ARE212 FINAL EXAM

ETHAN LIGON

This is the final exam for ARE212, covering material from the second half of the course taught in Spring 2021. 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 11am on Tuesday May 11.
- 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 Aaron in a discussion 'Tests of Normality' (`@32_f3`) the optimal weighting matrix can be written as a function of a single unknown parameter.")
- Please email files or links to `ligon@berkeley.edu`.
- If you have questions about the final I will look for these on the #are212-econometrics channel of the FY slack instance, but I do not intend to be continuously available on-line, so much better if you can ask questions early!

I. Identifying assumptions for regression

For each of the questions in this section provide a short answer and argument. Note the quality and concision of the argument matters much more than the answer!

(1) Evaluate the truth of following statement: "In the linear regression $y = X\beta + u$ the usual identifying assumption $\mathrm{E}(u|X) = 0$ implies $\mathrm{E}(h(X) \cdot u) = 0$ for any function h satisfying some regularity conditions related to measurability."

(2) Consider the same linear regression $y = X\beta + u$, but now suppose an alternative identifying assumption $\mathrm{E}(X|u) = 0$. Construct a simple estimator based on this alternative. Compare the usual and alternative identifying assumptions; are they equivalent? Is one stronger than the other?

(3) Suppose that $y = f(X) + u$ for some unknown but continuous function f. Suppose we want to use observed data on X to predict outcomes y, and seek a predictor $\hat{y}(X)$ which is "best" in the sense that the mean squared prediction error $\mathrm{E}(y - \hat{y}(X)|X)^2$ is

. *Date*: May 10, 2021.

minimized. What can we say about $\hat{y}$ and its relation to the conditional expectation $\mathrm{E}(y|X)$? Its relation to u?

II. 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?

III. Breusch-Pagan Extended

Consider a linear regression of the form

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

with (y, x) both scalar random variables, where it is assumed that (a.i) $\mathrm{E}(u \cdot x) = \mathrm{E}u = 0$ and (a.ii) $\mathrm{E}(u^2|x) = \sigma^2$.

(1) The condition a.i is essentially untestable, but Breusch and Pagan 1979 argue that one can test a.ii via an auxiliary regression $\hat{u}^2 = c + dx + e$, where the $\hat{u}$ are the residuals from the first regression, and the test of a.ii then becomes a test of $H_0 : d = 0$. Explain both why a.i is untestable, and the logic of the test of a.ii.

(2) Use the two conditions a.i and a.ii to construct a GMM version of the Breusch-Pagan test.

(3) What can you say about the performance or relative merits of the Bruesch-Pagan test versus your GMM alternative?

(4) Suppose that in fact that x is distributed uniformly over the interval $[0, 2\pi]$, and $\mathrm{E}(u^2|x) = \sigma^2(x) = \sigma^2 + \sin(x)$ (with σ^2 a constant greater than or equal to one), thus violating a.ii. What can you say about the performance of the Breusch-Pagan test in this circumstance? Can you modify your GMM test to provide a superior alternative?

(5) In the above, we've considered a test of a specific functional form for the variance of u. Suppose instead that we don't have any prior information regarding the form of $\mathrm{E}(u^2|x) = f(x)$. Discuss how you might go about constructing an extended version of the Breusch-Pagan test which tests for $f(x)$ non-constant.

(6) Show that you can use your ideas about estimating $f(x)$ to construct a more efficient estimator of β if $f(x)$ isn't constant. Relate your estimator to the optimal generalized least squares (GLS) estimator.

IV. Black Lives Matter

Fryer Jr 2019 uses data on encounters between police and civilians of different races in the US to explore how police use of force is related to a civilian's race. While Fryer finds that Black and Hispanic civilians are much more likely to "experience some form of force" from the police and while the probability of being shot by the police is much higher for a civilian who is Black or Hispanic, Fryer's most prominent result is that for "the most extreme use of force—officer-involved shootings—we find no racial differences either in the raw data or when contextual factors are taken into account."

Introducing some notation, let R denote the civilian's race; U some variables observed by the police officer prior to any interaction (e.g., observing "suspicious" behavior) but not the econometrician; D a binary variable indicating the event $(D = 1)$ of an encounter between a given civilian and a police officer; V a set of "contextual factors" related to the encounter and reported by the officer; and S the event that the civilian is shot by the officer. We can then express Fry's finding regarding shootings as not being able to reject either

$$\Pr(S|D = 1, R) = \Pr(S|D = 1) \tag{1}$$

or

$$\Pr(S|D = 1, V, R) = \Pr(S|D = 1, V). \tag{2}$$

(1) Durlauf and Heckman 2020 criticize this conclusion of Fryer's, on the grounds that D may be an endogenous variable. You needn't read their paper, but explain in your own words what sorts of endogeneity might undermine Fryer's conclusion that the probability of being shot by the police doesn't depend on race.

(2) Spell out conditions on (R, S, U, V, D) (perhaps using a causal diagram) which would suffice to interpret (1) and (2) as evidence that there are no racial differences in the victims of police shootings. In particular, what does one need to assume about $\Pr(D = 1|R, U)$?

(3) Consider an alternative ("driving while Black") model in which the police use race as a criterion for stopping or otherwise interacting with a given civilian. Compare the causal structure of this model with your answer to (2). Viewed through the lens of this model, how would one interpret Fry's failure to reject $\Pr(S|D = 1, R) = \Pr(S|D = 1)$?

(4) The Justice Department should care[1] about which (Fry's or the "driving while Black") is the better model. How might one go about testing one against the other?

1. The Equal Protection Clause of the fourteenth amendment to the US constitution is generally interpreted to require "equality before the law" for all persons subject to the jurisdiction of the various states, and the adoption of this amendment shortly after the Civil War is regarded as evidence that it was specifically intended to prevent discrimination against Black Americans.

V. Weighting of Linear IV Estimators

Consider the Linear IV model

$$y = X\beta + u \qquad \mathrm{E}(Z^\top u) = 0.$$

(1) Exploiting the moment condition, under what conditions does the estimator b_{IV} satisfying $Z^\top y = (Z^\top X)b_{IV}$ consistently estimate β?
(2) Suppose that Z has ℓ columns. Suppose that W is a symmetric, $\ell \times \ell$ full rank matrix, with a corresponding estimator b_W satisfying $WZ^\top y = W(Z^\top X)b_W$. Under what conditions will this estimator consistently estimate β?
(3) Describe the GMM criterion function that b_W minimizes.
(4) Consider Hansen's description of the two-stage least squares estimator (Section 12.12). What is W for this estimator? Under what conditions is this the optimally weighted GMM estimator?
(5) $W = I$ for the b_{IV} estimator described above. Under what conditions is b_{IV} the optimally weighted GMM estimator?
(6) For the estimator described in (2), suppose that W is diagonal, with diagonal elements proportional to $(1, 1/2, 1/4, ..., 2^{1-\ell})$. Under what conditions is the estimator b_W the optimally weighted estimator? Can you think of a practical example where the optimal weighting matrix might have this structure?

References

Breusch, Trevor S, and Adrian R Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society:* 1287–1294.

Durlauf, Steven N, and James J Heckman. 2020. An empirical analysis of racial differences in police use of force: a comment. *Journal of Political Economy* 128 (10): 3998–4002.

Fryer Jr, Roland G. 2019. An empirical analysis of racial differences in police use of force. *Journal of Political Economy* 127 (3): 1210–1261.