Supplement to "A Convenient Method for the Estimation of the Multinomial Logit Model with Fixed Effects"

Xavier D'Haultfœuille* Alessandro Iaria†

December 22, 2015

Proof of Proposition 1

$\widehat{\boldsymbol{\beta}}_{\text{sub}}$ is a conditional maximum likelihood estimator, hence by, e.g., Andersen (1970), it is asymptotically normal with asymptotic variance V_{as} equal to

$$V_{\text{as}} = E\left\{V\left[\left.\frac{\partial \ln\left(\Pr\left(\boldsymbol{Y}\,|\,D,\boldsymbol{X}\right)\right)}{\partial\boldsymbol{\beta}}\right| D,\boldsymbol{X}\right]\right\}^{-1}.$$

We now show that $V_{\text{as}} = \mathcal{I}(\widehat{\boldsymbol{\beta}}_{\text{sub}})^{-1}$. First,

$$\ln\left(\Pr\left(\boldsymbol{Y}\,|\,D,\boldsymbol{X}\right)\right) = \boldsymbol{Z}(\boldsymbol{Y})'\boldsymbol{\beta} - \ln\left(\sum_{\boldsymbol{k}\in D}\exp\left(\boldsymbol{Z}(\boldsymbol{k})'\boldsymbol{\beta}\right)\right).$$

Consequently,

$$\frac{\partial \ln\left(\Pr\left(\boldsymbol{Y}\,|\,D,X\right)\right)}{\partial\boldsymbol{\beta}} = \sum_{\boldsymbol{j}\in D}\boldsymbol{Z}(\boldsymbol{j})'\mathbb{1}\left\{\boldsymbol{Y}=\boldsymbol{j}\right\} - \frac{\sum_{\boldsymbol{k}\in D}\exp\left(\boldsymbol{Z}(\boldsymbol{k})'\boldsymbol{\beta}\right)\boldsymbol{Z}(\boldsymbol{k})}{\sum_{\boldsymbol{k}\in D}\exp\left(\boldsymbol{Z}(\boldsymbol{k})'\boldsymbol{\beta}\right)}.$$

As a result,

$$V\left(\left.\frac{\partial \ln\left(\Pr\left(\boldsymbol{Y}\,|\,D,\boldsymbol{X}\right)\right)}{\partial\boldsymbol{\beta}}\right| D\right) = \sum_{(\boldsymbol{j},\boldsymbol{j}')\in D^2}\boldsymbol{Z}(\boldsymbol{j})\boldsymbol{Z}(\boldsymbol{j})'\text{Cov}\left(\mathbb{1}\left\{\boldsymbol{Y}=\boldsymbol{j}\right\},\mathbb{1}\left\{\boldsymbol{Y}=\boldsymbol{j}'\right\}\middle|\, D,\boldsymbol{X}\right).$$

*CREST. Email address: xavier.dhaultfoeuille@ensae.fr
†CREST. Email address: alessandro.iaria@ensae.fr

Note that

$$\mathrm{Cov}\left(\mathbb{1}\left\{\boldsymbol{Y}=\boldsymbol{j}\right\},\mathbb{1}\left\{\boldsymbol{Y}=\boldsymbol{j}'\right\}\middle| D,\boldsymbol{X}\right)=\begin{cases}\Pr(\boldsymbol{Y}=\boldsymbol{j}|D,\boldsymbol{X})\left(1-\Pr(\boldsymbol{Y}=\boldsymbol{j}|D,\boldsymbol{X})\right) & \text{if } \boldsymbol{j}=\boldsymbol{j}'\\ -\Pr(\boldsymbol{Y}=\boldsymbol{j}|D,\boldsymbol{X})\Pr(\boldsymbol{Y}=\boldsymbol{j}'|D,\boldsymbol{X}) & \text{if } \boldsymbol{j}\neq\boldsymbol{j}'\end{cases}.$$

After some algebra, we obtain

$$V\left(\left.\frac{\partial\ln\left(\Pr\left(\boldsymbol{Y}|\,D,\boldsymbol{X}\right)\right)}{\partial\boldsymbol{\beta}}\right| D,\boldsymbol{X}\right)=\sum_{\boldsymbol{j}\in D}P_{\boldsymbol{j}D}\boldsymbol{Z}(\boldsymbol{j})\left(\boldsymbol{Z}(\boldsymbol{j})-\overline{\boldsymbol{Z}}_D\right)'.$$

Finally, by the law of iterated expectations,

$$V_{\text{as}}^{-1}=\sum_{\boldsymbol{j}\in\mathcal{P}}E\left[P_{\boldsymbol{j}}\boldsymbol{Z}(\boldsymbol{j})\boldsymbol{Z}(\boldsymbol{j})'\right]-E\left[\boldsymbol{Z}(\boldsymbol{j})E\left[P_{\boldsymbol{j}D}\overline{\boldsymbol{Z}}'_D|\boldsymbol{X}\right]\right].$$

To prove the second point, we use the expression

$$V_{\text{as}}^{-1}=E\left[\sum_{\boldsymbol{j}\in\mathcal{P}}P_{\boldsymbol{j}}\boldsymbol{Z}(\boldsymbol{j})\boldsymbol{Z}(\boldsymbol{j})'-\sum_{(\boldsymbol{j},\boldsymbol{k})\in\mathcal{P}^2}\boldsymbol{Z}(\boldsymbol{j})\boldsymbol{Z}(\boldsymbol{k})'P_{\boldsymbol{j}D}P_{\boldsymbol{k}D}\right] \tag{1}$$

and the fact that $\widehat{\boldsymbol{\beta}}_{CMLE}$ is a particular case of $\widehat{\boldsymbol{\beta}}_{\text{sub}}$, with $D=\mathcal{P}\left(\sum_{t=1}^T\boldsymbol{Y}_t\right)$. This implies that its asymptotic variance also satisfies (1) with this D. Then, it suffices to prove that

$$\sum_{(\boldsymbol{j},\boldsymbol{k})\in\mathcal{P}^2}\boldsymbol{Z}(\boldsymbol{j})\boldsymbol{Z}(\boldsymbol{k})'E\left[P_{\boldsymbol{j}D}P_{\boldsymbol{k}D}\middle|\boldsymbol{S},\boldsymbol{X}\right]>>\sum_{(\boldsymbol{j},\boldsymbol{k})\in\mathcal{P}^2}\boldsymbol{Z}(\boldsymbol{j})\boldsymbol{Z}(\boldsymbol{k})'P_{\boldsymbol{j}}^{\boldsymbol{S}}P_{\boldsymbol{k}}^{\boldsymbol{S}},$$

where $>>$ denotes the usual order on symmetric matrices, $\boldsymbol{S}=\sum_{t=1}^T\boldsymbol{Y}_t$ and $P_{\boldsymbol{j}}^{\boldsymbol{S}}=\Pr(\boldsymbol{Y}=\boldsymbol{j}|\boldsymbol{S},\boldsymbol{X})$. Equivalently, we must prove that matrix $\boldsymbol{M}=\sum_{(\boldsymbol{j},\boldsymbol{k})\in\mathcal{P}^2}\boldsymbol{Z}(\boldsymbol{j})\boldsymbol{Z}(\boldsymbol{k})'C_{\boldsymbol{jk}}$, with $C_{\boldsymbol{jk}}=\mathrm{cov}(P_{\boldsymbol{j}D},P_{\boldsymbol{k}D}|\boldsymbol{S},\boldsymbol{X})$, is symmetric positive. Consider a vector $\boldsymbol{\lambda}$ and an ordering $\boldsymbol{Z}^1,...,\boldsymbol{Z}^{2^{JT}}$ of the $(\boldsymbol{Z}(\boldsymbol{j}))_{\boldsymbol{j}\in\mathcal{P}}$. Let $\boldsymbol{a}=\left(\boldsymbol{\lambda}'\boldsymbol{Z}^1,...,\boldsymbol{\lambda}'\boldsymbol{Z}^{2^{JT}}\right)$. Some algebra shows that $\boldsymbol{\lambda}'\boldsymbol{M}\boldsymbol{\lambda}=\boldsymbol{a}'\boldsymbol{C}\boldsymbol{a}$, where $\boldsymbol{C}$ is the covariance matrix with typical term $C_{\boldsymbol{ij}}$. The result follows by positivity of $\boldsymbol{C}$.