Advanced Econometrics II
TA Session Problems No. 5

Agnieszka Borowska

03.02.2015

Note: this is only a draft of the problems discussed on Tuesday and might contain some typos or more or less imprecise statements. If you find some, please let me know.

1. ML testing
2. Reparametrisation

1 ML testing

- Model under the alternative: characterized by $l(\theta)$;
  Model under the null: with r (nonlinear) restrictions imposed.
- Restrictions depend on θ: $r(\theta)$, wlog $r(\theta) = 0$; r - smooth function.
- Model under the null: with $l(\theta)$ where $\theta \in \Theta \cap \{\theta : r(\theta) = 0\}$ (restricted parameter space).
- **Three classical tests:** asymptotically equivalent, $\overset{a}{\sim} \chi^2(r)$.

1.1 LR test

Test statistic: twice the difference between the unconstrained maximum value of the loglikelihood function and the maximum subject to the restrictions:

$$LR = 2\left(l(\hat{\theta}) - l(\tilde{\theta})\right) = 2\log\left(\frac{L(\hat{\theta})}{L(\tilde{\theta})}\right),$$

$\hat{\theta}$ – unrestricted MLE, $\tilde{\theta}$ – restricted MLE.
Note: invariant to reformulation of the restrictions [below].

1.2 Wald test

Depends only on the estimates of the unrestricted model.
Test statistic: a quadratic form in $r(\hat{\theta})$ and the inverse of its covariance matrix estimate, $\mathrm{Var}\,(\hat{\theta})$:

$$W = r^T(\hat{\theta})\left(R(\hat{\theta})\widehat{\mathrm{Var}}(\hat{\theta})R^T(\hat{\theta})\right)^{-1} r(\hat{\theta}),$$

where $R(\theta)$ is an $r \times k$ matrix with typical element $\partial r_j(\theta)/\partial\theta_i$ and $\widehat{\mathrm{Var}}(\hat{\theta})$ is an estimate of

$$\mathrm{Var}\,(r(\hat{\theta})) \overset{a}{=} R(\theta_0)\mathrm{Var}(\hat{\theta})R^T(\theta_0), \qquad \text{(delta method)}$$

for example the empirical Hessian (minus the inverse of the Hessian):

$$\widehat{\mathrm{Var}}(\hat{\theta}) = -H^{-1}(\hat{\theta}).$$

Note: not invariant to reformulation of the restrictions [homework].

1.3 LM test

Principle: based on the vector of Lagrange multipliers from a constrained maximization problem.
Practice: usually based on the gradient (score) vector of the unrestricted loglikelihood function, evaluated at the restricted estimates.

Test statistic: (in the score version)

$$LM = g^T(\tilde{\theta})\mathbf{I}^{-1}(\tilde{\theta})g(\tilde{\theta}),$$

where $g(\tilde{\theta})$ – the gradient estimate and $\mathbf{I}(\tilde{\theta})$ – the information matrix estimate (at the restricted MLE).
Note: invariant to reformulation of the restrictions [below].

2 Reparametrisation

The model specified by the likelihood function $l(\theta)$ is said to be **reparametrised** if the parameter vector θ is replaced by another parameter vector ϕ related to θ by a **one-to-one** relationship

$$\theta = T(\phi) \qquad \text{with inverse} \qquad \phi = T^{-1}(\theta).$$

2.1 Loglikelihood for reparametrised model

- The **loglikelihood function for the reparametrised model** is defined as

  $$l'(\phi) \equiv l(T(\phi)).$$

  Why does this definition make sense?
  The definition implies that the **joint densities** for y are the same under both parametrisations, which means that the reparametrisation does not change the underlying DGP yielding the sample y.

- The MLEs of $\hat{\phi}$ of the reparametrised model are related to the MLEs $\hat{\theta}$ of the original model by the relation

  $$\hat{\theta} = T(\hat{\phi}).$$

  Why? Since $\hat{\theta}$ - MLE for model specified by $l(\theta)$, we have

  $$l(\hat{\theta}) \geq l(\theta), \quad \forall\theta.$$

  Next, as T is a bijection and we have $\theta = T(\phi)$, so $\exists\tilde{\phi}$ such that $\hat{\theta} = T(\tilde{\phi})$, so the above inequality is equivalent to

  $$l(T(\tilde{\phi})) \geq l(T(\phi)), \quad \forall\phi.$$

  This implies

  $$l'(\tilde{\phi}) \geq l'(\phi), \quad \forall\phi.$$

  So in fact $\tilde{\phi} = T^{-1}(\theta)$ is MLE for the reparametrised model, which we will denote $\hat{\phi}$.

- Next, we will specify the relationship between the gradients and information matrices of the two models in terms of the derivatives of the components of θ with respect to those of ϕ.

  To get the relationship between the **gradients**, differentiate the defining identity $l'(\phi) \equiv l(T(\phi))$ with respect to ϕ. This results in

  $$g'(\phi) = J(\phi)g(\theta), \tag{1}$$

  where $J(\phi)$ is a $k \times k$ matrix with typical element $\partial T_j(\phi)/\partial\phi_i$. Due to the assumed invertability of the mapping T, we have

  $$g(\theta) = J^{-1}(\phi)g'(\phi), \tag{2}$$

  which is the required relationship between the gradients.

- To get the relationship between the **information matrices**, recall the result from the last TA session that the information matrix is equivalent to the **covariance matrix of the gradient vector**, i.e.

$$\mathbf{I}(\theta) = \mathbb{E}_\theta \left[g(\theta) g^T(\theta) \right].$$

Using (1) we can write

$$\begin{aligned} \mathbf{I}'(\phi) &= \mathbb{E}_\phi \left[g'(\phi)(g')^T(\phi) \right] \\ &= J(\theta)\mathbb{E}_\theta \left[g(\theta) g^T(\theta) \right] J^T(\theta) \\ &= J(\theta)\mathbf{I}(\theta)J^T(\theta), \end{aligned}$$

and

$$\begin{aligned} \mathbf{I}(\theta) &= \mathbb{E}_\theta \left[g(\theta) g^T(\theta) \right] \\ &= J^{-1}(\phi)\mathbb{E}_\phi \left[g'(\phi)(g')^T(\phi) \right] (J^T)^{-1}(\phi) \\ &= J^{-1}(\phi)\mathbf{I}'(\phi)(J^T)^{-1}(\phi), \end{aligned} \tag{3}$$

which describes the relationship between the information matrices.

2.2 Testing for reparametrised model

The set of r restrictions

$$r(\theta) = 0$$

can be adapted to the reparametrised model as follows

$$r'(\phi) \equiv r(T(\phi)) = 0.$$

Aim: show that both, the LR statistic and the LM statistic (in the efficient score form), are is **invariant** to whether the restrictions are tested for the original or the reparametrised model.

- **LR statistic**
Obvious, as the above results hold for both, the unrestricted estimates ($\hat{\theta}$ and $\hat{\phi}$), and the restricted estimates ($\tilde{\theta}$ and $\tilde{\phi}$). Hence, we have

$$\begin{aligned} LR &= 2(l(\hat{\theta}) - l(\tilde{\theta})) \\ &= 2(l(T(\hat{\phi})) - l(T(\tilde{\phi}))) \\ &= 2(l'(\hat{\phi}) - l'(\tilde{\phi}), \end{aligned}$$

which confirms that, indeed, the LR statistic is invariant under reparametrisation.

- **LM statistic**
Under the original parametrisation

$$LM = g^T(\tilde{\theta})\mathbf{I}^{-1}(\tilde{\theta})g(\tilde{\theta}).$$

Then, from (2) and (3) we can rewrite this in the following way

$$\begin{aligned} LM &= g^T(\tilde{\theta})\mathbf{I}^{-1}(\tilde{\theta})g(\tilde{\theta}) \\ &= (g')^T(\tilde{\phi})(J^T)^{-1}(\tilde{\phi}) \left(J^{-1}(\tilde{\phi})\mathbf{I}'(\tilde{\phi})(J^T)^{-1}(\tilde{\phi}) \right)^{-1} J^{-1}(\tilde{\phi})g'(\tilde{\phi}) \\ &= (g')^T(\tilde{\phi})(J^T)^{-1}(\tilde{\phi})J^T(\tilde{\phi})(\mathbf{I}')^{-1}(\tilde{\phi})J^{-1}(\tilde{\phi})J^{-1}(\tilde{\phi})g'(\tilde{\phi}) \\ &= (g')^T(\tilde{\phi})(\mathbf{I}')^{-1}(\tilde{\phi})g'(\tilde{\phi}), \end{aligned}$$

which is the LM statistic in the efficient score form for the reparameterized model. Hence, we obtain the desired result.