Econ 2120: 2020 final partial solutions

December 2020

Question 1

With Y_i and X_i binary, we have the following model:

$$P(Y=1 \mid X) = \frac{\exp\left(\beta_0 + \beta_1 X\right)}{1 + \exp\left(\beta_0 + \beta_1 X\right)}$$

(a) – MLE

As Y_i is binary, the conditional density has the following form

$$f(Y_i|X_i) = \left(\frac{\exp\left(\beta_0 + \beta_1 X_i\right)}{1 + \exp\left(\beta_0 + \beta_1 X_i\right)}\right)^{Y_i} \left(1 - \frac{\exp\left(\beta_0 + \beta_1 X_i\right)}{1 + \exp\left(\beta_0 + \beta_1 X_i\right)}\right)^{1-Y_i}$$

Now, let's make a substitution: $\frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)} \equiv p_1$ and $\frac{\exp(\beta_0)}{1+\exp(\beta_0)} \equiv p_0$. This permits us to express the likelihood in an easier way

$$L(Y|X) = \prod_i^n f(Y_i|X_i) = \prod_i^n \left(p_1 X_i + p_0(1-X_i)\right)^{Y_i} \left(1 - p_1 X_i - p_0(1-X_i)\right)^{1-Y_i}$$

$$l(Y|X) = \ln L(Y|X) = \sum_{i=1}^n \Big(Y_i \ln\left(p_1 X_i + p_0(1-X_i)\right) + (1-Y_i)\left(1 - p_1 X_i - p_0(1-X_i)\right)\Big)$$

FOC's yield

$$\sum_{i=1}^n \left(Y_i \frac{(1-X_i)}{p_1 X_i + p_0(1-X_i)} + (1-Y_i)(-1)\frac{(1-X_i)}{1 - p_1 X_i - p_0(1-X_i)}\right) = 0$$

$$\sum_{i=1}^n \left(Y_i \frac{X_i}{p_1 X_i + p_0(1-X_i)} + (1-Y_i)(-1)\frac{X_i}{1 - p_1 X_i - p_0(1-X_i)}\right) = 0$$

Following the first equation and using the fact that $\frac{Y_i X_i}{p_1 X_1 + p_0(1-X_i)} = \frac{Y_i X_i}{p_1}$,

$$\sum_{i=1}^{n}\left(Y_i\frac{(1-X_i)}{p_1X_i+p_0(1-X_i)}+(1-Y_i)(-1)\frac{(1-X_i)}{1-p_1X_i-p_0(1-X_i)}\right)=$$
$$\sum_{i=1}^{n}\left(\frac{Y_i(1-X_i)}{p_0}-\frac{(1-Y_i)(1-X_i)}{1-p_0}\right)=0$$
$$\frac{1-p_0}{p_0}=\frac{\sum_{i=1}^{n}(1-Y_i)(1-X_i)}{\sum_{i=1}^{n}Y_i(1-X_i)}$$
$$\hat{p}_0=\frac{\sum_{i=1}^{n}Y_i(1-X_i)}{\sum_{i=1}^{n}(1-X_i)}$$
$$\hat{p}_1=\frac{\sum_{i=1}^{n}Y_iX_i}{\sum_{i=1}^{n}X_i}$$

Under regularity conditions MLE estimates are consistent, $\hat{p}_i \to p_i$. Also $\forall g(\cdot)$ continuous (a.e.) $g(\hat{p}_i) \to g(p_i)$. Using this fact and assuming $p_i \in (0,1)$,

$$\hat{\beta}_0=\ln\left(\frac{p_0}{1-p_0}\right)\to\beta_0$$
$$\hat{\beta}_1+\hat{\beta}_1=\ln\left(\frac{p_1}{1-p_1}\right)\to\beta_0+\beta_1$$

Another possible way to get to the same estimates would be just to consider the separate MLE estimates for the conditional means of Y evaluated at particular values of X. The estimates for the conditional probabilities would then simply be the conditional means. Namely,

$$f(Y|X=1)=p_1^Y(1-p_1)^{1-Y}$$
$$\hat{p}_{1,MLE}=\frac{1}{\#X_i=1}\sum_{i:X_i=1}Y_i=\frac{\sum_{i=1}^{n}Y_iX_i}{\sum_{i=1}^{n}X_i}$$

To get the asymptotics for MLE estimates, we use GMM, as shown in the next point.

General remark: *For compactness, the problem here considered variable substitution with an idea that we later can use the continuous mapping theorem to have estimates of the original βs from the ps, but it is absolutely correct just to do calculations in terms of βs directly as many of you did. However, this substitution idea is useful to keep in mind, as it simplifies the calculations, splitting them into better manageable steps.*

b – GMM

MLE FOCs provide GMM orthogonality conditions (just-identified case, so no need for the weighting matrix). The estimates will correspond to those obtained in the previous point, but the usage of GMM framework permits us to obtain the asymptotic variance and distribution of the estimates.

$$\psi(W_i,p)=\begin{bmatrix}Y_i\frac{(1-X_i)}{p_0}+(1-Y_i)(-1)\frac{(1-X_i)}{1-p_0}\\ Y_i\frac{X_i}{p_1}+(1-Y_i)(-1)\frac{X_i}{1-p_1}\end{bmatrix}$$

Could easily check that $\mathbb{E}[\psi(W_i, \beta)] = 0$.

Then, standard asymptotics follow

$$\sqrt{n}(\hat{p} - p) \to \mathcal{N}(0, \Lambda)$$
$$\Lambda = \alpha' \Sigma \alpha$$
$$\alpha = \left[E\left[\frac{\partial \psi(W_i, \gamma)}{\partial a'} \right] \right]^{-1}$$
$$\Sigma = \mathrm{Cov}\left(\psi\left(W_i, \gamma\right)\right) = E\left[\psi\left(W_i, \gamma\right)\psi\left(W_i, \gamma\right)'\right]$$

Then, using the Delta method, we obtain asymptotics for β.

Other possible setup is to consider the fact that we have the specification for the conditional mean of Y_i: $P(Y_i = 1|X_i) = \mathbb{E}[Y_i|X_i] = \frac{\exp(\beta_0+\beta_1 X)}{1+\exp(\beta_0+\beta_1 X)}$.

$$V_i \equiv Y_i - \mathbb{E}[Y_i|X_i]$$
$$E[B_i V_i] = 0$$

where $B_i = g(X_i)$ is a vector with components that are functions of X_i (any one would satisfy orthogonality, as the model specifies the regression function – the projection error is orthogonal to any function of X_i). E.g. $B_i = [1, X_i]'$ for a just identified case. This yields a moment function

$$\psi(W_i, \beta) = B_i \left(Y_i - \frac{\exp\left(\beta_0 + \beta_1 X\right)}{1 + \exp\left(\beta_0 + \beta_1 X\right)} \right)$$

and all the asymptotic results follow.

c – Minimum Distance

One possible way to setup a MD problem is to consider just the conditional means, using the fact that $\mathbb{E}[Y_i|X_i] = P(Y_i = 1|X_i)$

$$h(\pi, \gamma) = \begin{bmatrix} \mathbb{E}[Y_i|X_i = 1] \\ \mathbb{E}[Y_i|X_i = 0] \end{bmatrix} - \begin{bmatrix} \frac{\exp(\beta_0+\beta_1)}{1+\exp(\beta_0+\beta_1)} \\ \frac{\exp(\beta_0)}{1+\exp(\beta_0)} \end{bmatrix}$$

which yields

$$\sqrt{n}(\hat{\gamma} - \gamma) \xrightarrow{d} \mathcal{N}(0, \Lambda)$$
$$\alpha = \left[\frac{\partial h(\pi, \gamma)}{\partial a'} \right]^{-1}$$
$$\Sigma = \frac{\partial h(\pi, \gamma)}{\partial \pi'} \Omega \frac{\partial h(\pi, \gamma)'}{\partial \pi}$$

where

$$\frac{\partial h(\pi,\gamma)}{\partial \pi'} = I$$

$$\frac{\partial h(\pi,\gamma)}{\partial a'} = \begin{pmatrix} \frac{\exp(\beta_0+\beta_1)}{(1+\exp(\beta_0+\beta_1))^2} & \frac{\exp(\beta_0+\beta_1)}{(1+\exp(\beta_0+\beta_1))^2} \\ \frac{\exp(\beta_0)}{(1+\exp(\beta_0))^2} & 0 \end{pmatrix}$$

$$\Omega = \begin{pmatrix} p_1(1-p_1) & 0 \\ 0 & p_0(1-p_0) \end{pmatrix}$$

General remark: *this is by far not the only way to setup the problem. In general, a conceptually easy setup for an MD procedure would be to consider some simple statistics, like means and variances (or the conditional versions thereof), and then relate them to the underlying parameters. However, as some of you did in the exam, it's also possible to consider more complicated statistics, e.g. OLS estimates, as the "unrestricted" targets and relate those to the restricted/modified model which uses the "structural" parameters.*

Question 2

Here we just remember the definition of the BLP and the conditions that are used to derive the coefficients. Coefficients β are found using the orthogonality conditions (projection error is orthogonal to the predictors – the property that was derived in solving the minimum norm problem/least squares problem)

$$\mathbb{E}[X_i(Y_i - X_i'\beta)] = 0$$

These orthogonality conditions provide a setup for a just-identified GMM.

$$\psi(W_i,\beta) = X_i(Y_i - X_i'\beta) = X_i U_i$$

this yields the estimate that is equal to the OLS estimate

$$\hat{\beta} = b_{ols} = \left(\sum X_i X_i'\right)^{-1} \sum X_i Y_i$$

while GMM asymptotic results provide the asymptotic variance for the estimates:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \alpha\Sigma\alpha')$$

$$\alpha = \left[E\left[\frac{\partial \psi(W_i,\gamma)}{\partial a'}\right]\right]^{-1}$$

$$\Sigma = \operatorname{Cov}(\psi(W_i,\gamma)) = E\left[\psi(W_i,\gamma)\,\psi(W_i,\gamma)'\right]$$

with

$$
\begin{aligned}
\alpha &= \mathbb{E}[X_i X_i']^{-1} \\
\Sigma &= \mathbb{E}[X_i U_i U_i' X_i'] = \mathbb{E}[U_i^2 X_i X_i']
\end{aligned}
$$

One remark is that in the just-identified case the weight matrix plays no role at all, so it does not appear in calculations (we don't need to minimize a weighted average, as we can set the sample moments exactly to zero – so no weighting matrix needed).