PRE-MIDTERM REVIEW QUESTIONS

(1) Suppose the parameter β can only take on the values 0 or 1. We have an unbiased estimator $\hat{\beta}$ available for β, and it has non-zero variance. Show that an estimator with strictly lower mean square error (MSE) can always be obtained by modifying $\hat{\beta}$ and that the modified estimator is no longer unbiased.

It has non-zero variance, so it is not just the trivial estimator that estimates β without error. Since it's unbiased and $P[\hat{\beta} = \beta] < 1$, it must have some non-zero probability of being less than one when $\beta = 0$. It also must have a non-zero probability of being greater than one when $\beta = 1$, by the same logic. But whether $\beta = 0$ or $\beta = 1$, we always bring our estimator closer to β if we set it to 1 whenever $\hat{\beta} > 1$, and similarly it always improves accuracy to set our estimate to 0 when $\hat{\beta} < 0$. But this truncated version of $\hat{\beta}$ is not unbiased. Since it always lies between 0 and 1, its expectation is above zero when $\beta = 0$ and below one when $\beta = 1$.

(2) You have found one of your favorite socks in your drawer, but realize the other half of the pair is in the drier. You go to the drier and pull out, one after another, m socks without finding one that matches. What is the likelihood function for the total number of socks in the drier? Flat-prior inference will not work here. Why? If you know that you have at most 50 socks, and thus that no more than 49 were in the drier to start with, and if you think any number between 1 and 49 could have been there to start with (i.e. a uniform prior over 1 to 49), what is your expected total original number of socks in the drier if you have pulled out 5 without finding the one you want. What is the expected number of additional socks you have to pull out before you find the one you want, given that so far you have pulled out $m = 5$ unsuccessfully?

If the number of socks in the drier is N, the probability of drawing m without matching the sock you are holding is

$$\prod_{j=1}^{m} \frac{N-j}{N-j+1} = \frac{N-m}{N} = 1 - \frac{m}{N}\,.$$

So the likelihood as a function of N is zero for $N \leq m$, then rises toward one as $N \to \infty$. So it does not have a finite sum as a function of N, and we cannot simply treat it as our posterior distribution (i.e. cannot use a flat prior). If $m = 5$ and we haev a flat prior on the integers 1 to 49, the posterior expected number of socks N is

$$\frac{\sum_{N=6}^{4} 9(N-5)}{\sum_{N=6}^{4} 9(1-5/N)} = 29.98$$

Conditional on the number of socks being N, the probability of finding the match (after the initial 5 failures) after j additional draws is 1/(N-5) for each of the numbers j from 1 to $N-5$. (This may be intuitively obvious to you. You can also see it by calculating the probability of a match as

$$\begin{aligned} & \frac{1}{N-5} && \text{on draw 1} \\ \frac{N-4}{N-5}\frac{1}{N-4} &= \frac{1}{N-5} && \text{on draw 2} \\ \frac{N-4}{N-5}\frac{N-3}{N-4}\frac{1}{N-3} &= \frac{1}{N-5} && \text{on draw 3} \\ & \dots && \text{etc.} \end{aligned}$$

That means the expected value of the number of additional draws, given N, is $(N-4)/2$, and since the expectation of N is 29.98, the expected number of additional draws is 12.99.

(3) x_j is either zero or one, with probability .5 on each value. $y_j = \beta x_j + z_j$, where z_j is 1 or -1 with equal probabilities and is independent of x_j. The data are independent across observations j.
   (a) Show that this setup makes the OLS estimator of β unbiased, consistent, and asymptotically normal.
   This is just a matter of noting that the standard assumptions $E[Y \mid X_j] = X_j\beta$, $\left|E[X_j'X_j]\right| \neq 0$ and i.i.d. data are met here. Here $E[X_j'X_j] = .5 \neq 0$, and since $E[z_j] = 0$ and z_j is independent of x_j, the $E[Y_j \mid X_j] = X_j\beta]$ assumption is also satisfied.
   (b) Show that there is a much better estimator than OLS for this model. Be sure to consider the case of $\beta = 0$ as well as other possible values of β, since the estimator behaves differently when $\beta = 0$.
   There are just four possible values for y_j: 1, -1, $\beta + 1$ and $\beta - 1$. Here's a good estimator: So long as the sample contains only $x_j = 0$ values, set $\hat{\beta} = 0$. As soon as there is a single $x_j = 1$ observation, set $\hat{\beta} = y_j$. As soon as there are two observations j, k with $x_j = x_k = 1$ and $y_j \neq y_k$, set $\hat{\beta} = (y_j + y_k)/2$, then leave the estimator unchanged as additional observations arrive. This estimator is exactly correct as soon as two observations with $x_j = 1$ and different values of y_j have occurred. Like OLS, it is unbiased conditional on the x's in any sample with at least one $x_j = 1$ observation. In samples with no $x_j = 1$ observations, OLS is undefined. This estimator is obviously biased in such samples, because it is 0 with probability one in such samples even if $\beta \neq 0$. When there is just one $x_j = 1$ observation, this estimator coincides with OLS. OLS is worse, though, because it takes the average of all the y_j's for $x_j = 1$ observations. In general, the number of $y_j = \beta + 1$ observations and the number of $y_j = \beta - 1$ observations will not be exactly equal, so OLS will not be perfectly accurate.

(4) I have a sample of 40 observations on quantity q_j and price p_j in a market, and I would like to estimate a demand curve $q_j = \beta_0 + \beta_1 p_j + \varepsilon_j$. Unfortunately, I have no data on a varible like weather to use as a supply-shifting instrumental variable in the demand equation. However, I have flipped a coin 40 times and recorded a variable z_j that I set to one whenever it came up heads and zero whenever it came up tails. Since it is random and has nothing to do with the y_j, x_j data, it surely is independent of ε_j. I realize that if the correlation of z_j with x_j is zero, this causes a "weak instruments" problem. But it seems I have been incredibly lucky — even though the true population correlation of z_j with x_j is surely zero, My particular sequence of coin flips has a sample correlation of .5 with x_j! So when I form $\hat{\beta} = (Z'X)^{-1}Z'Y$, there is no problem with near-singular $Z'X$. (Here Z and X are 40×2 matrices containing the constant vector (of ones) and a column of z values or 4x4 values, respectively.) Does this mean I can get reliable estimates using the coin flips as an instrument in this sample? Why or why not?

It doesn't. We need the instrument to "purge" the price variable of the component of it that is correlated with the demand shock ε_j. The coin flips will certainly have a true correlation with the demand shock of zero, but also a true correlation with the other component of price of zero. If the sample correlation of the coin flips with p is high, this could be due to a high sample correlation with either component of p. The instrument might look as if it is giving good results, therefore, but we know that in fact, because its true correlation with p is zero, this IV estimator is uninformative about the true demand curve.

(5) I want to estimate expenditure on food by state. I have a sample of individuals containing food expenditure for each individual and a set of dummy variables. Each dummy variable is 1 for individuals living in one state and zero for all other individuals, and there is one such variable for each state. I propose to estimate a regression on food expenditures on the full set of dummy variables (of course omitting the constant term, since it would be collinear with all the dummies). Explain why in this case multiple regression is pointless, and I might as well just estimate mean food expenditure in each state separately.

Consider the $X'X$ matrix. Each state dummy is one for observations in its state and zero elsewhere, so crossproducts of different state dummies are zero. The sum of squares of state dummy j is N_j, the number of observations in that state. So the $X'X$ matrix is diagonal, with $1/N_j$ down the diagonal. $X'y$ has as j'th element the sum of y_{ij} values for that state, where y_{ij} is food expenditure by individual i in state j. So $(X'X)^{-1}X'y$ is just the vector of state sample means. Setting up X as a matrix of dummy variables and using matrix algebra is unnecessary.

(6) Here's a really trivial limited dependent variable model: $P[y_j = 1] = p(x_j\beta)$, where x_j is a scalar variable (so there is no constant term). Suppose we only have two observations: $y_1 = 1$, $y_2 = 0$, $x_1 = 2$, $x_2 = 1$. The specifications we are considering are

$$\begin{aligned} \text{LPM}: &\qquad P[y_j = 1] = .5 + x_j\beta \\ \text{logit}: &\qquad P[y_j = 1] = \frac{e^{x_j\beta}}{1+e^{x_j\beta}} \\ \text{probit}: &\qquad P[y_j = 1] = \Phi(x_j\beta)\,, \end{aligned}$$

where as usual Φ is the normal cdf.

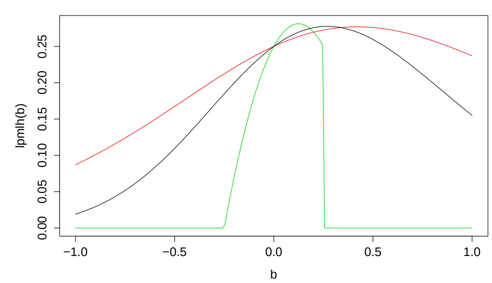
(a) Write down the three likelihood functions.

$$\begin{aligned} \text{LPM}: &\qquad (.5+2\beta)(.5-\beta) \quad \text{for } \beta \in (-.25, ..25)\,, \quad 0 \text{ otherwise.} \\ \text{logit}: &\qquad \frac{e^{2\beta}}{(1+e^{2\beta})(1+e^{\beta})} \\ \text{probit}: &\qquad \Phi(2\beta) * (1-\Phi(\beta)) \end{aligned}$$

(b) Why in this case is it natural to put the ".5" as the constant in the LPM?

The logit and probit models imply that $\beta = 0$ corresponds to a model that predicts $P[y_j = 1] = .5$ for all j. To make the LPM comparable, we want it to have the same property.

(c) Use the computer to graph the three likelihood functions. Could you tell which was logit and which probit just from the shapes of the curves? (On an exam, you woudn't be asked to do the computation, but you might be shown the curves and asked to tell which was which.)



The linear probability model is easy to distinguish from the others, because it has bounded support. (It's the green one.) The other two both are rather flat, with the probit (the black line) peaking at .27 and the logit peaking at .42. Though the peaks differ substantially, the likelihoods at the peaks do not differ much, so the models fit similarly well. Note that, as discussed in class, the derivative of $P[y_j = 1]$ with respect to β is different for the three models, so the values of β are not directly comparable across models. For logit and probit, the implied probabilities of the sample values $y_1 = 1]]$ and $y_2 = 0$ when β is at the MLE differ only in the third significant digit.

(d) What would happen to the likelihood functions if instead of $x_2 = 1$ we had $x_2 = -1$?
With $x_2 = 1$, $P[y_1 = 1 \mid x_1 = 2]$ is increasing in β and $P[y_2 = 0 \mid x_2 = 1]$ is decreasing in β for all three models, so the likelihoods have well-defined peaks. But if $x_2 = -1$, both probabilities increase in β, so the likelihoods for logit and probit simply increase toward one as β increases. The LPM's likelihood also is monotone increasing over the (-.25, .25) interval (which still is the range over which it does not imply impossible probabilities), but has a maximum at $\beta = .25$

(7) An exam would be likely to also have a question or two asking you to interpret or criticize estimation results, as the textbooks do when they discuss examples. Looking at a few of those examples should be part of your preparation.