**APPENDIX.** Derivatives of the simulated log-likelihood for the (generalized) true random effects model

---

To estimate the parameters of the TRE model, Greene (2005a,b) proposed adopting a procedure called 'maximum simulated likelihood'. It is the same as maximum likelihood except that, in short, simulated probabilities are used in lieu of the exact probabilities (Train, 2009).

In the case of the TRE model, conditional joint density function for the $i$-th observation, given the random individual-specific effect $w_i$, is

$$f(y_{i1}, \dots , y_{iT}|w_i) = \prod_{t=1}^{T} f(y_{it}|w_i) = \prod_{t=1}^{T} \frac{2}{\sigma}\phi\left(-\frac{\varepsilon_{it}}{\sigma}\right)\Phi\left(-\frac{\varepsilon_{it}}{\sigma}\lambda\right),$$

where $\varepsilon_{it} = y_{it} - x_{it}\beta - \sigma_w W_i$, $W_i = w_i/\sigma_w$ and $\phi(\Phi)$ is a standard normal probability (cumulative) density function.

Unconditional joint density function is written as:

$$f(y_{i1}, \dots , y_{iT}) = \int_{w_i} \prod_{t=1}^{T} f(y_{i1}, \dots, y_{iT}|w_i) f(w_i) dw_i$$

The log-likelihood function for the sample of size $N$ is:

$$L(\theta) = \prod_{i=1}^{N} f(y_{i1}, \dots, y_{iT}),$$

where $\theta = [\beta' \; \lambda \; \sigma \; \sigma_w]'$ is a vector of parameters.

It is not possible to directly maximise $L(\theta)$ and thus estimate $\theta$ because of the hidden variables $W_i$ ($i$=1,…, $N$). Let $L_S(\theta)$ be a simulated approximation to $L(\theta)$ in this sense that it is approximated through simulation. For this purpose, $L(\theta)$ is simulated by taking draws from $f(w_i)$, calculating $f(y_{i1}, \dots, y_{iT}|w_i)$ for each draw $W_{i,r}$ ($r$ = 1,…,$R$) and averaging the results (Train, 2009). In our application, estimation of the model was based on 500 Halton draws.

The simulated log-likelihood for the sample is $lnL_S = \sum_{i=1}^{N} \ln\left[\frac{1}{R}\sum_{r=1}^{R} P_{ir}(\theta)\right]$,

where $P_{i,r}(\theta) = \prod_{t=1}^{T} P_{it,r}(\theta)$, $P_{it,r}(\theta) = f\left(y_{it}\middle|w_{i,r}\right)$.

The value of the parameters $\theta$ that maximises $L_S(\theta)$ is called the maximum simulated likelihood estimator. In the numerical application, the derivatives of the simulated log-likelihood are calculated as follows (Greene, 2012, chapter 15):

$$\frac{\partial lnL_S(\theta)}{\partial\theta} = \sum_{i=1}^{N}\sum_{r=1}^{R} Q_{i,r}(\theta) g_{i,r}(\theta) = \sum_{i=1}^{N} \tilde{g}_i(\theta),$$

where

$$Q_{i,r}(\theta) = \frac{P_{i,r}(\theta)}{\sum_{r=1}^{R} P_{i,r}(\theta)},$$

$$g_{i,r}(\theta) = \sum_{t=1}^{T} \frac{\partial lnP_{it,r}(\theta)}{\partial\theta}.$$

Greene (2012, chapter 15) obtained the analytic expression of the second derivatives of the simulated log-likelihood function with respect to all parameters. The Hessian matrix takes the following form:

$$\frac{\partial^2 lnL_S}{\partial\theta\,\partial\theta'} = \sum_{i=1}^{N}\left\{\sum_{r=1}^{R} Q_{i,r}(\theta)\left[\frac{\partial g_{i,r}(\theta)}{\partial\theta'} + d_{i,r}(\theta)d_{i,r}(\theta)'\right]\right\},$$

where $d_{i,r}(\theta) = g_{i,r}(\theta) - \tilde{g}_i(\theta)$.

The difficulty in obtaining these derivatives of the simulated log-likelihood comes down to the calculation of derivatives for the auxiliary function $lnP_{it,r}(\theta)$. In particular, the latter are given by:

$$\frac{\partial lnP_{it,r}}{\partial\beta} = \left(\frac{\varepsilon_{it,r}}{\sigma^2} + M_{it,r}\frac{\lambda}{\sigma}\right)x'_{it},$$

$$\frac{\partial lnP_{it,r}}{\partial\sigma} = -\frac{1}{\sigma} + \frac{\left(\varepsilon_{it,r}\right)^2}{\sigma^3} + M_{it,r}\cdot\frac{\varepsilon_{it,r}}{\sigma^2}\lambda = \frac{1}{\sigma}\left(\left(\frac{\varepsilon_{it,r}}{\sigma}\right)^2 - M_{it,r}A_{it,r} - 1\right),$$

$$\frac{\partial lnP_{it,r}}{\partial\lambda} = -M_{it,r}\frac{\varepsilon_{it,r}}{\sigma},$$

$$\frac{\partial lnP_{it,r}}{\partial\sigma_w} = \left(\frac{\varepsilon_{it,r}}{\sigma^2} + M_{it,r}\frac{\lambda}{\sigma}\right)W_{i,r},$$

$$\varepsilon_{it,r} = y_{it} - x_{it}\beta - \sigma_w W_{i,r},$$

$$A_{it,r} = -\frac{\varepsilon_{it,r}}{\sigma}\lambda,$$

with the inverse Mills ratio denoted and defined by:

$$M_{it,r} = \frac{\phi\left(A_{it,r}\right)}{\Phi\left(A_{it,r}\right)}.$$

We also take the second order partial derivatives with respect to all pairs of parameters:

$$\frac{\partial^2 lnP_{it,r}}{\partial\beta\,\partial\beta'} = -\frac{1}{\sigma^2}\left(1 - \lambda^2 D_{it,r}\right)x'_{it}x_{it}$$

$$\frac{\partial^2 lnP_{it,r}}{\partial\sigma\,\partial\sigma} = -\frac{1}{\sigma^2}\left(3\left(\frac{\varepsilon_{it,r}}{\sigma}\right)^2 - A_{it,r}\left(A_{it,r}D_{it,r} + 2M_{it,r}\right) - 1\right)$$

$$\frac{\partial^2 lnP_{it,r}}{\partial\lambda^2} = -\left(\frac{\varepsilon_{it,r}}{\sigma}\right)^2 M_{it,r}\left(A_{it,r} + M_{it,r}\right) = -\left(\frac{\varepsilon_{it,r}}{\sigma}\right)^2 \left(-D_{it,r}\right)$$

$$\frac{\partial^2 lnP_{it,r}}{\partial(\sigma_w)^2} = -\left(\frac{W_{i,r}}{\sigma}\right)^2 \left(1 - \lambda^2 D_{it,r}\right)$$

$$\frac{\partial^2 lnP_{it,r}}{\partial\beta\partial\sigma} = \frac{1}{\sigma^2}\left(-2\frac{\varepsilon_{it,r}}{\sigma} - \lambda\left(M_{it,r} + A_{it,r}D_{it,r}\right)\right)x'_{it}$$

$$\frac{\partial^2 lnP_{it,r}}{\partial\beta\partial\lambda} = \frac{1}{\sigma}\left(M_{it,r} + A_{it,r}D_{it,r}\right)x'_{it}$$

$$\frac{\partial^2 lnP_{it,r}}{\partial\beta\partial\sigma_w} = \frac{W_{i,r}}{\sigma^2}\left(\lambda^2 D_{it,r} - 1\right)x'_{it}$$

$$\frac{\partial^2 lnP_{it,r}}{\partial\sigma\partial\lambda} = \frac{\varepsilon_{it,r}}{\sigma^2}\left(M_{it,r} + A_{it,r}D_{it,r}\right)$$

$$\frac{\partial^2 lnP_{it,r}}{\partial\sigma\partial\sigma_w} = \frac{W_{i,r}}{\sigma^2}\left(-2\frac{\varepsilon_{it,r}}{\sigma} - \lambda\left(M_{it,r} + A_{it,r}D_{it,r}\right)\right)$$

$$\frac{\partial^2 lnP_{it,r}}{\partial\lambda\partial\sigma_w} = \frac{W_{i,r}}{\sigma}\left(M_{it,r} + A_{it,r}D_{it,r}\right),$$

where $D_{it,r}$ is the derivative of the inverse Mills ratio with respect to its argument.

In the GTRE model, the composed error term is defined as $\varepsilon_{it,r} = y_{it} - x_{it}\beta - \sigma_w W_{i,r} + \sigma_h\left|H_{i,r}\right|$ and the vector of parameters $\theta$ contains one additional parameter $\sigma_h$. It should be noted that the random variable $H_{i,r}$ can be treated in the same way as the standardised individual-specific effect $W_{i,r}$, except that there is a plus sign before the first one. Consequently, for the GTRE model, the first and the second derivatives of $lnP_{it,r}(\theta)$ with respect to $\sigma_h$ are analogous to those for the TRE model when we calculate the partial derivatives with respect to $\sigma_w$. Obviously, they have the opposite sign, except for their own second derivatives.