Preston Mui — mui@berkeley.edu

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

3 Binomial-Beta Learning

(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:

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

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_3^2 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.