Micro-econometrics: Notes on Week 7 Exercise

Julia Giese

Julia.Giese@Nuffield.ox.ac.uk

Some remarks:

- If you are planning to do empirical time series modelling either in your DPhil or in your job, I would recommend a summer school at the University of Copenhagen, entitled: ***"Econometric Methodology and Macroeconomics Applications - the Cointegrated VAR Model"***. It is run by Sœren Johansen, Katarina Juselius and Anders Rahbek and takes place from 4-24 August 2008. Besides technical lectures, you get the chance to work on your data set with the experts for three weeks – and have a paper written by the end of it!

  For further information talk to me or visit `www.econ.ku.dk/Summerschool` .

Notes on suggested solutions:[1]

1.4. We are asked for the variance of the estimator $\hat{\theta}$. One way to do this is to recognise that the asymptotic variance of $\hat{\theta}$ is given by the inverse of the information matrix (i.e. minus the expectation of the second derivative of the likelihood function, as seen in class and in Victoria's solutions. Another way goes as follows:

$$
\begin{aligned}
Var(\hat{\theta}) &= Var\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) && (1)\\
&= \frac{1}{N^2} Var\left(\sum_{i=1}^{N} x_i\right) && (2)\\
&= \frac{1}{N^2}\sum_{i=1}^{N} Var\left(x_i\right) \quad [x_i \text{ independent}] && (3)\\
&= \frac{1}{N^2}\sum_{i=1}^{N} \left[E\left(x_i^2\right) - E\left(x_i\right)^2\right]. && (4)
\end{aligned}
$$

From (4) we can see we need to also find out what $E(x_i)^2$ is. Integrating by parts:

$$
\begin{aligned}
E(x_i)^2 &= \int_0^{\infty} x_i^2 \frac{1}{\theta} \exp\left(-\frac{x_i}{\theta}\right) dx && (5)\\
&= \left[-x_i^2 \exp\left(-\frac{x_i}{\theta}\right)\right]_0^{\infty} + \int_0^{\infty} 2x_i \exp\left(-\frac{x_i}{\theta}\right) dx && (6)\\
&= 0 + 2\theta \int_o^{\infty} x \frac{1}{\theta} \exp\left(-\frac{x_i}{\theta}\right) dx && (7)\\
&= 2\theta^2. && (8)
\end{aligned}
$$

[1] Many thanks to James Reade for providing notes from last year.

Now the last line (to (8)) uses the earlier integration by parts result in 1.1, and so we can use (8) in (4):

$$Var(\hat{\theta}) = \frac{1}{N^2}\sum_{i=1}^{N}\left[E\left(x_i^2\right) - E\left(x_i\right)^2\right] \tag{9}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\left[2\theta^2 - \theta^2\right] = \frac{\theta^2}{N}. \tag{10}$$

Note that this is the asymptotic variance of $\hat{\theta}$. In a large sample we would approximate the variance by using $\hat{Var}(\hat{\theta}) = \frac{\hat{\theta}^2}{N}$ as the true θ is unknown.

1.5. We are now told that $x_i \sim \mathsf{IN}\left[\theta, \sigma^2\right]$, and asked to comment on the estimators we have, which are the estimators for θ and for $Var(\hat{\theta})$, i.e. $\hat{\theta}$ and $\hat{\theta}/N$ respectively. The properties we should check are bias, consistency and efficiency. So beginning with $\hat{\theta}$,

$$E(\hat{\theta}) = E\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) = \frac{1}{N}\sum_{i=1}^{N} E\left(x_i\right) = \theta, \tag{11}$$

so our estimator is unbiased. Consistency:

$$plim(\hat{\theta}) = plim\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) \overset{SLLN}{=} E(x_i) = \theta, \tag{12}$$

hence our estimator is asymptotically unbiased, but we also need to check the variance.

$$Var(\hat{\theta}) = Var\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right) \tag{13}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} Var\left(x_i\right) \quad [x_i \text{ independent}] \tag{14}$$

$$= \frac{1}{N^2}N\sigma^2 = \frac{\sigma^2}{N}. \tag{15}$$

Hence the variance of θ goes to zero asymptotically, which together with asymptotic unbiasedness implies consistency.

We know that the maximum likelihood estimator attains the Cramer-Rao lower bound, the minimum variance estimator for θ, and we can show that $\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is the maximum likelihood estimator for θ also when the distribution is normal. Hence the correctly specified estimator, i.e. that using the Normal distribution, call it $\hat{\theta}_N$ must attain this, and so (15) must be the Cramer-Rao lower bound. So given that we found in (10) that when using the exponential distribution, the maximum likelihood estimator $\hat{\theta}_E$ has variance θ^2/N, then it must be that:

$$Var(\hat{\theta}_N) \leq Var(\hat{\theta}_E), \tag{16}$$

with equality only when $\theta = \sigma$. Thus unless $\theta = \sigma$, $\hat{\theta}_E$ is inefficient.

Turning next to the estimator for $Var(\hat{\theta})$, $\hat{Var}(\hat{\theta})$, then we can see that comparing (10) to (15) that if we use the exponential distribution, then again unless $\theta = \sigma$, we will get a different estimate for $Var(\hat{\theta})$ ($\hat{\theta}^2/N$ and not correct $\hat{\sigma}^2/N$). Hence $\hat{Var}(\hat{\theta})$ is biased.

2.1 **The Probit model.** Y_i is such that:

$$\mathsf{P}\left(y_i = 1 \,|x_i; \beta\right) = F(x_i'\beta). \tag{17}$$

Such a model could be used for an outcome of an event, y_i is binary, or dichotomous. Such as buying a good or not, participating in the labour market or not. x_i denote observed characteristics of individuals that influence the outcome.

2.2. The likelihood function considers the likelihood of observing particular parameters given the data. We can think of the likelihood as the probability of y_i, and so we can use the binary nature of y_i to write:

$$\mathsf{P}\left(y_i \,|x_i; \beta\right) = \mathsf{P}\left(y_i = 1 \,|x_i; \beta\right)^{y_i} \mathsf{P}\left(y_i = 0 \,|x_i; \beta\right)^{y_i - 1}. \tag{18}$$

Independence of the N observations means that the joint density of the N observations is the product of the density of the N observations:

$$\begin{aligned} L(\beta; x) &= \prod_{i=1}^{N} \mathsf{P}\left(y_i = 1\right)^{y_i} \mathsf{P}\left(y_i = 0\right)^{1-y_i}. && (19) \\ &= \prod_{i=1}^{N} F(x_i'\beta)^{y_i}(1 - F(x_i'\beta))^{1-y_i}. && (20) \end{aligned}$$

We can then take logs:

$$\log L(\beta; x) = \sum_{i=1}^{N} \left\{ y_i \log F\left(x_i'\beta\right) + (1 - y_i) \log\left[1 - F(x_i'\beta)\right] \right\}. \tag{21}$$

Having got the log-likelihood function in (21) we can now proceed to the score by differentiating. We note that convention tells us that:

$$F'(x_i'\beta) = f(x_i'\beta)x_{ik}, \tag{22}$$

when differentiation is with respect to the k^{th} element of β, so:

$$\begin{aligned} \frac{\partial \log L(\beta; x)}{\partial \beta_k} &= \sum_{i=1}^{N} \left\{ \frac{y_i}{F(x_i'\beta)} f(x_i'\beta)x_{ik} + \frac{(1 - y_i)}{1 - F(x_i'\beta)} \left[-f(x_i'\beta)x_{ik}\right] \right\} && (23) \\ &= \sum_{i=1}^{N} \left\{ \frac{y_i f(x_i'\beta)x_{ik}}{F(x_i'\beta)} - \frac{f(x_i'\beta)x_{ik}}{1 - F(x_i'\beta)} + \frac{y_i f(x_i'\beta)x_{ik}}{1 - F(x_i'\beta)} \right\} && (24) \\ &= \sum_{i=1}^{N} \left\{ \frac{\left[y_i - F(x_i'\beta)\right] f(x_i'\beta)x_{ik}}{F(x_i'\beta)\left[1 - F(x_i'\beta)\right]} \right\}. && (25) \end{aligned}$$

2.3. Next we add the distributional assumption of normality. So:

$$F(x_i'\beta) = \Phi(x_i'\beta). \tag{26}$$

There is no need to use the definition of the PDF or CDF of the standard normal distribution, this only complicates matters. So substituting (26) into (20):

$$\begin{aligned} L(\beta; x) &= \prod_{i=1}^{N} \mathsf{P}\left(y_i = 1\right)^{y_i} \mathsf{P}\left(y_i = 0\right)^{1-y_i}. && (27) \\ &= \prod_{i=1}^{N} \Phi(x_i'\beta)^{y_i}(1 - \Phi(x_i'\beta))^{1-y_i}. && (28) \end{aligned}$$

$$\log L(\beta;x) = \sum_{i=1}^{N} \left\{ y_i \log \Phi\left(x_i'\beta\right) + (1-y_i)\log\left[1-\Phi(x_i'\beta)\right] \right\}. \tag{29}$$

Then differentiating to find the score, we note the standard normal equivalent of (22):

$$\Phi'(x_i'\beta) = \phi(x_i'\beta)x_{ik},$$

and use this to find the score:

$$\frac{\partial \log L(\beta;x)}{\partial \beta_k} = \sum_{i=1}^{N} \left\{ \frac{y_i}{\Phi(x_i'\beta)}\phi(x_i'\beta)x_{ik} + \frac{(1-y_i)}{1-\Phi(x_i'\beta)}\left[-\phi(x_i'\beta)x_{ik}\right] \right\} \tag{30}$$

$$= \sum_{i=1}^{N} \left\{ \frac{\left[y_i - \Phi(x_i'\beta)\right]\phi(x_i'\beta)x_{ik}}{\Phi(x_i'\beta)\left[1-\Phi(x_i'\beta)\right]} \right\}. \tag{31}$$

2.4. A latent variable is an unobservable, non-binary, variable that determines the binary outcome of y_i, e.g.:

$$\begin{aligned} y_i &= 1 \quad \text{if} \quad y_i^* > 0, \\ y_i &= 0 \quad \text{if} \quad y_i^* \leq 0. \end{aligned} \tag{32}$$

where $y_i^* = x_i'\beta + \varepsilon_i$ with x_i still being individual characteristics. In a model where y_i denote participation or non-participation in the labour market, y_i^* could be the wage level. The person would go to work if the wage is above her reservation level. In that example, x_i could be the number of children in the household.

We can then show that the form of this model takes that above and we can use the score and likelihood functions above:

$$\mathsf{P}\left(y_i = 1\,|x_i\right) = \mathsf{P}\left(y_i^* > 0\,|x_i\right) = \mathsf{P}\left(\varepsilon_i > -x_i'\beta\right) = \mathsf{P}\left(\varepsilon_i/\sigma > -x_i'(\beta/\sigma)\right) = F(x_i'\beta),$$

so we have our $F(x_i'\beta)$ as before. We also note that we can only estimate β/σ, and hence need to normalise $\sigma = 1$ to identify β.