskedastic how could you exploit this information? How would the resulting estimator compare with two-stage-least squares? What can you say about the relative efficiency of this estimator versus two-stage least squares?