FINAL REVIEW PROBLEMS 4

1. In the normal linear model, the least squares estimator b is independent of the sum of squared residuals SSR conditional on $X = x$. What do we use this for?

2. If the $n \times 1$ random vector W has the (multivariate) normal distribution $\mathcal{N}(\mu, \Sigma)$, then what is the distribution of the scalar random variable $\alpha' W$ (where α is $n \times 1$ and not random)?

3. In the normal linear model, what is the distribution of the least-squares estimator? What is the distribution of the sum of squared residuals?

4. In the normal linear model, how do we construct a confidence interval for a linear combination of the coefficients?

5. Give an example where we would be interested in a confidence region for two or more linear combinations of the coefficients. Under the normal linear model, how do you determine whether a given point is in such a confidence region?

6. Suppose that you have a way to take independent draws from a standard normal distribution. How could you use this to obtain draws from a chi-square distribution? From a t distribution? From an F distribution?

7. Show that multiplication by an orthogonal matrix preserves the least-squares inner product.

8. How is the QR decomposition used to simplify derivations in the normal linear model?

9. Suppose that we have observations from a random sample, (Y_i, X_i) for $i = 1, \ldots, n$, where Y_i is scalar and X_i is $K \times 1$. Let $E^*(Y_i \mid X_i) = X_i'\beta$. We have been able to obtain a confidence interval for a linear combination $l'\beta$ such that the coverage probability converges to .95 as $n \to \infty$. Briefly discuss how we have been able to relax assumptions in the normal linear model in doing this. How does the confidence interval differ from the one in the normal linear model?

10. Suppose that we have observations from a random sample, (Y_i, X_i) for $i = 1, \ldots, n$, where Y_i is $H \times 1$ and X_i is $H \times K$.

(a) Provide a definition of the linear predictor $E^*(Y_i \mid X_i) = X_i\beta$. (Note that Y_i is a vector.) Provide a formula for β.

(b) Show that the least-squares estimator of β is consistent. What assumptions are needed?

(c) Provide explicit instructions for a research assistant to construct a .95 confidence interval for a linear combination $l'\beta$. What assumptions are needed in order for the coverage probability to converge to .95?

11. Consider longitudinal data on a random sample of N firms followed over T periods.

(a) Provide minimal assumptions in order for the Mundlak estimator to have a limit distribution as $N \to \infty$.

(b) Provide minimal assumptions in order for the first-differences estimator to have a limit distribution as $N \to \infty$.

(c) Provide assumptions under which the Mundlak estimator and the first-differences estimator have the same probability limit as $N \to \infty$.

12. We have developed method of moments inference that exploits the orthogonality conditions in the following framework: $Q_i = R_i\gamma + V_i$, $E(W_i V_i) = 0$. Suppose that we have longitudinal data on earnings for a random sample of n individuals: $(Y_{i1}, \ldots, Y_{iT})$ $(i = 1, \ldots, n)$. Consider the following autoregression model:

$$E(Y_{it} \,|\, Y_{i1}, \ldots, Y_{i,t-1}, A_i) = \lambda + \gamma Y_{i,t-1} + A_i$$

$(t = 2, \ldots, T)$.

(a) Explain how the method of moments framework can be used to provide a (asymptotic, as $n \to \infty$) confidence interval for γ.

(b) Suppose that $T = 3$. Provide a consistent estimator for γ that is a ratio of sample covariances.

(c) Suppose that $T = 4$. Explain how a weight matrix is used to deal with the overidentification that results from there being more than one orthogonality condition for the estimation of the scalar parameter γ.

(d) Provide enough detail on the construction of an optimal weight matrix so that a research assistant, who knows Matlab, could use the data to do the calculations.

(e) Consider the weight matrix that would be optimal under homoskedasticity (so depends only on second moments, not fourth moments). Provide enough detail on the construction of this weight matrix so that a research assistant, who knows Matlab, could use the data to do the calculations. Can this weight matrix be used to obtain a confidence interval for γ that is valid in large samples even if there is heteroskedasticity? Explain.

13. Suppose that we have observations from a random sample, (Y_i, Z_i) for $i = 1, \ldots, n$, where Y_i is a binary random variable that takes on the values 0 and 1. The $K \times 1$ vector X_i is obtained from a function of Z_i that we specify: $X_i = f(Z_i)$.

(a) How is the probit approximation (based on X_i) to the regression function $E(Y_i \,|\, Z_i)$ defined?

(b) Sketch the argument that nonlinear least squares provides a consistent estimator for the coefficients in the probit approximation.

(c) Show how the GMM results can be used to obtain the limit distribution for the estimator in (b).

14. How is the information inequality used to provide intuition for the consistency of the maximum-likelihood estimator?

15. Explain how the GMM results can be used to obtain the limit distribution for the maximum-likelihood estimator.

16. Consider the following panel probit model:

$$Y_{it}^* = X_{it}'\alpha + V_i + \epsilon_{it},$$

with V_i, ϵ_{i1}, …, ϵ_{iT} mutually independent, $V_{it} \sim \mathcal{N}(0, \sigma_v^2)$, and $\epsilon_{it} \sim \mathcal{N}(0, 1)$. We observe

$$Y_{it} = \begin{cases} 1, & \text{if } Y_{it}^* \geq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Suppose that you have maximum-likelihood estimates of α and σ_v. Discuss how they can be used to construct an estimate of the (predictive) effect of X_{it} on Y_{it}.

17. Suppose that you have observations from a random sample, (Y_{i1}, Y_{i2}) for $i = 1, \ldots, n$, where Y_{i1} and Y_{i2} are scalar. Consider the 3×1 vector $\hat{\sigma}$ formed from the sample variance of Y_1, the sample covariance between Y_1 and Y_2, and the sample variance of Y_2.

(a) Provide the limit distribution of $\hat{s}$.

(b) Provide a consistent estimator for the asymptotic covariance matrix of $\hat{s}$. Give enough detail so that a research assistant, who knows Matlab, could use the data to do the calculations.

18. You have the data from a random sample, Y_i for $i = 1, \ldots, n$, where the vector Y_i is 3×1: $Y_i' = (\, Y_{i1} \quad Y_{i2} \quad Y_{i3} \,)$. Let $\hat{\sigma}$ denote the 6×1 vector of distinct sample covariances, formed from the lower triangle of the sample covariance matrix

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})(Y_i - \bar{Y})'.$$

Consider the following variance-components model:

$$Y_{it} = \mu_t + V_i + U_{it} \qquad (t = 1, 2, 3),$$

where the latent random variables V_i, U_{i1}, U_{i2}, U_{i3} are all mutually independent with mean 0. Furthermore, the variance of U_{it} is constant:

$$\mathrm{Var}(U_{it}) = \sigma_U^2 \qquad (t = 1, 2, 3).$$

Let σ_V^2 denote the variance of V_i. (μ_t is a parameter, giving the mean of Y_{it}.)

(a) Explain how the minimum distance framework can be applied to obtain a confidence interval for σ_V^2.

(b) Explain how to obtain a confidence interval for the variance ratio

$$\rho = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_U^2}.$$

Provide enough detail so that a research assistant, who knows Matlab, could use the data to do the calculation.

(c) Consider the weight matrix for minimum distance that would be optimal if $Y_i \sim \mathcal{N}(\mu, \Sigma)$. Could this weight matrix be constructed just using the statistics $\bar{Y}$ and $\hat{\sigma}$? Explain. (It is not necessary to provide an explicit formula for this weight matrix.)

19. Suppose that you have data from a random sample, (Y_i, X_i) for $i = 1, \ldots, n$, where Y_i is scalar and X_i is $K \times 1$. Assume that

$$E(Y_i \,|\, X_i) = X_i'\beta, \quad \mathrm{Var}(Y_i \,|\, X_i) = \exp(X_i'\gamma).$$

Let

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} X_1' \\ \vdots \\ X_n' \end{pmatrix}, \quad b = (X'X)^{-1}X'Y, \quad e = Y - Xb.$$

Consider the following nonlinear least-squares estimator for γ:

$$\hat{\gamma} = \arg\min_a \sum_{i=1}^{n} [e_i^2 - \exp(X_i'a)]^2.$$

(a) Explain how the GMM framework can be used to obtain a limit distribution for $\hat{\gamma}$.

(b) If β were known, we could use $U_i^2 = (Y_i - X_i'\beta)^2$ in the nonlinear least-squares estimator:

$$\hat{\gamma} = \arg\min_a \sum_{i=1}^{n} [U_i^2 - \exp(X_i'a)]^2.$$

Does this estimator have the same limit distribution as the estimator in (a) that uses e_i^2? Explain.

20. How can an instrumental variable model be used to deal with omitted variable bias? What are the requirements on the instrumental variable? How can the orthogonality condition ($\perp$) framework be applied to obtain inferences?

21. In what sense is two-stage least squares an optimal estimator?

22. Consider a job training program that is intended to raise the future earnings of the participants. We want to evaluate the effectiveness of the program. The target population consists of high school dropouts. A random sample of size $n = 5000$ is taken from this population and everyone in the sample is invited to participate in the program, which takes six months. In addition, 2500 of the sample members are chosen at random and offered a subsidy of \$500 per month if they participate in the program. For each of the n sample members, we observe B, S, T, and Y: B = earnings for the individual in 1970, before the program took place; $S = 1$ if the individual is offered the subsidy, $S = 0$ otherwise; $T = 1$ if the individual enrolls in the program, $T = 0$ otherwise; Y = earnings for the individual in 1975, two years after the program took place. Suppose that everyone who enrolls in the program completes it.

(a) Define a treatment effect of the program on earnings. You may assume that the treatment effect is the same for all the individuals.

(b) Consider two least-squares regressions: (i) Y on a constant and T; (ii) Y on a constant, T, and B. Which of these regressions would be better for estimating the treatment effect in (a)? Explain your reasoning. Under what conditions would the preferred regression provide a consistent estimate of the treatment effect?

(c) Explain how you could use the information on the subsidy to construct an instrumental variables estimator of the treatment effect. Sketch the argument that motivates this estimator. What are the advantages and disadvantages of this estimator relative to your preferred regression estimator in (b)?

23. Explain how random assignment deals with omitted variable bias. Explain how panel data can be used to deal with omitted variable bias. What are the advantages and disadvantages of these two approaches to dealing with omitted variable bias?

24. Consider estimating the demand function in the market for Frozen Orange Juice Concentrate (FOJC) in the U.S. There are n annual observations on market price (p) and quantity (q). In addition, you have data on annual Florida rainfall in inches (r) and on U.S. per capita income (m).

(a) A least-squares regression of $\log q$ on $\log p$ gives a slope coefficient b. Is b likely to be biased upward or downward as an estimate of the demand elasticity? Explain.

(b) Build a model that incorporates all the available data (i.e., p, q, r, and m). Explain the role played by each of the variables and the assumptions you are using.

(c) Show how the model in (b) motivates a two-stage least squares estimator for the demand elasticity.