PhD Econometrics 1: Study Questions Class 5
Imperial College London
Hormoz Ramian

Solutions

Question 1 If $u_t \sim \mathcal{N}$ with zero mean and variance σ^2, then the density of u_t is:

$$f(u_t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u_t^2}{2\sigma^2}} \tag{1}$$

We need to construct the density of y_t to set up the ML objective function. However, note that y_t is a nonlinear function of u_t

$$\log\left(\frac{y_t}{1-y_t}\right) = \ldots + u_t \tag{2}$$

In this problem, first we substitute instead of u_t in the density of u_t:

$$f\left[\log\left(\frac{y_t}{1-y_t}\right) - \boldsymbol{X}_t\boldsymbol{\beta}\right] = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left\{\log\left(\frac{y_t}{1-y_t}\right)-\boldsymbol{X}_t\boldsymbol{\beta}\right\}^2}{2\sigma^2}} \tag{3}$$

Second, we need to work out an additional term called the Jacobian term:

$$\frac{d}{dy_t}\log\left(\frac{y_t}{1-y_t}\right) = \frac{d}{dy_t}(\log(y_t) - \log(1-y_t)) = \frac{1}{y_t(1-y_t)} \tag{4}$$

The density of y_t is constructed using:

$$\text{density}(y_t) = f\left[\log\left(\frac{y_t}{1-y_t}\right) - \boldsymbol{X}_t\boldsymbol{\beta}\right] \times \left|\frac{d}{dy_t}\log\left(\frac{y_t}{1-y_t}\right)\right| \tag{5}$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left\{\log\left(\frac{y_t}{1-y_t}\right)-\boldsymbol{X}_t\boldsymbol{\beta}\right\}^2}{2\sigma^2}} \times \frac{1}{y_t(1-y_t)} \tag{6}$$

where |...| is the absolute value operator. We can now take log of the last expression and setup the log-likelihood function $\ell(\boldsymbol{y};\boldsymbol{\beta},\sigma)$ or $\ell(\boldsymbol{y};\boldsymbol{\beta},\sigma^2)$. Hence the log-likelihood function is:

$$\ell(\boldsymbol{y},\boldsymbol{\beta},\sigma) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \sum_{t\in T}\log y_t - \sum_{t\in T}\log(1-y_t) - \frac{1}{2\sigma^2}\sum_{t\in T}(\log(y_t/(1-y_t))\boldsymbol{X}_t\boldsymbol{\beta})^2$$

Noting that $-\frac{n}{2}\log(2\pi)$ is just a constant and inconsequential through the maximization, then:

$$\ell(\boldsymbol{y},\boldsymbol{\beta},\sigma) \propto -\frac{n}{2}\log\sigma^2 - \sum_{t\in T}\log y_t - \sum_{t\in T}\log(1-y_t) - \frac{1}{2\sigma^2}\sum_{t\in T}(\log(y_t/(1-y_t))\boldsymbol{X}_t\boldsymbol{\beta})^2$$

Question 2 Since the null hypothesis is linear restriction, we can rewrite it as:

$$H_0 : \boldsymbol{R\theta} = \boldsymbol{r} \tag{7}$$

where $\boldsymbol{R}$ is a $q \times k$ matrix, $\boldsymbol{\theta} = \begin{bmatrix}\beta_1 & \beta_2 & \beta_3\end{bmatrix}'$ and $\boldsymbol{r}$ is a $q \times 1$ vector of known constants (q is the number of restrictions and k is the number of parameters):

$$\boldsymbol{R} = \begin{bmatrix}0 & 0 & 1\end{bmatrix}, \quad \boldsymbol{r} = \begin{bmatrix}-1\end{bmatrix} \tag{8}$$

Likelihood Ratio (LR) Test — Let $\widehat{\boldsymbol{\theta}}$ denote the ML estimates of true $\boldsymbol{\theta}$ without applying the constraint and let $\widetilde{\boldsymbol{\theta}}$ denote the ML estimates when the constraint is applied to the ML estimation. Moreover, let $\ell(\widehat{\boldsymbol{\theta}})$ and $\ell(\widetilde{\boldsymbol{\theta}})$ denote log-likelihood functions the unrestricted and restricted cases, respectively. Then the formal LR test uses the following test-statistic:

$$LR = 2(\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widetilde{\boldsymbol{\theta}})) \sim \chi^2_q \tag{9}$$

This can alternatively be written as:

$$LR = n\log\left(1+\frac{\widetilde{\boldsymbol{e}}'\widetilde{\boldsymbol{e}}-\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}\right) \tag{10}$$

This test requires knowledge of both the unrestricted and restricted models which could be costly or computationally intensive for studies with large datasets or complex models. The likelihood ratio can be written as:

$$\begin{aligned}
LR &= n\log\left(1+\frac{\widetilde{\boldsymbol{e}}'\widetilde{\boldsymbol{e}}-\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}\right) &(11)\\
&= n\log\left(1+\frac{(\boldsymbol{R}\widehat{\boldsymbol{\theta}}-\boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\theta}}-\boldsymbol{r})}{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}\right) &(12)\\
&= n\log\left(1+\frac{(\widehat{\beta}_3+1)[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\widehat{\beta}_3+1)}{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}\right) &(13)
\end{aligned}$$

Assuming our former notation for $(\boldsymbol{X}'\boldsymbol{X})^{-1}=\{c_{ij}\}$ and that $\boldsymbol{R}=\begin{bmatrix}1 & 0 & 1\end{bmatrix}$, then:

$$[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'] = c_{33} \tag{14}$$

Substituting this into equation (13):

$$\begin{aligned}
LR &= n\log\left(1+\frac{(\widehat{\beta}_3+1)[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\widehat{\beta}_3+1)}{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}\right)\\
&= n\log\left(1+\frac{(\widehat{\beta}_3+1)^2}{c_{33}\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}}\right)\\
&= n\log\left(1+\frac{(\widehat{\beta}_3+1)^2}{nc_{33}\widehat{\sigma}^2}\right)
\end{aligned}$$

Hence, the null hypothesis will be rejected if this value exceeds the critical value from the chi-squared distribution at a certain confidence level.

Wald (W) Test — This test only relies on the unrestricted model and uses the asymptotic Normality properties of $\widehat{\boldsymbol{\theta}}$ with asymptotic variance $\mathcal{I}(\boldsymbol{\theta})$ with the following test-statistic (with linear constraint):

$$\begin{aligned}
W &= \left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}-\boldsymbol{r}\right)'\left(\boldsymbol{R}I^{-1}(\widehat{\theta})\boldsymbol{R}\right)^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}-\boldsymbol{r}\right)\sim\chi^2_q &(15)\\
&= \frac{\left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}-\boldsymbol{r}\right)'\left(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\right)^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}-\boldsymbol{r}\right)}{\widehat{\sigma}^2} &(16)\\
&= \frac{(\widehat{\beta}_3+1)^2}{c_{33}\widehat{\sigma}^2} &(17)
\end{aligned}$$

A large value of W will lead to rejection of the null hypothesis.

Lagrange Multiplier (LM) Test — Contrary to the Wald test, LM uses restricted model with a $(q\times 1)$ vector of Lagrange multipliers $\boldsymbol{\mu}$ defined for the following constrained optimization problem:

$$\ell(\widetilde{\boldsymbol{\theta}}) = \ell(\widehat{\boldsymbol{\theta}})-\boldsymbol{\mu}'\left(\boldsymbol{R}\widehat{\boldsymbol{\theta}}-\boldsymbol{r}\right) \tag{18}$$

When the restriction is inconsequential, then imposing it should have negligible effect on the restricted log-likelihood function and the restricted and unrestricted log-likelihood functions should be statistically equal, or alternatively that elements inside $\boldsymbol{\mu}$ are very small or zero. However, when the restriction is consequential, the log-likelihood function value of the restricted model falls below the unrestricted one[1], and elements inside $\boldsymbol{\mu}$ increase, measuring the costliness of the constraint.

Denoting the gradients with $\boldsymbol{g}$, then the test-statistic is:

$$LM \;=\; \boldsymbol{g}'(\widetilde{\boldsymbol{\theta}})I^{-1}(\widetilde{\boldsymbol{\theta}})\boldsymbol{g}(\widetilde{\boldsymbol{\theta}}) \;\sim\; \chi^2_q \tag{19}$$

which simplifies to the following in case of linear restrictions:

$$LM \;=\; n\frac{\widetilde{\boldsymbol{e}}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widetilde{\boldsymbol{e}}}{\widetilde{\boldsymbol{e}}'\widetilde{\boldsymbol{e}}'} \tag{20}$$

Recall that $\boldsymbol{X'}\widetilde{e} = \boldsymbol{R}'(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})$ then:

$$\widetilde{\boldsymbol{e}}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widetilde{\boldsymbol{e}}' = \left(\boldsymbol{R}\widehat{\boldsymbol{\theta}} - \boldsymbol{r}\right)'\left(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}\right)^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\theta}} - \boldsymbol{r}\right) \tag{21}$$

therefore:

$$LM \;=\; n\frac{(\widehat{\beta}+1)^2}{c_{33}\left(\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}} + (\widehat{\beta}+1)^2 c_{33}^{-1}\right)} \tag{22}$$

$$\;=\; n\frac{(\widehat{\beta}+1)^2}{nc_{33}\widehat{\sigma}^2 + (\widehat{\beta}+1)^2} \;\sim\; \chi^2_1 \tag{23}$$

The derivatives of the log-likelihood function should be flat with respect to each parameter when the model is unrestricted. Conversely, when the constraint is applied, we expect the derivatives to be non-zero values because the constraint prevents the estimators to freely search for the flat tangents. As a result, deviation of the scores from zero can be used to measure the impact of constraint and thus its statistical significance. This methodology mainly works the the derivative and is more commonly known as the score test (interchangeable name for LM).

Finite Sample Characterization Asymptotically, three tests are equivalent with the same asymptotic distribution but in finite samples, they may give different conclusions. We can establish the following order for these test-statistics for finite samples and given only linear restrictions:

$$LM \leq LR \leq W \tag{24}$$

Question 3 There are two ways to solve this question.

Solution I: Without loss of generality, assume that $\boldsymbol{\beta}_1$ is $k_1 \times 1$ and $\boldsymbol{\beta}_2$ is $k_2 \times 1$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ is a column vector of size $k_1 + k_2 = k$. The null and alternative hypotheses are:

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0}_{k_2}$$

$$H_A : \boldsymbol{\beta}_2 \neq \mathbf{0}_{k_2}$$

The vectorized null hypothesis can be written as:

$$\underset{k_2\times k}{\boldsymbol{R}}\underset{k\times 1}{\boldsymbol{\beta}} \;=\; \underset{k_2\times 1}{\boldsymbol{r}}$$

$$\begin{bmatrix}\mathbf{0}_{k_2\times k_1} & \boldsymbol{I}_{k_2}\end{bmatrix}\boldsymbol{\beta} \;=\; \mathbf{0}_{k_2\times 1}$$

Simplifying give:

$$\boldsymbol{\beta}_2 \;=\; \mathbf{0}_{k_2\times 1}$$

[1]Recall that in MLE, we maximize the objective function and therefore higher log-likelihood function value reflects a better fit. This is in contrast with the OLS, where we minimize the objective function (SSR) and lower SSR is an indication of better fit.

The Wald-statistic is:

$$W = \frac{(\boldsymbol{R}\widehat{\boldsymbol{\theta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\theta}} - \boldsymbol{r})}{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}} \tag{25}$$

Let $\boldsymbol{X}'\boldsymbol{X} = \left(\begin{bmatrix}\boldsymbol{x}_1' \\ \boldsymbol{x}_2'\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2\end{bmatrix}\right)$ and $\det(\boldsymbol{X}'\boldsymbol{X})$ be a scalar function:

$$\begin{aligned}
W &= \frac{\widehat{\boldsymbol{\beta}}_2'\left\{\begin{bmatrix}\boldsymbol{0}_{k_2\times k_1} & \boldsymbol{I}_{k_2}\end{bmatrix}\left(\begin{bmatrix}\boldsymbol{x}_1' \\ \boldsymbol{x}_2'\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{0}_{k_2\times k_1} & \boldsymbol{I}_{k_2}\end{bmatrix}'\right\}^{-1}\widehat{\boldsymbol{\beta}}_2}{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}} \\
&= \frac{\widehat{\boldsymbol{\beta}}_2'\left\{\begin{bmatrix}\boldsymbol{0}_{k_2\times k_1} & \boldsymbol{I}_{k_2}\end{bmatrix}\left(\begin{bmatrix}\boldsymbol{x}_1'\boldsymbol{x}_1 & \boldsymbol{x}_1'\boldsymbol{x}_2 \\ \boldsymbol{x}_2'\boldsymbol{x}_1 & \boldsymbol{x}_2'\boldsymbol{x}_2\end{bmatrix}\right)^{-1}\begin{bmatrix}\boldsymbol{0}_{k_2\times k_1}' \\ \boldsymbol{I}_{k_2}'\end{bmatrix}\right\}^{-1}\widehat{\boldsymbol{\beta}}_2}{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}} \\
&= \frac{\widehat{\boldsymbol{\beta}}_2'\left\{\begin{bmatrix}\boldsymbol{0}_{k_2\times k_1} & \boldsymbol{I}_{k_2}\end{bmatrix}\begin{bmatrix}\boldsymbol{x}_1'\boldsymbol{x}_1 & \boldsymbol{x}_1'\boldsymbol{x}_2 \\ \boldsymbol{x}_2'\boldsymbol{x}_1 & \boldsymbol{x}_2'\boldsymbol{x}_2\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{0}_{k_1\times k_2} \\ \boldsymbol{I}_{k_2}\end{bmatrix}\right\}^{-1}\widehat{\boldsymbol{\beta}}_2}{\widehat{\boldsymbol{e}}'\widehat{\boldsymbol{e}}} \\
&= \frac{\det(\boldsymbol{X}'\boldsymbol{X})}{n\widehat{\sigma}^2}\widehat{\boldsymbol{\beta}}_2'\left\{\begin{bmatrix}\boldsymbol{0}_{k_2\times k_1} & \boldsymbol{I}_{k_2}\end{bmatrix}\operatorname{Adj}\left(\begin{bmatrix}\boldsymbol{x}_1'\boldsymbol{x}_1 & \boldsymbol{x}_1'\boldsymbol{x}_2 \\ \boldsymbol{x}_2'\boldsymbol{x}_1 & \boldsymbol{x}_2'\boldsymbol{x}_2\end{bmatrix}\right)\begin{bmatrix}\boldsymbol{0}_{k_1\times k_2} \\ \boldsymbol{I}_{k_2}\end{bmatrix}\right\}^{-1}\widehat{\boldsymbol{\beta}}_2
\end{aligned}$$

Solution II: Recall the FWL theorem is readily applicable to show that:

$$\widehat{\boldsymbol{\beta}}_2 = (\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x_2})^{-1}\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{y} \tag{26}$$

then the Wald-statistic is:

$$W = \frac{1}{\widehat{\sigma}^2}\boldsymbol{y}'\boldsymbol{M}_1\boldsymbol{x}_2(\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{x_2})^{-1}\boldsymbol{x}_2'\boldsymbol{M}_1\boldsymbol{y} \tag{27}$$

The Wald statistic is indeed a deterministic, strictly increasing function of the conventional F-statistic, $W = \frac{nk_2}{n-k}F$. Note that, if the Wald-statistic used s^2 instead of $\widehat{\sigma}^2$, the relationship between W and F would be even simpler: $W = k_2F$. Although this result can only be shown for the case of zero restrictions, it is true for arbitrary linear restrictions, because we can always re-parameterize a linear regression model so that arbitrary linear restrictions become zero restrictions.

Question 4 The ML estimator is driven from the likelihood function which is equal to:

$$\mathcal{L}(\theta; x) = f(x; \theta) \tag{28}$$

where the right hand size is the joint probability density function of the random variable x given the parameter $\{\theta_1, \theta_2\}$. However, we have to take $\theta_1 \leq x \leq \theta_2$ as a constraint into account as the support for this uniform distribution is determined by the parameters, and that we wish the parameters to construct a density function which covers are sample $\{x_i\}_{i=1}^N$:

$$\begin{aligned}
\theta_1^*, \theta_2^* &= \arg\max\ [\mathcal{L}(\theta; x)] \\
&= \arg\max\ \prod_{i=1}^N \frac{1}{\theta_2 - \theta_1} \\
&= \arg\max\ \frac{1}{(\theta_2 - \theta_1)^N} \\
&\propto \arg\min\ (\theta_2 - \theta_1)^N \\
&\propto \arg\min\ (\theta_2 - \theta_1)^N \\
&\propto \arg\min\ \theta_2 - \theta_1
\end{aligned}$$

subject to $\theta_1 \leq x \leq \theta_2$. Note that without the constraint the solution would be to set $\theta_1^* = \theta_2^*$ as this maximizes the likelihood function but this would result in a degenerate probability density function condenses on one value which is not desired. However, the simplified likelihood function subject to $\theta_1 \leq x \leq \theta_2$ picks the boundary values such that $\theta_1^* = \min\{x_i\}_{i=1}^N$ and $\theta_2^* = \max\{x_i\}_{i=1}^N$.