PhD Econometrics 1: Study Questions Class 6
Imperial College London
Hormoz Ramian

Solutions

Question 1

(1.1) There a are number of assumptions in place when working with different aspects of ML and its properties. The basic assumptions, however, are: The sample is drawn from a specific density with known functional form, which is continuous in $\boldsymbol{\beta} \in \boldsymbol{B} \subseteq \mathbb{R}$. The parameter space $\boldsymbol{B}$ is compact: without nonconvexities e.g. jumps, gaps, etc. and it contains its boundaries. When working with asymptotic properties, this can be replaced with assuming that the true $\boldsymbol{\beta}$ which generated the data belongs to the interior of the parameter space $\boldsymbol{B}$. The log-likelihood function $\ell(\boldsymbol{\beta}) := \log(L(\boldsymbol{\beta}))$ is continuously differentiable twice, its expectations and and first two derivative exits. The information matrix $\mathcal{I}$ is positive definite and is increasing in sample size n.

(1.2) The (average) score vector is $0 = \frac{\partial \bar{\ell}(\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = \frac{1}{n}\frac{\partial \ell(\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}}$ which can be re-expressed in term so of the following first order Taylor expansion:

$$0 \;=\; \frac{\partial \bar{\ell}(\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} \;=\; \frac{\partial \bar{\ell}(\boldsymbol{\beta})}{\partial^2 \boldsymbol{\beta}} + \frac{\partial \bar{\ell}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + \text{Remainder term } o_p(1) \tag{1}$$

Re-write s:

$$-\left\{\frac{1}{n}\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right\} \times \sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \;\approx\; \frac{1}{\sqrt{n}}\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \tag{2}$$

where we can apply the CLT to the random variable $\frac{1}{\sqrt{n}}\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ to establish the asymptotic normality of the score.

$$\frac{1}{\sqrt{n}}\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathcal{I}) \tag{3}$$

Assuming that the information matrix in invertible, use results in (2) and (3) to write:

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathcal{I}^{-1} \times \mathcal{I} \times \mathcal{I}^{-1}) \tag{4}$$

(1.3) Before applying the definition of the test statistics to this model, recall the following:

$$\begin{aligned}
\widehat{\beta} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \;\sim\; \mathcal{N}(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}) \\
\ell(\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2) &= -\frac{n}{2}[1 + \log(2\pi)] - \frac{n}{2}\log \widehat{\sigma}^2 \\
\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma^2}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\
\mathcal{I}^{-1} &= diag\left[\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}, \frac{2\sigma^4}{n}\right] \\
\widehat{\sigma}^2 &= \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\
\boldsymbol{\theta}' &:= (\boldsymbol{\beta}', \sigma^2)
\end{aligned}$$

Note that, first, $\mathcal{I}$ is a block diagonal matrix, and second, the score vector is asymptotically $\mathcal{N}(\mathbf{0}, \mathcal{I})$, and third that the two block elements inside $\mathcal{I}$ are independent. We can focus on the first block element when defining LM-statistic for $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ The large sample Wald statistic measures the distance between the hypothesized parameter vector $\boldsymbol{\beta}_0$ against

$\widehat{\boldsymbol{\beta}}$ (quadratically) scaled by inverse of the variance term in the following way:

$$\begin{aligned} W &= \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)' \left\{\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}\right\}^{-1} \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \\ &= \frac{1}{\widehat{\sigma}^2} \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)' (\boldsymbol{X}'\boldsymbol{X}) \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \end{aligned}$$

For the LM-statistic we need to construct H_0-restricted ML objective function to obtain $\widetilde{\sigma}^2 := \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)$ which is a random variable because:

$$\begin{aligned} LM &= \left\{ \frac{[\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})]'}{\sigma^2} \left[\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}\right] \frac{[\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})]}{\sigma^2} \right\} \Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\ &= \widetilde{\sigma}^2 [\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})]'(\boldsymbol{X}'\boldsymbol{X})^{-1}[\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})] \\ &= \widetilde{\sigma}^2 \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)' (\boldsymbol{X}'\boldsymbol{X}) \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \end{aligned}$$

Lastly, denote $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$:

$$LR = 2\left(\frac{n}{2}\log \widetilde{\sigma}^2 - \frac{n}{2}\log \widehat{\sigma}^2\right) = n \log\left(\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2}\right)$$

(1.4)

$$W = \frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} LM$$

noting that

$$\begin{aligned} \widetilde{\sigma}^2 &= \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0)'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}_0) \\ &= \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0))'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)) \\ &= \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) + \frac{1}{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{2}{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) \\ &= \widehat{\sigma}^2 + \frac{\widehat{\sigma}^2}{n} W + 0 \end{aligned}$$

since $\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \mathbf{0}_k$. Then,

$$\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} = 1 + \frac{1}{n} W$$

but comparing with LM-statistic give:

$$LM = \frac{W}{1 + W/n} \leq W$$

but comparing with LM-statistic give:

$$\frac{1}{n} LR = \log\left(1 + \frac{1}{n} W\right) \leq \frac{1}{n} W$$

Question 2

(2.1) Multiplying out the objective function $\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} Q_n(\boldsymbol{\theta})$ noting that

$$\begin{aligned} Q_n(\boldsymbol{\theta}) &= \left\{\frac{1}{n}\boldsymbol{u}(\boldsymbol{\theta})'\boldsymbol{Z}\right\}' \boldsymbol{W}_n \left\{\frac{1}{n}\boldsymbol{u}(\boldsymbol{\theta})'\boldsymbol{Z}\right\} \\ &= \frac{1}{n^2}\left[\boldsymbol{y}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y} + \boldsymbol{\theta}'\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\theta} - 2\boldsymbol{y}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\theta}\right] \end{aligned}$$

Differentiating w.r.t. $\boldsymbol{\theta}$ and equating to zero (evaluated at $\widehat{\boldsymbol{\theta}}_{GMM} = \widehat{\boldsymbol{\theta}}$) yields,

$$\begin{aligned} 0 &= -\frac{2}{n^2}\left(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\widehat{\boldsymbol{\theta}}_n - \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{y}\right) \\ &= -\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\widehat{\boldsymbol{\theta}}_n + \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{y}\right) \end{aligned}$$

provided that $\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}$ is non-singular, the estimator is given by:

$$\widehat{\boldsymbol{\theta}}_n = \left\{\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\right\}^{-1}\left\{\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{y}\right)\right\}$$

(2.2) Substituting for $\boldsymbol{y}$ gives:

$$\widehat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} + \left\{\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\right\}^{-1}\left\{\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{u}\right)\right\} \tag{5}$$

then (Slutsky's theorem):

$$\text{plim } \widehat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} + \left\{\text{plim }\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\text{plim }(\boldsymbol{W}_n)\,\text{plim }\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\right\}^{-1} \times \tag{6}$$

$$\left\{\text{plim }\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\text{plim }(\boldsymbol{W}_n)\,\text{plim }\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{u}\right)\right\} \tag{7}$$

where $\boldsymbol{W}$ is a positive definite symmetric matrix defined below as:

$$\text{plim } (\boldsymbol{W}_n) = \boldsymbol{W}$$

The limiting behaviour of the other matrices in can be deduced from the WLLN:

$$\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i' \xrightarrow{p} \mathbb{E}[\boldsymbol{z}_i\boldsymbol{x}_i']$$

$$\frac{1}{n}\boldsymbol{Z}'\boldsymbol{u} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{u}_i' \xrightarrow{p} \mathbb{E}[\boldsymbol{z}_i\boldsymbol{u}_i']$$

At this stage the population moment and identification conditions become important. The identification condition states that $\mathbb{E}[\boldsymbol{z}_i\boldsymbol{x}_i']$ must be of rank k such that the inverse form $\mathbb{E}[\boldsymbol{x}_i\boldsymbol{z}_i']\boldsymbol{W}\mathbb{E}[\boldsymbol{z}_i\boldsymbol{x}_i']$ exists. Then define:

$$\boldsymbol{M} = \left\{\left(\boldsymbol{W}^{\frac{1}{2}}\mathbb{E}[\boldsymbol{z}_i\boldsymbol{x}_i']\right)'\left(\boldsymbol{W}^{\frac{1}{2}}\mathbb{E}[\boldsymbol{z}_i\boldsymbol{x}_i']\right)\right\}^{-1}\left(\boldsymbol{W}^{\frac{1}{2}}\mathbb{E}[\boldsymbol{z}_i\boldsymbol{x}_i']\right)'\boldsymbol{W}^{\frac{1}{2}}$$

Re-arranging equation (7) using $\boldsymbol{M}$ and that $\mathbb{E}[\boldsymbol{z}_i\boldsymbol{u}_i'] = \boldsymbol{0}$, bring the consistency of $\widehat{\boldsymbol{\theta}}_n$ as:

$$\text{plim } \widehat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} + \boldsymbol{M}\mathbb{E}[\boldsymbol{z}_i\boldsymbol{u}_i'] = \boldsymbol{\theta} \tag{8}$$

the asymptotic distribution of the random variable $\widehat{\boldsymbol{\theta}}_n$ can be shown by re-arranging equation (5) in the following way:

$$\frac{1}{\sqrt{n}}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) = \left\{\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\right\}^{-1}\left\{\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{\sqrt{n}}\boldsymbol{Z}'\boldsymbol{u}\right)\right\} \tag{9}$$

before applying the CLT, we should assume that the $\boldsymbol{z}_i\boldsymbol{u}_i$ is an i.i.d. random sequence as this simplifies cross-dependencies (no covariances in term $\boldsymbol{S}$ below) to write:

$$\frac{1}{\sqrt{n}}\boldsymbol{Z}'\boldsymbol{u} = \frac{1}{\sqrt{n}}\sum_{i=1}^{T}\boldsymbol{z}_i\boldsymbol{u}_i \xrightarrow{d} \mathcal{N}(0, S) \tag{10}$$

$$\boldsymbol{S} = \lim_{n\to\infty}\text{var}\left[\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{u}_i\right] \tag{11}$$

and the mean of this distribution follows from the population moment condition. Re-arranging equation (9) gives:

$$\frac{1}{\sqrt{n}}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{M}\boldsymbol{S}\boldsymbol{M}') \tag{12}$$

Note that when $k = q$ then $\boldsymbol{M}$ reduces to $\mathbb{E}[\boldsymbol{z}_i\boldsymbol{x}_i']$ and that $\boldsymbol{M}\boldsymbol{S}\boldsymbol{M}' = \{\mathbb{E}[\boldsymbol{z}_i\boldsymbol{x}_i']\}^{-1}\,\mathbb{E}[\boldsymbol{x}_i\boldsymbol{z}_i']$

(2.3) (This part is not examinable) From equation (12) an approximate large sample $100(1-\alpha)\%$

confidence interval for $\theta_{0,i}$ is:

$$\widehat{\theta}_{GMM,n,j} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\widehat{V}_{n,jj}}{n}} \tag{13}$$

where $\widehat{V}_{n,jj}$ is the j-j^{th} element of a consistent estimator of $\boldsymbol{MSM'}$ and $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution. A consistent estimator of $\boldsymbol{MSM'}$ can be obtained from consistent estimators of its components because by Slutsky's Theorem if

$$\widehat{\boldsymbol{M}}_n \xrightarrow{p} \boldsymbol{M} \tag{14}$$

$$\widehat{\boldsymbol{S}}_n \xrightarrow{p} \boldsymbol{S} \tag{15}$$

then $\widehat{\boldsymbol{M}}_n \widehat{\boldsymbol{S}}_n \widehat{\boldsymbol{M}}_n \xrightarrow{p} \boldsymbol{MSM'}$. One candidate for $\widehat{\boldsymbol{M}}_n$ is $\left(\frac{1}{n}\boldsymbol{X'Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z'X}\right)\left(\frac{1}{n}\boldsymbol{X'Z}\right)\boldsymbol{W}_n$ as it converges in probability to $\boldsymbol{M}$ in previous part, however, to construct $\widehat{\boldsymbol{S}}_n$, assuming an independently and identically distributed sequence with a mean of zero $\boldsymbol{z}_i\boldsymbol{u}_i$ and also that $\mathbb{E}[u_i u_s \boldsymbol{z}_i \boldsymbol{z}_s'] = \mathbb{E}[u^2 \boldsymbol{zz}]\ \forall_{i=s}$ and that $\mathbb{E}[u_i u_s \boldsymbol{z}_i \boldsymbol{z}_s'] = \mathbb{E}[u^2 \boldsymbol{zz}] = 0\ \forall_{i\neq s}$ then:

$$\boldsymbol{S} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{s=1}^{n} \mathbb{E}[u_i u_s \boldsymbol{z}_i \boldsymbol{z}_s'] = \mathbb{E}[u^2 \boldsymbol{zz}'] \tag{16}$$

which can be estimated consistently by,

$$\widehat{\boldsymbol{S}}_n = \frac{1}{n} \sum_{i=1}^{n} \widehat{u}_i^2 \boldsymbol{z}_i \boldsymbol{z}_i' \tag{17}$$

where $\widehat{u}_i = y_i - \boldsymbol{x}_i' \widehat{\boldsymbol{\theta}}_n$.