Solution to Exercise 10.15

⋆10.15 The model specified by the loglikelihood function $\ell(\boldsymbol{\theta})$ is said to be reparametrized if the parameter vector $\boldsymbol{\theta}$ is replaced by another parameter vector $\boldsymbol{\phi}$ related to $\boldsymbol{\theta}$ by a one to one relationship $\boldsymbol{\theta} = \boldsymbol{\Theta}(\boldsymbol{\phi})$ with inverse $\boldsymbol{\phi} = \boldsymbol{\Theta}^{-1}(\boldsymbol{\theta})$. The loglikelihood function for the reparametrized model is then defined as $\ell'(\boldsymbol{\phi}) \equiv \ell(\boldsymbol{\Theta}(\boldsymbol{\phi}))$. Explain why this definition makes sense.

Show that the maximum likelihood estimates $\hat{\boldsymbol{\phi}}$ of the reparametrized model are related to the estimates $\hat{\boldsymbol{\theta}}$ of the original model by the relation $\hat{\boldsymbol{\theta}} = \boldsymbol{\Theta}(\hat{\boldsymbol{\phi}})$. Specify the relationship between the gradients and information matrices of the two models in terms of the derivatives of the components of $\boldsymbol{\theta}$ with respect to those of $\boldsymbol{\phi}$.

Suppose that it is wished to test a set of r restrictions written as $\boldsymbol{r}(\boldsymbol{\theta}) = \mathbf{0}$. These restrictions can be applied to the reparametrized model in the form $\boldsymbol{r}'(\boldsymbol{\phi}) \equiv \boldsymbol{r}(\boldsymbol{\Theta}(\boldsymbol{\phi})) = \mathbf{0}$. Show that the LR statistic is invariant to whether the restrictions are tested for the original or the reparametrized model. Show that the same is true for the LM statistic (10.69).

The definition in the first part of the question makes sense because it implies that the joint densities for $\boldsymbol{y}$ are exactly the same in both parametrizations. This in turn implies that the DGPs are exactly the same.

That the ML estimates $\hat{\boldsymbol{\phi}}$ of the reparametrized model are related to the estimates $\hat{\boldsymbol{\theta}}$ of the original model by the relation $\hat{\boldsymbol{\theta}} = \boldsymbol{\Theta}(\hat{\boldsymbol{\phi}})$ follows from the fact that the inequality

$$\ell(\hat{\boldsymbol{\theta}}) \geq \ell'(\boldsymbol{\theta}) \text{ for all } \boldsymbol{\theta}$$

is equivalent to the inequality

$$\ell\big(\boldsymbol{\Theta}(\hat{\boldsymbol{\phi}})\big) \geq \ell\big(\boldsymbol{\Theta}(\boldsymbol{\phi})\big) \text{ for all } \boldsymbol{\phi},$$

and this in turn implies that

$$\ell'(\hat{\boldsymbol{\phi}}) \geq \ell'(\boldsymbol{\phi}) \text{ for all } \boldsymbol{\phi}.$$

The relationship between the gradients of the two models may be obtained by differentiating the identity $\ell'(\boldsymbol{\phi}) \equiv \ell(\boldsymbol{\Theta}(\boldsymbol{\phi}))$ with respect to the components of $\boldsymbol{\phi}$ and using the chain rule. The result is

$$\boldsymbol{g}'(\boldsymbol{\phi}) = \boldsymbol{J}(\boldsymbol{\phi})\boldsymbol{g}(\boldsymbol{\theta}), \tag{S10.33}$$

where $\boldsymbol{J}(\boldsymbol{\phi})$ is a $k \times k$ matrix with typical element $\partial\Theta_j(\boldsymbol{\phi})/\partial\phi_i$. Since the mapping $\boldsymbol{\Theta}$ is invertible, we also have

$$\boldsymbol{g}(\boldsymbol{\theta}) = \boldsymbol{J}^{-1}(\boldsymbol{\phi})\boldsymbol{g}'(\boldsymbol{\phi}). \tag{S10.34}$$

For the information matrix, we start from the definition used in Exercise 10.5, according to which

$$\boldsymbol{I}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}\big(\boldsymbol{g}(\boldsymbol{\theta})\boldsymbol{g}^\top(\boldsymbol{\theta})\big).$$

Then, from (S10.33),

$$\begin{aligned}\boldsymbol{I}'(\boldsymbol{\phi}) &= \mathrm{E}_{\boldsymbol{\phi}}\big(\boldsymbol{g}'(\boldsymbol{\phi})\boldsymbol{g}'^\top(\boldsymbol{\phi})\big)\\ &= \boldsymbol{J}(\boldsymbol{\theta})\mathrm{E}_{\boldsymbol{\theta}}\big(\boldsymbol{g}(\boldsymbol{\theta})\boldsymbol{g}^\top(\boldsymbol{\theta})\big)\boldsymbol{J}^\top(\boldsymbol{\theta}) \qquad \text{(S10.35)}\\ &= \boldsymbol{J}(\boldsymbol{\theta})\boldsymbol{I}(\boldsymbol{\theta})\boldsymbol{J}^\top(\boldsymbol{\theta}).\end{aligned}$$

Similarly,

$$\begin{aligned}\boldsymbol{I}(\boldsymbol{\theta}) &= \mathrm{E}_{\boldsymbol{\theta}}\big(\boldsymbol{g}(\boldsymbol{\theta})\boldsymbol{g}^\top(\boldsymbol{\theta})\big)\\ &= \boldsymbol{J}^{-1}(\boldsymbol{\phi})\mathrm{E}_{\boldsymbol{\theta}}\big(\boldsymbol{g}'(\boldsymbol{\phi})\boldsymbol{g}'^\top(\boldsymbol{\phi})\big)(\boldsymbol{J}^\top)^{-1}(\boldsymbol{\phi}) \qquad \text{(S10.36)}\\ &= \boldsymbol{J}^{-1}(\boldsymbol{\phi})\boldsymbol{I}'(\boldsymbol{\phi})(\boldsymbol{J}^\top)^{-1}(\boldsymbol{\phi}).\end{aligned}$$

Equations (S10.33) and (S10.34) give the relationships between the gradients for the two parametrizations, and equations (S10.35) and (S10.36) give the relationships between the information matrices.

That the LR statistic is invariant under reparametrization is obvious. The result we proved above applies to the restricted estimates $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\phi}}$ just as it does to the unrestricted ones $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$. Therefore,

$$2\big(\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})\big) = 2\big(\ell'(\hat{\boldsymbol{\phi}}) - \ell'(\tilde{\boldsymbol{\phi}})\big).$$

Proving that (10.69), the efficient score form of the LM statistic, is invariant is only a little harder. In the original parametrization, we have

$$\mathrm{LM} = \boldsymbol{g}^\top(\tilde{\boldsymbol{\theta}})\boldsymbol{I}^{-1}(\tilde{\boldsymbol{\theta}})\boldsymbol{g}(\tilde{\boldsymbol{\theta}}).$$

Using equations (S10.34) and (S10.36), this can be rewritten as

$$\begin{aligned}&\boldsymbol{g}'^\top(\tilde{\boldsymbol{\phi}})(\boldsymbol{J}^\top)^{-1}(\tilde{\boldsymbol{\phi}})\big(\boldsymbol{J}^{-1}(\tilde{\boldsymbol{\phi}})\boldsymbol{I}'(\tilde{\boldsymbol{\phi}})(\boldsymbol{J}^\top)^{-1}(\tilde{\boldsymbol{\phi}})\big)^{-1}\boldsymbol{J}^{-1}(\tilde{\boldsymbol{\phi}})\boldsymbol{g}'(\tilde{\boldsymbol{\phi}})\\ &\quad= \boldsymbol{g}'^\top(\tilde{\boldsymbol{\phi}})(\boldsymbol{J}^\top)^{-1}(\tilde{\boldsymbol{\phi}})\boldsymbol{J}^\top(\tilde{\boldsymbol{\phi}})(\boldsymbol{I}')^{-1}(\tilde{\boldsymbol{\phi}})\boldsymbol{J}(\tilde{\boldsymbol{\phi}})\boldsymbol{J}^{-1}(\tilde{\boldsymbol{\phi}})\boldsymbol{g}'(\tilde{\boldsymbol{\phi}})\\ &\quad= \boldsymbol{g}'^\top(\tilde{\boldsymbol{\phi}})(\boldsymbol{I}')^{-1}(\tilde{\boldsymbol{\phi}})\boldsymbol{g}'(\tilde{\boldsymbol{\phi}}).\end{aligned}$$

Since the last line here is the efficient score form of the LM statistic in the reparametrized model, we have proved that this form of the LM statistic is invariant to reparametrization.