Preston Mui — mui@berkeley.edu

Econ-240a Part 2 Review Sheet (partial) Answers

3

(a) Show that $b_0 = \beta_0 + \rho \frac{\sigma_U}{\sigma_V} \frac{1}{\mu^2+1}$: The MSE minimizing predictor of Y given X is

$$\begin{aligned} E^*[Y|X] &= E^*[\alpha_0|X] + E^*[\beta_0 X|X] + E^*[U|X] \\ &= \alpha_0 + \beta_0 X + E^*[U|X] \end{aligned} \tag{1}$$

Trick is to get $E^*[U|X] = f + gX$. By the best linear predictor, $g = \mathbb{C}(X,U)\mathbb{V}(X)^{-1}$ and $f = \mu_U - g\mu_X$. Solving first for g:

$$\begin{aligned} g = \mathbb{C}(X,U)\mathbb{V}(X)^{-1} &= \mathbb{C}(\eta_0 + Z'\pi_0 + V, U)\mathbb{V}(\eta_0 + Z'\pi_0 + V)^{-1} \\ &= (\rho\sigma_V\sigma_U)\Big(\mathbb{V}(Z'\pi_0) + \mathbb{V}(V)\Big)^{-1} \\ &= \rho\frac{\sigma_U}{\sigma_V}\Big(\pi_0'\mathbb{V}(Z)\pi_0/\sigma_V^2 + 1\Big)^{-1} = \rho\frac{\sigma_U}{\sigma_V}\Big(\mu^2+1\Big)^{-1} \end{aligned}$$

Ignoring f because it will not be a function of X, plugging back into (1) gives

$$E^*[Y|X] = \alpha_0 + \beta_0 X + f + gX = (\alpha_0 + f) + \Big(\beta_0 + \rho\frac{\sigma_U}{\sigma_V}(\mu^2+1)^{-1}\Big)X$$

$b_0 \neq \beta_0$ when $\rho \neq 0$ because when $\rho \neq 0$, the error term U is correlated with V, which feeds into the variable X, so the error is correlated with the independent variable. In regards to Card and Krueger, if schooling resources (Z) affect educational attainment (X), then U may be correlated with X.

(b) From the properties of multivariate normal distributions, $E[U|V=v] = \mu_U + \rho\sigma_V\sigma_U(\sigma_V^2)^{-1}(v-\mu_V) = \rho\frac{\sigma_U}{\sigma_V}v$. So,

$$\begin{aligned} E^*[Y|X,V] &= \alpha_0 + \beta_0 X + E^*[U|X,V] \\ &= \alpha_0 + \beta_0 X + E[E^*[U|X,V]|Z] \\ &= \alpha_0 + \beta_0 X + E[E^*[U|V]|Z] \\ &= \alpha_0 + \beta_0 X + \rho\frac{\sigma_U}{\sigma_V}V \end{aligned}$$

We need to include Z because the expectation of U given X, V could in theory depend on X. But, applying the law of iterated expectations over Z, we can ignore the X because X is a function of Z and V. Then, note that the joint distribution of U and V is the same regardless of the value of Z. Intuitively, the coefficient on V captures the part of X that affects that portion of U. This linear predictor is not well-defined if π_0 because if $\pi_0 = 0$, then X and V are perfectly collinear.

(i) The H matrix for this test is given by $H = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$; the θ_0 matrix is given by $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. That is, testing that the coefficients on D_{FMA}, D_{MJJ}, and D_{ASO} are all equal to 0. The Wald statistic is given by

$$\begin{aligned} W_0 &= N(\hat\theta - \theta_0)'(H\Lambda H')^{-1}(\hat\theta - \theta_0) \\ &= 32587\begin{pmatrix} -0.3372 & -0.2711 & 0.0406 \end{pmatrix}(H\Lambda H')^{-1}(\hat\theta - \theta_0) \end{aligned}$$

First, examine $(H\Lambda H')^{-1}$ which is simply Λ after taking off the first row and column. There is kind of a problem here in that inverting a 3 x 3 matrix by hand is annoying, so I will skip that and say just compare W_0 to the critical values of the χ^2_3 distribution.

(ii) This is simply testing whether or not the Coefficient on $\hat V$ is equal to 0. If it is, then both equations (1) and (3) provide the same functional form and coefficients. So, this is a simple t-test on the coefficient of $\hat V$. The t-stat is $\frac{0.0076}{0.0302} > 3$ so we can reject the null that the coefficients are the same.

10

(a) $E^*[Y|G] = \alpha + \beta G + 2E^*[A|G] + E^*[U|G] = \alpha + \beta G + 2(\delta_0 + \delta_1 G) = \alpha + 2\delta_0 + (\beta + 2\delta_1)G$

(b) $\alpha + 2\delta_0 = 35000$ and $\beta + 2\delta_1 = 30000$.

(c) $\delta_1 = 1300 - 1100 = 200$; $\delta_0 = 1100$.

(d) According to the model, the expected earnings cost of dropping out is β. Using the calculations from above, $\beta = 30000 - 400 = 29600$. However, if I am dropping out specifically to spend more time on Telegraph Avenue, presumably the cost is higher because I will be jobless.

(e) Using $E^*[Y|G]$, the expected earnings of the Free Speech Movement neighbor are $\alpha + 2\delta_0 = 35000$ and the expected earnings of the other neighbor are $\alpha + 2\delta_0 + \beta + 2\delta_1 = 65000$.

15

(a) β_l is the simple average of $m(X_i)$ for all $X_i \in B_l$.

(b) The LHS is Mean-Squared-Error simply because MSE = Squared Bias + Variance. So, the problem resolves to showing that the LHS and RHS are equal. First off, I believe there is an error in the statement, and that the RHS should divide the first term by 4. In any case, here is how I showed it.

Without loss of generality, order the observations such that $X_1 < X_2 < X_3 < \cdots < X_N$. Further, denote N_l as the number of observations within a given strata l.

First, consider the bias of $\hat{m}(\bar{x}) = \hat{\beta}_L$. since $\hat{\beta}_L$ will be the observed mean of Y_i for all i such that $X_i \in B_L$, we can write

$$\begin{aligned}
E[\hat{m}(\bar{x})] &= E\Big[\frac{1}{N_L}\sum_{j=N-N_L+1}^{N} Y_j\Big] \text{ (Don't freak out this is just the average } Y \text{ for the observations in bin } L) \\
&= \frac{1}{N_L}\sum_{j=N-N_L+1}^{N} E[Y_j] \\
&= \frac{1}{N_L}\sum_{j=N-N_L+1}^{N} m(X_j)
\end{aligned}$$

For every $m(X_j)$ we can form an approximation to $m(X_j)$ using $m(\bar{x})$ and $m'(\bar{x})$. Note that the distance (in X units) between X_N and X_j is $\frac{N-j}{N_L-1}h$. So, the approximation is $m(X_j) \approx m(\bar{x}) - m'(\bar{X})\frac{N-j}{N_L-1}h$. Pluggin this approximation in,

$$\begin{aligned}
E[\hat{m}(\bar{x})] &= \frac{1}{N_L}\sum_{j=N-N_L+1}^{N}\left[m(\bar{x}) - m'(\bar{X})\frac{N-j}{N_L-1}h\right] \\
&= m(\bar{x}) - \frac{1}{N_L}\sum_{j=N-N_L+1}^{N}\left[m'(\bar{X})\frac{N-j}{N_L-1}h\right] \\
&= m(\bar{x}) - m'(\bar{x})h\frac{1}{N_L}\sum_{j=N-N_L+1}^{N}\left[\frac{N-j}{N_L-1}\right] \\
&= m(\bar{x}) - m'(\bar{x})h\frac{1}{N_L}\frac{0+1+\cdots+N_L-1}{N_L-1} \\
&= m(\bar{x}) - m'(\bar{x})h\frac{1}{N_L}\frac{(N_L-1)N_L/2}{N_L-1} \\
&= m(\bar{x}) - m'(\bar{x})\frac{h}{2}
\end{aligned}$$

So, bias squared is

$$\Big(E[\hat{m}(\bar{x})] - m(\bar{x})\Big)^2 = \Big(m(\bar{x}) - m'(\bar{x})\frac{h}{2} - m(\bar{x})\Big)^2 = m'(x)^2\frac{h^2}{4}$$

The variance of $\hat{m}(\bar{x})$ will be

$$\begin{aligned}
\mathbb{V}(\hat{m}(\bar{x})) &= \mathbb{V}(\frac{1}{N_L}\sum_{j=N-N_L+1}^{N} Y_j) \\
&= \frac{1}{N_L^2}\sum_{j=N-N_L+1}^{N} \mathbb{V}(Y_j) \\
&= \frac{\sigma^2}{N_L}
\end{aligned}$$

So the question is what is N_L? Well this is N divided by L. But $L = (\bar{x} - \underline{x})/h$. So, $\mathbb{V}(\hat{m}(\bar{x})) = \frac{\sigma^2(\bar{x}-\underline{x})}{Nh}$.
The intuition of this problem is that if you have a really small bin then your bias is very very small because you are only taking observations close to $\bar{x}$. But, if you have a small bin then you have fewer observations so your variance is much higher. So, that is the tradeoff. Taking the first-order conditions of the RHS (assuming the statement as written in the review sheet is CORRECT) and rearranging yields

$$h^* = \Big(\frac{(\bar{x}-\underline{x})\sigma^2}{2Nm'(\bar{x})^2}\Big)^{1/3}$$

(c) From the functional form of Y given X derive $\hat{m}'(x) = \hat{\pi}_1 + 2\hat{\pi}_2 x$. Then, plugging in for the optimal bin length from above gives

$$\hat{h^*} = \Big(\frac{(\bar{x}-\underline{x})\hat{\sigma}^2}{2N(\hat{\pi}_1 + 2\hat{\pi}_2 x)^2}\Big)^{1/3}$$

Since $L = (\bar{x} - \underline{x})h^{-1}$,

$$\begin{aligned}
L_N &= (\bar{x}-\underline{x})\Big(\frac{(\bar{x}-\underline{x})\hat{\sigma}^2}{2N(\hat{\pi}_1 + 2\hat{\pi}_2 x)^2}\Big)^{-1/3} \\
&= 8(\bar{x}-\underline{x})^{2/3}\Big(\frac{\hat{\sigma}^2}{(\hat{\pi}_1 + 2\hat{\pi}_2 x)^2}\Big)^{-1/3} N^{1/3}
\end{aligned}$$

The floor function applies to give us an integer number of bins.

(d) Here is a simple but not-all-that-formal argument: The function in this section is an estimate of $m(X)$. For any two X_i, X_j such that $i, j \in B_l$, the implied $m'(X_i)$ is $\frac{m(X_j)-m(X_i)}{X_j - X_i} = \gamma_l$. A little more formally, say i and $i+1$ are both in B_i, then

$$\begin{aligned}
E[m'(x_i)] &\approx E\Big[\frac{m(X_{i+1}) - m(X_i)}{X_{i+1} - X_i}\Big] \\
&= \frac{E[m(X_{i+1})] - E[m(X_i)]}{X_{i+1} - X_i} \\
&= \frac{\gamma_l X_{i+1} - \gamma_l X_i}{X_{i+1} - X_i} = \gamma_l
\end{aligned}$$

This relies on the linear predictor being close to the conditional mean of Y on X. Also the distance between X_{i+1} and X_i gets smaller as N increases, so it gets closer to the definition of the derivative.