University of Illinois
Spring 2010

Dept. of Economics
Sung Y. Park

ECON 507

Problem Set 1

1. Expand the matrix product $\mathbf{X} = \{[\mathbf{AB} + (\mathbf{CD})'][(\mathbf{EF})^{-1} + \mathbf{GH}]\}'$. Assume that all matrices are square and that $\mathbf{E}$ and $\mathbf{F}$ are nonsingular.

2. Prove that for $nK \times 1$ column vectors $\mathbf{x}_i$, $i = 1, 2, \cdots, n$ and some nonzero vector $\mathbf{a}$,
$$\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{a})(\mathbf{x}_i - \mathbf{a})' = \mathbf{X}'\mathbf{M}^0\mathbf{X} + n(\bar{\mathbf{x}} - \mathbf{a})(\bar{\mathbf{x}} - \mathbf{a})',$$
where the ith row of $\mathbf{X}$ is $\mathbf{x}_i'$ and $\mathbf{M}^0 = \mathbf{I} - \mathbf{1}\mathbf{1}'/n$.

3. Prove that $tr(\mathbf{A} \otimes \mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B})$.

4. Compute the characteristic roots of
$$\mathbf{A} = \begin{pmatrix} 2 & 4 & 3 \\ 4 & 8 & 6 \\ 3 & 6 & 5 \end{pmatrix}.$$

5. Suppose that $\mathbf{x} = \mathbf{x}(z)$, where z is a scalar. What is $\partial[\mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{B}\mathbf{x}]/\partial z$?

6. For the matrix
$$\mathbf{X}' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 4 & -2 & 3 & -5 \end{pmatrix}$$
compute $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{M} = (\mathbf{I} - \mathbf{P})$. Verify that $\mathbf{MP} = \mathbf{0}$. Let
$$\mathbf{Q} = \begin{pmatrix} 1 & 3 \\ 2 & 8 \end{pmatrix}.$$
   (a) Compute the $\mathbf{P}$ and $\mathbf{M}$ based on $\mathbf{XQ}$ instead of $\mathbf{X}$.
   (b) What are the characteristic roots of $\mathbf{M}$ and $\mathbf{P}$?

7. Suppose that $\mathbf{A}$ is $n \times n$ matrix of the form
$$\mathbf{A} = (1 - \rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}',$$
where $\mathbf{1}$ is a column ones and $0 < \rho < 1$. Write down the format of $\mathbf{A}$ explicitly for $n = 4$. Find all the characteristic roots and vector of $\mathbf{A}$.

8. Suppose that you have two independent unbiased estimators of the same parameter θ, say $\hat{\theta}_1$ and $\hat{\theta}_2$, with different variances v_1 and v_2. What linear combination $\hat{\theta} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2$ is the minimum variance unbiased estimator of θ?

9. Prove that the least squares intercept estimator in the classical regression model is the minimum variance linear unbiased estimator.

10. Suppose that the classical regression model applied but that the true of the constant is zero. Compare the variance of the least squares slope estimator computed without a constant term to that of the estimator computed with an unnecessary constant term.

11. Suppose that you estimate a multiple regression first with then without a constant. Whether the R^2 is higher in the second case than the first will depend in part on how it is computed. Using the (relatively) standard method

$$R^2 = 1 - \frac{\mathbf{e}'\mathbf{e}}{\mathbf{y}'\mathbf{M}^0\mathbf{y}},$$

which regression will have a higher R^2?

12. Let e_i be the ith residual in the ordinary least squares regression of $\mathbf{y}$ on $\mathbf{X}$ in the classical regression model, and let ϵ_i be the corresponding true disturbance. Prove that $plim(e_i - \epsilon_i) = 0$.

13. In the least squares regression of $\mathbf{y}$ on a constant and $\mathbf{X}$, to compute the regression coefficients on $\mathbf{X}$, we can first transform $\mathbf{y}$ to deviations from the mean $\bar{\mathbf{y}}$ and, likewise, transform each column of $\mathbf{X}$ to deviations from the restrictive column mean; second, regress the transformed $\mathbf{y}$ on the transforms $\mathbf{X}$ without a constant. Do we get the same result if we only transform $\mathbf{y}$? What if we only transform $\mathbf{X}$?

14. For the classical normal regression model $y = X\beta + \epsilon$ with no constant term and k regressors, what is

$$\text{plim} F[K, n-k] = \text{plim} \frac{R^2/K}{(1-R^2)/(n-K)}$$

assuming that the true value of β is zero? What is the exact expected value?

15. Consider the equality between the two forms of the information matrix

$$\mathcal{I}(\theta_0) = E_0\left(\frac{\partial \log f(Y|\theta_0)}{\partial \theta}\frac{\partial \log f(Y|\theta_0)}{\partial \theta'}\right) \quad \text{and}$$

$$\mathcal{J}(\theta_0) = -E_0\left(\frac{\partial^2 \log f(Y|\theta_0)}{\partial\theta\partial\theta'}\right),$$

where E_0 denotes the expectation with respect to the true distribution of Y. It is very reasonable to test the equality between the two forms of the information matrix. It is well-known information matrix test. Consider $\mathcal{C}(\theta) = \mathcal{I}(\theta) - \mathcal{J}(\theta)$. The quantities of interest are the elements of the $\mathcal{C}(\theta)$. Let q be the number of such quantities, which are indexed by l so that $l = 1, \cdots, q$ with $q \le p(p+1)/2$, where p denotes the dimension of parameter vector, i.e., $\theta \in \Theta \subset \mathcal{R}^p$. The l-th quantity corresponding to the (i, j)-th element is

$$a_l(y, \theta) = \frac{\partial \log f}{\partial\theta_i}\frac{\partial \log f}{\partial\theta_j} + \frac{\partial^2 \log f}{\partial\theta_i\partial\theta_j}, \quad l = 1, \cdots, q.$$

Let $a(y, \theta)$ denote the vector with components $a_l(y, \theta)$, $l = 1, \cdots, q$. Let

$$\hat{a}_n(\hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n} a(y_i, \hat{\theta}_n),$$

where $\hat{\theta}_n$ is the pseudo maximum likelihood estimator of θ. Now consider the model

$$y_i = \theta_1 + u_i,$$

where the u_i's are independently and normally distributed with zero mean and common variance θ_2

(a) What is the log-likelihood for the i-th observation?

(b) Derive and interpret $\hat{a}_n(\hat{\theta}_n)$.

(c) It is well-known that, under the null hypothesis, the statistic $\sqrt{n}\hat{a}_n(\hat{\theta}_n)$ is asymptotically distributed as a centered normal distribution with variance covariance matrix

$$\mathbf{V} = E_0\left(a(Y, \theta_0) + \mathbf{A}(\theta_0)\mathcal{J}^{-1}(\theta_0)\frac{\partial \log f(Y|\theta_0)}{\partial\theta}\right)$$
$$\left(a(Y, \theta_0) + \mathbf{A}(\theta_0)\mathcal{J}^{-1}(\theta_0)\frac{\partial \log f(Y|\theta_0)}{\partial\theta}\right)',$$

where

$$\mathbf{A}(\theta_0) = E_0\frac{a(Y, \theta_0)}{\partial\theta'}.$$

Derive the test statistic for the information matrix equality.

16. Consider the data generating process (DGP) characterized by $\mathbf{Y} = \mathbf{X}\beta + \epsilon$, where $\epsilon \sim N(\mathbf{0}_{25}, \sigma^2 \mathbf{I}_{25})$. The fixed $\mathbf{X}$ matrix will be (25×2), where the first column will simple be a column of 1's, and the second column will be 25 iid outcomes from a uniform distribution on the interval $[0, 10]$. The β vector will equal to $\beta = (1, 2)'$. All experiment will be conducted conditional on the $\mathbf{X}$-matrix you generated in the previous step.

    (a) Let $\sigma^2 = 1$; generate 500 outcomes of the LS estimator of β. Calculate the empirical distribution function for $\hat{\beta}_2$. Report the exercise for $\sigma^2 = 1$ and $\sigma^2 = 10$. Discuss the relationship between σ^2 and the LS estimator of β_2.

    (b) Repeat the preceding simulation exercise, calculating EDFs for the estimator S^2 of σ^2 given by $S^2 = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})/(n - k)$. Discuss the relationship between σ^2 and the estimator S^2 of σ^2.

    (c) Can you show the consistency result of $\hat{\beta}$ by simulation (changing the sample size)? Discuss how fast does $\hat{\beta}$ converge to β.