Spring, 2019
Hidehiko Ichimura

Review Questions I Answers

In the first half of the spring term, we have studied the linear regression model, the OLS estimator for the coefficients in the linear regression model, and the basics of asymptotic analysis.

The main points I want you to understand are

1. *how we can specify the linear regression model using various forms as regressors such as cross terms and dummy variables,*
2. *what the conditions are, under which the coefficient have a causal interpretation,*
3. *what the OLS estimator is, when it is well defined, and why the method is conditionally unbiased and consistent at an intuitive level,*
4. *concrete conditions under which the estimators are consistent and asymptotically normal,*
5. *how to conduct inferences using the finite sample results and asymptotic results,*
6. *how to conduct hypothesis tests using the finite sample results and asymptotic results, and*
7. *what are the concrete empirical situations under which various assumptions do not hold, and*
8. *what happens to the OLS estimator when the basic assumptions do not hold.*

I want to emphasize that there are common set of tools underlying these issues. They are

1. *matrix algebra,*
2. *the concept of conditional expectation operator, and*
3. *asymptotic analysis.*

They continue to play critical roles for instrumental variable method, panel data analysis, maximum likelihood estimation, non-parametric and semiparametric analysis, etc. Tools are explained in class and discussed in the textbook and also in the lecture note. Please make sure to understand them. Once you obtain the tools, the eight issues above become applications of the tools. In turn, understanding the tools require examination of how they can be applied.

Of course concrete empirical issues motivate all these analysis. Without knowing the actual empirical issues, why the theoretical problems examined are interesting will not be understood. Some examples of empirical issues are given in the last part of the questions below.

Basics

Here is a list of questions you may be able to use to review the material covered in the first half of the course. The course covered some more material, but going through the questions will give you an idea about the outline of the course.

1. *What is the interpretation of the coefficient in the standard linear regression model? What are the conditions under which this can be done?*

A short answer is, under the correct specification assumption (additive error term and the assumption that $E(u|X) = 0$), the coefficient can be used to make a causal inference of a change in any of the regressors. Whether a single coefficient is involved or more coefficients need to be considered simultaneously, depends on the specification of the model.

To think about this more, let me explain the background in general first and then specialize the discussion to the linear regression model.

We consider the case to examine the average effect of a change in X_2 from x_2 to x_2' holding X_1 and the unobserved variable, then under the condition that X_2 and ϵ are indepenent, given X_1, it is possible to estimate this.

To see this, consider the model: $Y = m(X_1, X_2, \epsilon)$ in which X_1, X_2 and Y are observed but ϵ is not observed. Suppose we want to evaluate the effect of changing X_2 from x_2 to x_2' holding X_1 at x_1 so that we want to study

$$m(x_1, x_2', \epsilon) - m(x_1, x_2, \epsilon).$$

Since we don't observe ϵ, we cannot hold ϵ at the same value. The best we can observe is the averaged version, the average treatment effect:

$$E[m(x_1, x_2', \epsilon) - m(x_1, x_2, \epsilon)] = E[m(x_1, x_2', \epsilon)] - E[m(x_1, x_2, \epsilon)].$$

The last expression can be estimated by $E[Y|X_1 = x_1, X_2 = x_2'] - E[Y|X_1 = x_1, X_2 = x_2]$ only under the additional assumption that X_2 and ϵ are independent given X_1. To see this:

$$\begin{aligned}
&E[Y|X_1 = x_1, X_2 = x_2'] - E[Y|X_1 = x_1, X_2 = x_2] \\
&= E[m(X_1, X_2', \epsilon)|X_1 = x_1, X_2 = x_2'] - E[m(X_1, X_2, \epsilon)|X_1 = x_1, X_2 = x_2] \\
&= E[m(x_1, x_2', \epsilon)|X_1 = x_1, X_2 = x_2'] - E[m(x_1, x_2, \epsilon)|X_1 = x_1, X_2 = x_2] \\
&= E[m(x_1, x_2', \epsilon)|X_1 = x_1] - E[m(x_1, x_2, \epsilon)|X_1 = x_1] \\
&= E[m(x_1, x_2', \epsilon) - m(x_1, x_2, \epsilon)|X_1 = x_1].
\end{aligned}$$

We have used the conditional independence assumption to derive the equality in the second to the last expression. The last expression can be integrated over x_1 using the marginal distribution of X_1 to obtain $E[m(x_1, x_2', \epsilon) - m(x_1, x_2, \epsilon)]$.

In the context of linear regression model, $Y = X_1'\beta_1 + X_2'\beta_2 + \epsilon$, the coefficients characterizes the conditional mean function, so they can be used to discuss the average causal effect, under the same assumption. In the simple case that X_2 is not a random vector, but a random variable, it enters in linearly in the model, and there is no interaction term of X_2 and other variables, then the coefficient of X_2 can be interpreted as the average treatment effect of a unit change in X_2.

To see this, note that $E(Y|X_1 = x_1, X_2 = x_2) = x_1'\beta_1 + x_2\beta_2 + E(\epsilon|X_1 = x_1, X_2 = x_2)$, so that, denoting $x_2' = x_2 + 1$,

$$\begin{aligned}
&E[Y|X_1 = x_1, X_2 = x_2'] - E[Y|X_1 = x_1, X_2 = x_2] \\
&\quad = x_2'\beta_2 + E(\epsilon|X_1 = x_1, X_2 = x_2') - x_2\beta_2 - E(\epsilon|X_1 = x_1, X_2 = x_2) \\
&\quad\quad = \beta_2 + E(\epsilon|X_1 = x_1, X_2 = x_2 + 1) - E(\epsilon|X_1 = x_1, X_2 = x_2) = \beta_2.
\end{aligned}$$

The last equality follows under the assumption that ϵ and X_2 are independent given X_1 so that

$$E(\epsilon|X_1 = x_1, X_2 = x_2) = E(\epsilon|X_1 = x_1, X_2 = x_2').$$

Since the linear regression model assumes even stronger condition, i.e. $E(\epsilon|X_1 = x_1, X_2 = x_2) = 0$ for any x_1 and x_2 in the support of (X_1, X_2), by the same argument, the average causal effect of changes in any regressor can be obtained.
In a more complicated case in which X_2 enters in polynomial terms or enter with product terms with other regressors, then the average causal effect of a unit change in X_2 will involve more coefficients but that is the only difference.
Note that what we really need for estimating the average treatment effect of X_2 on Y is the conditional independence of unobserved variables that affect the outcome and X_2 given X_1, but that additive error allows us to weaken that requirement to conditional mean independence.

2. *What is the OLS estimator of the coefficients in the linear regression model and what is the condition under which it exists?*
Let the linear regression model be, for $i = 1, ..., n$,

$$y_i = x_i'\beta + u_i$$

where y_i and ε_i are scalars and x_i and β are vectors of size K. OLS estimator of β (slope coefficients) is defined as the minimizer of the following objective function over b:

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - x_i'b\right)^2.$$

The solution, i.e. the OLS estimator, is uniquely given by $\left(\sum_{i=1}^{n} x_i x_i'\right)^{-1}\sum_{i=1}^{n} x_i y_i$ when $\sum_{i=1}^{n} x_i x_i'$ is invertible.

3. *Why is the OLS residual vector orthogonal to each of the regressors?*
The OLS estimator, denoted $\hat{\beta}$, satisfies the first order condition corresponding to the minimization problem defining the OLS estimator, so that

$$-2\frac{1}{n}\sum_{i=1}^{n} x_i(y_i - x_i'\hat{\beta}) = 0.$$

Since the OLS residual is defined as $\hat{u}_i = y_i - x_i'\hat{\beta}$, the above first order condition implies

$$\sum_{i=1}^{n} x_i\hat{u}_i = 0,$$

which means that $(\hat{u}_1, \ldots, \hat{u}_n)'$ is orthogonal to each element vector of regressors, $(x_{j1}, \ldots, x_{jn})'$ for $j = 1, \ldots, K$.

4. *Explain intuitively why the objective function of the OLS estimator identifies the coefficients in the linear model.*
The conditional mean function is the best predictor of Y given X and the linear regression model parametrises the conditional mean function as a linear combination of regressors. The least square objective function mimics the mean square prediction error in a finite sample for a trial slope coefficient. Thus the OLS estimator mimics the smallest prediction error property of the conditional mean function in a finite sample.
Note that by the law of large number (LLN) the objective function converges to $E\left[(y_i - x_i'b)^2\right]$.

When $y_i = x_i'\beta + u_i$ holds,

$$\begin{aligned} &E\left[\left(y_i - x_i'b\right)^2\right] \\ &= E\left[\left(x_i'\beta - x_i'b\right)^2\right] + E\left[\left(x_i'\beta - x_i'b\right)u_i\right] + E\left(u_i^2\right). \end{aligned}$$

When $E(u_i|x_i) = 0$ holds, the expression equals (the first term is written differently for the later argument)

$$\begin{aligned} &E\left[\left(\beta' x_i - b' x_i\right)\left(x_i'\beta - x_i'b\right)\right] + E\left(u_i^2\right) \\ &= E\left[(\beta - b)' x_i x_i' (\beta - b)\right] + E\left(u_i^2\right) \\ &= (\beta - b)' E\left(x_i x_i'\right)(\beta - b) + E\left(u_i^2\right). \end{aligned}$$

Thus the limiting function is minimized at $b = \beta$ if $E(x_i x_i')$ is invertible. So when the four conditions, assumption justifying the LLN, the specification $y_i = x_i'\beta + u_i$, $E(u_i|x_i) = 0$ and invertibility of $E(x_i x_i')$ hold, we can justify consistency of the OLS estimator.

5. *What does the OLS estimator estimate under misspecification?*
   By the same argument as above, by defining $U = Y - E(Y|X)$, so that $E(u|X) = 0$, one can obtain the decomposition

   $$E[(Y - X'b)^2] = E(U^2) + E[(m(X) - X'b)^2].$$

   So minimiation of the left-hand side yields the minimization of the second term in the right-hand side. It implies that the OLS estimator yields the best approximation to the unknown conditional mean function of Y on X using the mean squared loss function, within a class of models expressed as a linear combination of regressors.

6. *What are the sufficient conditions for the OLS estimator to be conditionally unbiased?*
   In order for the OLS estimator to exist, X, the $n \times K$ matrix whose ith row is x_i', needs to have rank K. The additional assumptions are $y_i = x_i'\beta + \varepsilon_i$ (the specification is correct), and that $E(\varepsilon_i|x_1, ..., x_n) = 0$. The latter condition holds if $E(\varepsilon_i|x_i) = 0$ and the sampling is independent. (Can you explain why?)

7. *What are the sufficient conditions for the OLS estimator to be consistent?*

   $y_i = x_i'\beta + \varepsilon_i$, $E(x_i x_i')$ is finite and invertible, $E(x_i\varepsilon_i) = 0$ and random sampling.

8. *What are the sufficient conditions for the OLS estimator to be asymptotically normal?*

   The only difference between establishing consistency and asymptotic normality is whether to use LLN or the central limit theorem (CLT). The additional assumptions required is the finiteness of the second moment of y_i given x_i.

9. *Assume random sampling and the model*

   $$y_i = \alpha + x_i'\beta + \varepsilon_i.$$

   Under the assumptions that justify the OLS estimator to be unbiased, discuss the relationship between $\alpha + x_i'\beta$ and the conditional mean function of the dependent variable y_i conditional on the regressors x_i.

   They are the same because

   $$\begin{aligned} E(y_i|x_i) &= E\left(\alpha + x_i'\beta + \varepsilon_i|x_i\right) \\ &= \alpha + x_i'\beta + E(\varepsilon_i|x_i) = \alpha + x_i'\beta. \end{aligned}$$

10. *Is the OLS estimator consistent and asymptotically normal under heteroskedasticity? Explain your answer.*

    The arguments above show that there is no need for the error term to be homoskedastic so OLS estimator can be consistent and asymptotically normal under heteroskedasticity so long as a LLN is applicable to $n^{-1}\sum_{i=1}^{n} x_i\varepsilon_i$ and a CLT is applicable to $n^{-1/2}\sum_{i=1}^{n} x_i\varepsilon_i$. (Can you verify the Lindeberg condition when there is heteroskedasticity under some general restriction on the form of heteroskedasticity?)

11. *When there is heteroskedasticity what happens to OLS estimator?*

    So long as the rest of the assumptions discussed above hold, the OLS is still unbiased, consistent and asymptotically normal. OLS estimator's asymptotic variance-covariance matrix is no longer the usual form. Since most statistical packages use the form only valid under homoskedasticity, the standard errors reported by the packages are based on an inconsistent estimator of the asymptotic variance-covariance matrix. In addition, the OLS estimator is no longer BLUE.

12. *When is the conditional variance of the OLS estimator of a coefficient small?*

    The conditional variance can be decomposed in the following way (See the lecture note for the explanation.):

    $$V(\hat{\beta}_1|\mathbf{X}) = \frac{N^{-1}\sum_{i=1}^{N}\hat{v}_{1i}^2\sigma^2(x_i)}{N^{-1}\sum_{i=1}^{N}\hat{v}_{1i}^2 N^{-1}\sum_{i=1}^{N}\sigma^2(x_i)} \frac{N^{-1}\sum_{i=1}^{N}\sigma^2(x_i)\cdot N^{-1}}{(1-R_1^2)\hat{V}(x_1)}.$$

    The first term is one if the sample covariance of $\hat{v}_{1i}^2$ and $\sigma^2(x_i)$ for $i = 1, \ldots, N$ is zero, greater than one if it is positive, and less than 1 if it is negative. The rest of the expression consists of 4 factors:

    (a) The sample average of the conditional variances of the dependent variable $N^{-1}\sum_{i=1}^{N}\sigma^2(x_i)$. (conditional variance of $\hat{\beta}_1$ is larger as they are larger)

    (b) Sample variance of the regressor under consideration, $\hat{V}(x_1)$. (conditional variance of $\hat{\beta}_1$ is smaller as it is larger)

    (c) R^2 of the auxiliary regression, $\hat{R}_1$. (conditional variance of $\hat{\beta}_1$ is larger as it is larger)

    (d) The sample size, N. (conditional variance of $\hat{\beta}_1$ is smaller as it is larger)

13. *What should you do if you think there is heteroskedasticity in the context of a linear regression model?*

    There are two ways to approach the problem. Given that the major problem with the OLS estimator is the use of inconsistent variance-covariance matrix estimator, a solution is to estimate the variance-covariance matrix consistently allowing for the heteroskedasticity. Another approach is to model the heteroskedasticity and estimate the coefficient by the GLS estimator or the estimated GLS estimator.

14. *What are the basic assumptions to justifying the OLS estimator to be BLUE?*

    In addition to the conditions discussed above, the variance-covariance matrix of the error terms need to be diagonal with a constant variance.

15. *What happens to the OLS estimator when the error term and some of the regressors are correlated?*

    The OLS estimator will be biased and inconsistent.

16. *Give five cases in which the error term and some of the regressors are correlated.*

    (1) The functional misspecification, (2) measurement error among regressors, (3) simultaneity, (4) sample selection, and (5) presense of lagged dependent variable among regressors and serial correlation in the error term.

17. *Explain three different ways to test the null hypothesis: $C\beta = A$ where C is an $r \times K$ matrix of full row rank.*

    Under homoskedasticity assumption we have discussed Wald approach, Likelihood Ratio approach, and Lagrangean Multiplier approach. Make sure you know what they are and that you can use them with proper degrees of freedom. When there is heteroskedasticity, we have seen how the test statistic needs to be modified for the Wald approach. You should be able to use this as well.

18. *Explain the circumstances in which you would use each one of the three tests.*

    Wald type test is useful when the unrestricted estimator is already computed. LM type test is useful when restricted estimator is much easier to compute that the unrestricted estimator. When both unrestricted and restricted estimators are available, computing the test statistic itself is simple for the likelihood ratio type test.

19. *Let $\hat{b}$ be an estimator of β and that $\sqrt{n}\left(\hat{b} - \beta\right)$ converge in distribution to a normal random vector with zero mean and variance covariance matrix of V. Let $\hat{V}$ be a consistent estimator of V.*

    (a) *How will you construct a 95% confidence interval for $c'\beta$ for a given vector c?*

    Under the conditions specified $\sqrt{n}c'\left(\hat{b} - \beta\right)/\sqrt{c'\hat{V}c}$ converges in distribution to the standard normal random variable. Thus the 95% interval for $c'\beta$ is

    $$c'\hat{b} - 1.96\sqrt{c'\hat{V}c}/\sqrt{n} \leq c'\beta \leq c'\hat{b} + 1.96\sqrt{c'\hat{V}c}/\sqrt{n}.$$

    (b) *How will you construct a Wald type test of $c'\beta = a$ for a given vector c and a constant value a using 5% significance level?*

    If the alternative is $c'\beta \neq a$ then the null is rejected if $\sqrt{n}\left(c'\hat{b} - a\right)/\sqrt{c'\hat{V}c} \leq -1.96$ or $1.96 \leq \sqrt{n}\left(c'\hat{b} - a\right)/\sqrt{c'\hat{V}c}$.

    If the alternative is $c'\beta > a$ then the null is rejected if $1.64 \leq \sqrt{n}\left(c'\hat{b} - a\right)/\sqrt{c'\hat{V}c}$.

    If the alternative is $c'\beta < a$ then the null is rejected if $-1.64 \geq \sqrt{n}\left(c'\hat{b} - a\right)/\sqrt{c'\hat{V}c}$.

    (c) *How will you construct a 95% confidence region for $C'\beta$ for a given ($r \times k$, $r < k$) matrix C?*

    Under the conditions specified, $n\left(C'\left(\hat{b} - \beta\right)\right)'\left[C'\hat{V}C\right]^{-1}C'\left(\hat{b} - \beta\right)$ converges in distribution to the chi-squared distribution with the degrees of freedom equal to the rank of matrix C. Thus the 95% confidence region is

    $$\left\{\beta \in R^k; n\left(C'\left(\hat{b} - \beta\right)\right)'\left[C'\hat{V}C\right]^{-1}C'\left(\hat{b} - \beta\right) \leq h\right\}$$

    where h is a constant value which is defined as $\Pr\left\{\chi \leq h\right\} = 0.95$ and that χ is the chi-squared random variable with degrees of freedom equal to the rank of matrix C.

(d) *How will you construct a Wald type test of $C'\beta = A$ for a given $(r \times k,\ r < k)$ matrix C and a constant value vector $(r \times 1)$ A where rows of C are linearly independent using 5% significance level?*

Using the same notations as above, the null hypothesis is rejected if

$$n\left(C'\hat{b} - A\right)'\left[C'\hat{V}C\right]^{-1}\left(C'\hat{b} - A\right) > h.$$

- *This problem highlights the fact that all the inference problems and hypothesis testing frameworks we use in this course have the same structure. The differences are in the construction of $\hat{b}$ and $\hat{V}$ as well as the meaning of β. This problem also makes clear that if r is two or higher, we lose the sense of direction in testing the null against a particular alternative.*
- *You need to be able to conduct inferences and hypothesis tests in a specific context where an estimator and c (and a) or C (and A) are given.*

20. *How do you estimate the variance covariance matrix of an OLS estimator under homoskedasticity?*

    Using the same notations as in lectures, $\hat{\sigma}^2\left(n^{-1}\sum_{i=1}^n x_i x_i'\right)^{-1}$ where $\hat{\sigma}^2 = n^{-1}\sum_{i=1}^n\left(y_i - x_i'\hat{b}_{OLS}\right)^2$. In place of σ^2, $(n-K)^{-1}\sum_{i=1}^n\left(y_i - x_i'\hat{b}_{OLS}\right)^2$ could be used.

21. *How do you estimate the variance covariance matrix of an OLS estimator under heteroskedasticity?*

    We can estimate it by, $\left(n^{-1}\sum_{i=1}^n x_i x_i'\right)^{-1}\left(n^{-1}\sum_{i=1}^n \hat{u}_i^2 x_i x_i'\right)\left(n^{-1}\sum_{i=1}^n x_i x_i'\right)^{-1}$, where $\hat{u}_i = y_i - x_i'\hat{b}_{OLS}$.

Basics II

Here are more detailed questions.

Asymptotics

The topics we studied include (1) basics such as various concepts of convergence, $O_P(1)$, $o_p(1)$ concepts (2) LLN, CLT.

Please make sure you have clear ideas on how to use $O_P(1)$, $o_p(1)$ concepts.

1. *Explain the difference between different consistency concepts of an estimator.*

   Convergence in probability, almost sure convergence , and convergence in the rth moment for $r \geq 1$ are the three concepts of consistency. Please review your note from the lectures.

2. *Explain the differences among consistency, unbiasedness, and asymptotic unbiasedness of an estimator.*

   Please review your note from the lectures.

3. *Why is asymptotic normality of an estimator a useful property of an estimator?*

   Please review your note from the lectures.

4. *Prove the following:*

(a) $O_P(1) \cdot o_p(1) = o_p(1)$ *always.*

Let $X_n = O_P(1)$ and $Y_n = o_p(1)$. The only two cases $X_n \cdot Y_n \neq o_p(1)$ are when $\|X_n\|$ is large and when $\|Y_n\|$ is not small. But both probabilities are small.

(b) $O_P(1) + O_P(1) = O_P(1)$ *always.*

Let $X_n = O_P(1)$ and $Y_n = O_p(1)$. The only two cases $X_n + Y_n \neq O_p(1)$ are when $\|X_n\|$ is large and when $\|Y_n\|$ is large. But both probabilities are small.

(c) $o_p(1) + o_p(1) = o_p(1)$ *always.*

Let $X_n = o_P(1)$ and $Y_n = o_p(1)$. The only two cases $X_n + Y_n \neq o_p(1)$ are when $\|X_n\|$ is not small or when $\|Y_n\|$ is not small. But both probabilities are small.

5. *Assume that for each i, $X_{ni} = o_p(1)$ or $X_{ni} = O_P(1)$ as $n \to \infty$. In view of the last two results above, an induction argument implies that for any finite J, $\sum_{i=1}^{J} X_{ni} = o_p(1)$ if each $X_{ni} = o_p(1)$ and $\sum_{i=1}^{J} X_{ni} = O_p(1)$ if each $X_{ni} = O_p(1)$. Show this.*

   Repeated application of (b) or (c) finite times yields the result.

6. *However, when J also goes to infinity as $n \to \infty$, either of the results does not hold. Explain this by constructing an example for each case.*

   Consider the case where $X_{ni} = 1$. Then clearly $X_{ni} = O_P(1)$ but $\sum_{i=1}^{n} X_{ni} = \sum_{i=1}^{n} 1 = n \to \infty$.

   Consider the case where $X_{ni} = 1/n$. Then clearly $X_{ni} = o_p(1)$ but $\sum_{i=1}^{n} X_{ni} = \sum_{i=1}^{n} 1/n = 1$ and does not converge to zero.

7. *If for each i, $X_{ni} = o_p(1)$ as $n \to \infty$, then for any finite J, $\sum_{i=1}^{J} X_{ni}/J = o_p(1)$ as $n \to \infty$, but show that $\sum_{i=1}^{n} X_{ni}/n$ is not necessarily $o_p(1)$ as $n \to \infty$. (This is really the same problem as the problem above for $o_p(1)$. However, stating this way may be slightly more surprising.)*

   Let $X_{ni} = 1\{|U_i| < h_n\}/h_n$, where U_i is uniform distribution on $[-0.5, 0.5]$ and that h_n converges to zero deterministically. Then clearly $X_{ni} = o_p(1)$ for each i, but $E(\sum_{i=1}^{n} X_{ni}/n) = E[1\{|U_i| < h_n\}/h_n] = 2$. Variance converges to zero if $nh_n \to \infty$ so that the average converges to 2 in probability.

8. *Consider the following random coefficient model under random sampling:*

$$y_i = x_i'\beta_i$$

   where $x_i = (1, \tilde{x}_i')'$ and $\tilde{x}_i$ and β_i are independent random vectors with finite second moments. Show that when $E(x_i x_i')$ is non-singular, OLS estimator consistently estimates $E(\beta_i)$.

   $y_i = x_i'\beta_i = x_i'E(\beta_i) + \varepsilon_i$ where $\varepsilon_i = x_i'[\beta_i - E(\beta_i)]$. Since $E(\varepsilon_i) = 0$ and $E\{\varepsilon_i x_i\} = 0$, the OLS of y_i on x_i consistently estimates $E(\beta_i)$ under random sampling.

9. *Let $\hat{b}_1$ and $\hat{b}_2$ be estimators of $\beta_0 \in R^K$ and that $\sqrt{n}(\hat{b}_j - \beta_0)$ converges in distribution to $N(0, V_j)$ for $j = 1, 2$.*

   (a) *Derive the asymptotic variances of $\sqrt{n}[g(\beta_j) - g(\beta_0)]$, for a continuously differentiable function g, $g : R^K \to R$ for $j = 1, 2$. (Use the Delta-method covered in Antonio's class.)*

   Using the Delta method, $\sqrt{n}[g(\hat{b}_j) - g(\beta_0)]$ converges in distribution to a normal random variable with zero mean and variance $\partial g(\beta_0)/\partial\beta' V_j \partial g(\beta_0)/\partial\beta$.

(b) *Sometimes an estimator $\hat{b}_1$ is said to be more efficient than $\hat{b}_2$ if $V_2 - V_1$ is positive definite. Use the results in (a) to explain why this is a reasonable way to rank estimators.*

The result above indicates that when $V_2 - V_1$ is positive definite, any nonlinear transformation of β_0 by a continuously differentiable function g can be estimated by $g\left(\hat{b}_1\right)$ with a smaller asymptotic variance than $g\left(\hat{b}_2\right)$.

10. *Show that if X_n converges in distribution to a normal random variable with mean 0 and variance 1 and Z_n converges in probability to a constant number $A \neq 0$, then $X_n \cdot Z_n$ converges in distribution to a normal random variable with mean 0 and variance A^2.*

The result follows from the continuous mapping theorem.

OLS

1. *Suppose you consider a linear regression model of y_i on a constant term and a regressor x_i and z_i. Let d_i denote a dummy variable that takes value 1 if the ith person is a female and 0 if the person is a male.*

   (a) *Formulate a linear regression model that does not allow any difference between male and female observations.*

   (b) *Formulate a linear regression model that allows for a difference between male and female observations only in the constant term.*

   (c) *Formulate a linear regression model that allows for differences between male and female observations in the constant term and the coefficient on x_i. Coefficient on z_i is assumed to be the same between male and female.*

2. *Discuss how you will assess the omitted variable bias of the OLS estimator in the context of the linear regression model.*

   - *This problem expects an answer in general terms but you should be able to execute the omitted variable bias computation in a specific context.*

3. *Explain the difference in the interpretation of the β coefficient in the following linear regression model*

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

   where the conditional expectation of ε_i given x_i is zero.

4. *Discuss the difficulties we need to face when there is multicollinearity problem in the linear regression model.*

5. *Why is the OLS estimator inefficient when there is heteroskedasticity?*

6. *What is the consequence of omitting a variable not correlated with any of the included regressors?*

7. *What is the consequence of including a variable whose coefficient in a linear regression model is zero? Discuss this in two cases: including a variable that is not correlated with the error term and including a variable that is correlated with the error term.*

8. *Suppose there is only one regressor to examine the relationship to a dependent variable. Since linear regression analysis requires specifying a functional form, we might just wish to graphically analyze the data by plotting the dependent variable against the regressor. Critically evaluate this approach in the two different contexts:*

(a) *the residual given x_i in the linear regression model has zero conditional mean, and*

(b) *the residual and x_i are correlated.*

Applied

1. *Suppose we wish to estimate the wheat production function specified below using a random survey sample of wheat farmers in the US:*

$$\log y_i = \alpha + \beta \cdot \log fertilizer_i/acre + \gamma \cdot \log hours\_of\_labor_i/acre + \theta \cdot \log capital_i/acre + \varepsilon_i$$

where y_i is bushel per acre.

(a) *What does the error term represent? Give at least three different reasons why we may have an error term in this particular context.*

(b) *What assumptions do we need to justify using OLS for (1) consistent estimation of the coefficients, and (2) consistent estimation of OLS standard errors?*

(c) *In this particular context, discuss if the assumptions can be justified assuming that there is no misspecification in the systematic effects of inputs but that at least a part of the residual term represent the land quality and farmers choose input level knowing their land quality.*

(d) *How will you form the confidence interval for the OLS estimator of β?*

(e) *Suppose you learned that only about a half of the requested survey was returned. What is a crucial assumption that justifies the approach you discussed above?*

2. *Suppose we wish to estimate the earnings equation specified below using a random survey sample of workers in a firm:*

$$\log y_i = \alpha + \beta \cdot \log \text{age}_i + \gamma \cdot \log \text{experience}_i + \delta \cdot \log \text{education}_i + \theta \cdot \text{gender}_i + \varepsilon_i$$

where y_i is earnings per year. Assume that gender variable takes value 1 or 0 depending on whether the ith person is a male or female, respectively. Assume that the error term conditional on regressors have mean zero.

(a) *How will you test if there is no sex discrimination in this firm under homoskedasticity?*

(b) *How will you test if there is no sex discrimination in this firm under heteroskedasticity?*

(c) *If you suspect that the sex discrimination may take form related with return to age or experience or education. How will you test this? Specify a model you use and explain the test statistic with degrees of freedom.*

3. *Suppose currently a subsidy is given to the head of a household and we want to measure the changes in the household's consumption behavior for the households with husbands being designated as the head of household when the subsidy is given to the wife. We will measure the changes in their consumption behavior by the changes in the mean of the share of their food expenditure in their total expenditure.*

(a) *Discuss the difficulty of using observational data to measure the effect.*

(b) *Describe the randomized experiment needed to measure this effect.*

(c) Suppose the government does not want to randomly choose households whether to give the subsidy to wives or husbands but willing to randomly choose areas and give every wives or every husbands in the area the subsidy. Explain how this kind of sampling will allow us to measure the effect.

4. Suppose you wish to measure the price elasticity of gas consumption per year. You collect a random sample of households and collect gas price per kWh (kilo watt for an hour) paid (denoted by p) and amount consumed (denoted by y) along with household information and income (denoted by a vector x). Assume that a gas company is chosen by a household. In what follows, for simplicity, assume there is no nonlinear pricing. Assume that there is the following relationship between these variables

$$y = \alpha + \beta \cdot p + x'\theta + \varepsilon$$

where α, β, and θ are unknown constant parameters and ε is an unobserved variable.

(a) State conditions under which OLS estimator of this regression yields consistent estimator of α, β, and θ.

(b) Which of the assumptions you stated in (a) may be affected by the fact a gas company is chosen by a household? Explain why.

5. Suppose you wish to measure the effect of holding different health insurance on total health expenditure (denote it by y). The total health expenditure includes expenditure by a patient as well as that by the insurance company. Assume that the only difference among health insurances is the co-payment ratio (denote it by c); the percentage of expenditure you need to pay out of your own pocket. For example, if the person is not covered by a health insurance, then $c = 1$, and if the person is fully covered $c = 0$. Denote other variables by a vector x. Suppose there is the following relationship among the variables

$$\log y = \alpha + \beta \cdot c + x'\theta + \varepsilon,$$

where ε denotes an unobserved variable.

(a) Discuss how you will estimate parameters α, β, and θ in this model.

(b) Discuss how you will construct the 95% confidence interval of β.

(c) Discuss the difference of the model above and the model which includes cross terms of c and elements in vector x.

6. Two program evaluation problems are described below each with a proposed estimation method. For each evaluation problem, discuss briefly (in one or two lines for each question) what a potential problem with the proposed estimation is and why. For each issue you may find more than one problem with the described method. Choose the one you think is the more important problem. Assume that data are collected as described and that there is no problem regarding the sample size to apply large sample approximations.

(a) One wishes to measure the average effect of attending college on earnings. Using a random sample from US population one estimates the coefficients β_0,...,β_4 in the following regression using the OLS estimator:

$$\text{earnings} = \beta_0 + \beta_1 \cdot \text{college education} + \beta_2 \cdot \text{experience} + \beta_3 \cdot \text{age} + \beta_4 \cdot \text{gender} + \varepsilon,$$

where ε represents an unobserved variable.

(b) *Suppose there are only two types of coffee shops, one type of shops sells for \$3 a cup and the other for \$2 a cup. Assume the shops sell just coffee and there is only one size of a cup of coffee. Assume that each shop is a local monopoly. Let p denotes the price variable which takes value 1 if the shop sells \$3 a cup coffee and 0 if the shop sells a \$2 a cup coffee. The effect of the price change is measured as the OLS estimate of the coefficient on the p dummy variable on the following regression model:*

$$sales = \alpha + \beta \cdot p + \varepsilon$$

where ε represents an unobserved variable.

7. *Suppose you wish to measure the effect of a de-worming program on school attendance (number of days) in a developing country. The de-worming program randomly selects 100 schools and offer free de-worming medication to all students in those schools. As a comparison group comparable data are collected from randomly selected 100 schools and not given the treatment. To make the problem simpler, assume that whether a student takes the medication or not does not affect the effect of the medication on other students.*

   (a) *If every student in treated schools takes the medication and every student in untreated schools does not take the medication, how can we estimate the average impact of the medication on school attendance? (If you think this is too easy, you are correct so do not be concerned.)*

   (b) *If every student in treated schools does not necessarily take the medication and every student in untreated schools does not take the medication, how can we estimate the average impact of the medication on school attendance for those who take the medication?*

   (c) *Under the same condition as in b., how can we estimate the average impact of the medication on school attendance?*

8. *Suppose you wish to measure the effect of education on life expectancy. Suppose in 1910 in one area the compulsory schooling increased from 6 years to 7 years. You have census data on those who were raised in this area. You want to use the change in the compulsory schooling law to estimate the effect of education on life expectancy.*

   (a) *How will you estimate the average effect of a year increase in education from 6 years for those students who would have received just 6 years of education without the compulsory schooling law change?*

   (b) *What are the assumptions you need to make to justify the method you describe in a?*

   (c) *What kind of additional information do you need to examine if the assumptions you make in b. are valid?*