371H Exam Grading Key

April 28, 2016

Question 1: Short answer

(A) Logistic regression is used when the outcome y in our regression model is binary (yes/no). It is generally superior to the linear probability model because it ensures that the predicted probability that $y = 1$ will always fall between 0 and 1.

(B) False, because $E(u(X)) = 0.5 \cdot \log(900) + 0.5 \cdot \log(1600)$, which is not equal to the value given in the statement.

(C) False: underestimating the covariance will mean underestimating the variance of the linear combination. Because we think the portfolio is less variable than it really is, we will underestimate the probability of a large loss.

(D) Bayes' rule holds that

$$P(A \mid B) = P(A) \cdot \frac{P(B \mid A)}{P(B)} .$$

We are given $P(A) = 0.15$, $P(B \mid A) = 0.9$, and $P(B \mid \text{ not } A) = 0.05$. From these, we can use the addition and multiplication rules to calculate

$$P(B) = P(B, A) + P(B, \text{ not } A) = 0.15 \cdot 0.9 + 0.85 \cdot 0.05 .$$

Therefore we know that

$$P(A \mid B) = 0.15 \cdot \frac{0.9}{0.15 \cdot 0.9 + 0.85 \cdot 0.05}$$

(which you don't need to actually calculate).

Question 2: Essay

There are five major concepts here. You had to describe a clear simulation procedure that correctly incorporated and explained all five of these elements. Rather than write out an essay, let me highlight these main elements:

Value at risk: for the 5% level, VaR is the 5% quantile of the profit/loss distribution after two weeks or 10 trading days.

Monte Carlo simulation: to approximate the probability distribution of profit/loss after ten days, we will repeatedly simulate the trajectory of a portfolio over that horizon.

Compounding daily returns. To get one trajectory (i.e. draw from the probability distribution of profit/loss after 10 days), we must compound (i.e. chain together in a for loop) 10 individual sets of daily returns for the three assets. Each individual day's set of returns is a draw from the joint distribution.

Daily returns for our three assets are probably correlated. We must model them with a joint distribution, rather than three independent distributions.

The joint distribution of daily returns is complicated. We often must use a technique like bootstrap resampling to approximate this joint distribution.

Question 3

The general strategy for this question was to use the estimated coefficients/standard errors in the tables, together with the results of the permutation tests for various pairs of models, to choose a good model. Once a model is chosen, then it becomes straightforward to reason about the statements.

Statements 1-3. These statements are all wrapped up together. In light of the evidence, I end up believing that 1-2 are true and 3 is false. But the most important thing here is that you look to the right models and go through the right process to address the question, not that you agree with me about the answer. There are many approaches you could have taken here.

Let's take statement 1: *the JEC cartel, when it was operational, fundamentally shifted the demand curve for rail shipping up or down, by changing the constant A (holding other relevant factors constant).*

On balance this seems true. This statement is asking about the cartel dummy variable in the regression with log quantity as a response. There are several varieties of correct answer here. For example, you could compare models with cartel to those without. In both of the direct comparisons involving this variable, it looks like cartel shifts the curve:

i. Model 3 beats Model 2 ($R^2 = 0.254$ for Model 3 in tails of permutation-test null distribution in Figure 5, cartel coefficient in model 3 more than twice its standard error.)

ii. Model 7 beats Model 6 ($R^2 = 0.324$ for Model 7 in tails of null distribution in Figure 8, cartel coefficient in model 7 more than twice its standard error.)

The story is similar for statement 2: *the cartel, when it was operational, fundamentally altered the demand curve for rail shipping by changing the price elasticity of demand β (holding other relevant factors constant).* This statement is asking about the interaction between log price and cartel (i.e. did the activation of the cartel dummy variable change the slope of the line). In both of the direct comparisons involving this interaction variable, it looks like cartel changes the slope:

i. Model 4 beats Model 3 ($R^2 = 0.264$ for Model 4 in tails of permutation-test null distribution in Figure 6, cartel:price interaction in model 4 roughly twice its standard error.)
ii. Model 8 beats Model 7 ($R^2 = 0.341$ for Model 8 in tails of null distribution in Figure 9, both interaction and cartel coefficient in model 8 more than twice their standard errors.)

You could also just select what you perceive to be the best model and draw a conclusion on the basis of that model. To me Model 8 looks like the best on the basis of the pairwise tests. This model includes both cartel and an an interaction between cartel and log price, implying that statements 1-2 are both true and statement 3 is false.

The one model where you can make a good case against statements 1-2 is in Model 4. Here, the cartel–price interaction is about twice its standard error, and the cartel dummy variable is about 1.5 times its standard error. If you had a conservative alpha level, you would fail to reject the null hypothesis that both coefficients on these variables are zero. (Note: one thing that doesn't make much sense is to use Model 4 to say that statement 2 is true but that statement 1 is false. Remember, we can't have the interaction term in Model 4 without also having the corresponding main effects.)

The problem with this argument is that Model 8 really seems better than Model 4. For example, comparing the $R^2 = 0.341$ for Model 8 to the null distribution in the right panel of Figure 10 makes it hard to leave out the seasonal effects. And in Model 8, once we adjust for confounding due to these seasonal effects, both the cartel dummy variable and interaction look significant (although not necessarily with a very conservative alpha level—though you would need to argue this explicitly).

Statement 3 says the following. *The cartel, when it was operational, did make railroad shipping more expensive for consumers, and therefore affected the quantity demanded. But, holding other relevant factors constant, the cartel's operation did not fundamentally alter the demand curve in either of the ways described in statements 1 or 2.* But if statements 1 and 2 are true, then 3 is false.

A side note to be aware of in this problem: there are a few irrelevant distractions in the figures. For example, in the left panel of Figure 10, we see the sampling distribution of the log(price):cartel interaction term. This is not relevant for deciding the result of the permutation test.

Statement 4: *There are significant seasonal effects in the time series of price (P), adjusting for other* *relevant factors.* There are clear seasonal variations in price, but the statement is false: once we adjust for covariates, the remaining seasonal effects appear insignificant. In other words: ice and cartel are confounded with season, and once we adjust for these confounders, the seasonal effects vanish. The relevant test is Model 5 versus Model 4. We start with Model 4 as the "null hypothesis" here because it provides a baseline set of factors but doesn't include season. Moreover, we clearly need all the factors included in Model 4. For example, model 4 clearly beats Model 2: see Figure 2, where the R^2 for Model 4 is not even close to the null distribution. Moreover, in Model 4, all the coefficients are at least three times their standard errors.

In comparing Model 5 to Model 4, we see that R^2 for Model 5 is 0.579, versus 0.559 for Model 4. But $R^2 = 0.579$ is completely consistent with the null hypothesis of no seasonal effects, as can be seen in permutation test for season shown in the right panel of Figure 4. Therefore under any reasonable α level we would fail to reject the hypothesis that there are no remaining seasonal effects once ice, time, and the cartel variables are included.

Statement 5: *There are significant seasonal effects in the time series of quantity (Q), adjusting for other relevant factors.* This looks true. The relevant comparison is Model 8 versus Model 4, in which season looks significant. (See Figure 10, in which comparing $R^2 = 0.341$ for Model 8 to the null distribution in the right panel of Figure 10 makes it hard to leave out the seasonal effects.)

You can also get almost full credit by looking at the coefficients in Table 3 and observing that many of the monthly dummies are individually significant (i.e. different from the baseline month of January) and that there seems to be some variability in these coefficients across the 4-week periods. However, the permutation test is really a cleaner way of addressing this question.

Statement 6: *The presence of ice on the Great Lakes fundamentally shifted the demand curve for rail shipping up or down, by changing the constant A (holding other relevant factors constant).* This statement is clearly true: in every single model with log quantity as a response, the coefficient on the ice dummy variable is very large compared to its standard error. When there was ice on the Great Lakes, the demand curve for rail shipping shifted up at all price levels.

Statement 7: *The presence of ice on the Great Lakes fundamentally altered the demand curve for rail shipping by changing the price elasticity of demand β (holding other relevant factors constant).* This statement is undecidable. We'd need to look at a model involving an interaction between ice and log(price) to decide this question.