Simultaneous Bayesian Modelling of a Panel of Parametric Income Distributions for Grouped Data

Bastian Gribisch & Ilya Zarubin

submitted by:
Ilya Zarubin

zarubin@wiso.uni-koeln.de

1 Model I

Assumptions

- for a fixed cross sectional unit i at some time point t:
  $\bar{\boldsymbol{n}}_{it} = \left(n_{it}^{(1)}, \ldots, n_{it}^{(M_{it})}\right)$: number of observations in the M_{it} income groups

- we consider a panel of observations where the number of income groups may vary over both indices but typically changes only with i i.e. $M_{it} \equiv M_i$

Assumption 1: let each $\bar{\boldsymbol{n}}_{it} \overset{ind.}{\sim} \mathrm{MNL}_{it}\left(\pi_{it}^{(1)}, \ldots, \pi_{it}^{(M_{it})}\right)$ conditional on regressors x_{it}
hence the joint likelihood function is

$$L\left(\{\theta_{it}\}; \{\bar{\boldsymbol{n}}_{it}\}\right) = \prod_{i=1}^{N} \prod_{t=1}^{T} \mathrm{MNL}_{it}(\pi_{it}^{(1)}, \ldots, \pi_{it}^{(M_{it})}) \ ,$$

where each $\pi_{it}^{(k)} = \left(F_y(z_{it}^{(k)}; \theta_{it}) - F_y(z_{it}^{(k-1)}; \theta_{it})\right)$ and some parametric income distribution $F_y(y; \theta)$ with P-variate parameter vector $\theta = (\theta^{(1)}, \ldots, \theta^{(P)})$.

- key point of distributional regression: link each parameter component $\theta_{it}^{(p)}$ to a *structured additive predictor* $\eta_{it}^{(p)}$ via a response function $h_{(p)}$

$$\theta_{it}^{(p)} = h_p(\eta_{it}^{(p)}) \ ,$$

where $h^{(p)}$'s range equals $\theta_{it}^{(p)}$'s possible parameter values e.g. $\exp(\cdot)$ for $\theta_{it}^{(p)} > 0$
The predictor specification allows each $\theta_{it}^{(p)}$ to depend on J_p different effects e.g. linear or random effects denoted as $x_{(it)}$'s or u_i, respectively

$$\eta_{it}^{(p)} = x'_{(it,1)}\beta_1^{(p)} + \ldots + x'_{(it,J_p)}\beta_{J_p}^{(p)} + u_i^{(p)} \ ,$$

- Number of parameters: $J_1 \times J_2 \times \ldots \times J_P$ different (multivariate) β's and $N \times P$ different random effects u_i.

Assumption 2: To specify an income distribution function we assume a rather general four-parameter GB2:

$$F(y; a, b, p, q) = B(d; p, q) = \frac{\int_0^d t^{p-1}(1-t)^{q-1} dt}{\mathrm{B}(p, q)} \ , \ d = \frac{(y/b)^a}{1 + (y/b)^a} \ ,$$

where each parameter is strictly positive $a, b, p, q > 0$

- Hence, in a panel setting the parameters are $a_{it} = \theta_{it}^{(1)}$, $b_{it} = \theta_{it}^{(2)}$, $p_{it} = \theta_{it}^{(3)}$, $q_{it} = \theta_{it}^{(4)}$; , and all are equipped with strictly positive response function $\forall i, t$:

$$
\begin{aligned}
a_{it} &= h\left(\eta_{it}^{(a)}\right) = \exp\left(\eta_{it}^{(a)}\right) \ , \quad b_{it} = h\left(\eta_{it}^{(b)}\right) = \exp\left(\eta_{it}^{(b)}\right) \ , \\
p_{it} &= h\left(\eta_{it}^{(p)}\right) = \exp\left(\eta_{it}^{(p)}\right) \ , \quad q_{it} = h\left(\eta_{it}^{(q)}\right) = \exp\left(\eta_{it}^{(q)}\right) \ ,
\end{aligned}
$$

and corresponding additive predictors

$$
\begin{aligned}
\eta_{it}^{(a)} &= \left(x_{it}^{(a)}\right)' \beta^{(a)} + u_i^{(a)} \ , \quad \eta_{it}^{(b)} = \left(x_{it}^{(b)}\right)' \beta^{(b)} + u_i^{(b)} \ , \\
\eta_{it}^{(p)} &= \left(x_{it}^{(p)}\right)' \beta^{(p)} + u_i^{(p)} \ , \quad \eta_{it}^{(q)} = \left(x_{it}^{(q)}\right)' \beta^{(q)} + u_i^{(q)} \ ,
\end{aligned}
$$

where the β's may be multivariate and u_i's are random effects.

Bayesian model formulation:

We follow Klein *et al.* (2013) and consider a Bayesian setting:

$$
p\left(\boldsymbol{\beta}, \boldsymbol{U}, \boldsymbol{\sigma}, \boldsymbol{\tau} | \bar{\boldsymbol{N}}, \boldsymbol{X}\right) \propto p\left(\bar{\boldsymbol{N}} | \boldsymbol{\beta}, \boldsymbol{U}, \boldsymbol{X}\right) p\left(\boldsymbol{\beta} | \boldsymbol{\tau}\right) p\left(\boldsymbol{\tau}\right) p\left(\boldsymbol{U} | \boldsymbol{\sigma}\right) p(\boldsymbol{\sigma}) \ ,
$$

where the components are

- deterministic regressors $\boldsymbol{X} = (\boldsymbol{x}_{it})_{i,t}^{N,T} = \left(x_{it}^{(a)}, x_{it}^{(b)}, x_{it}^{(p)}, x_{it}^{(q)}\right)_{i,t}^{N,T}$
- the (multivariate) panel data $\bar{\boldsymbol{N}} = (\bar{\boldsymbol{n}}_{it})_{i,t}^{N,T} = \left(n_{it}^{(1)}, \ldots, n_{it}^{(M_{it})}\right)_{i,t}^{N,T}$
- the (multivariate) parameters $\boldsymbol{\beta} = \left(\boldsymbol{\beta}_a, \boldsymbol{\beta}_b, \boldsymbol{\beta}_p, \boldsymbol{\beta}_q\right)$
- beta prior variances $\boldsymbol{\tau} = (\tau_a, \tau_b, \tau_p, \tau_q)$
- and random effects $\boldsymbol{U} = (\boldsymbol{U}_i)_{i=1}^{N} = \left(u_i^{(a)}, u_i^{(b)}, u_i^{(p)}, u_i^{(q)}\right)_{i=1}^{N}$
- random effect prior variances $\boldsymbol{\sigma} = (\sigma_a, \sigma_b, \sigma_p, \sigma_q)$

and where the corresponding prior and likelihood specifications are

- $p\left(\boldsymbol{\beta}|\tau\right) = p\left(\boldsymbol{\beta_a}|\tau_a\right) p\left(\boldsymbol{\beta_b}|\tau_b\right) p\left(\boldsymbol{\beta_p}|\tau_p\right) p\left(\boldsymbol{\beta_q}|\tau_q\right)$, with each $\boldsymbol{\beta}|\tau \sim N(0, \tau^2)$
- $p\left(\boldsymbol{U}\right) = \prod_{i=1}^{N} p(\boldsymbol{U}_i|\boldsymbol{\sigma})$ and $p(\boldsymbol{U}_i|\boldsymbol{\sigma}) = p\left(u_i^{(a)}|\sigma_a\right) p\left(u_i^{(b)}|\sigma_b\right) p\left(u_i^{(p)}|\sigma_p\right) p\left(u_i^{(q)}|\sigma_q\right)$, with each $u_i|\sigma \sim N(0, \sigma^2)$
- $p\left(\bar{\boldsymbol{N}}|\boldsymbol{\beta}, \boldsymbol{U}, \boldsymbol{X}\right) = \prod_{i=1}^{N} \prod_{t=1}^{T} \mathrm{MNL}_{it}\left(\pi_{it}^{(1)}, \ldots, \pi_{it}^{(M_{it})}\right)$ is the joint multinom. likelihood, $\pi_{it}^{(k)} = \left(F_{\mathrm{GB2}}(z_{it}^{(k)}; a_{it}, b_{it}, p_{it}, q_{it}) - F_{\mathrm{GB2}}(z_{it}^{(k-1)}; a_{it}, b_{it}, p_{it}, q_{it})\right)$

2 Estimation

2.1 MAP estimation (maximization w.r.t. posterior kernel)

2.1.1 Including maximization over RE

$$
\begin{aligned}
\Theta_{\text{MAP}} &= \underset{\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{\sigma},\boldsymbol{\tau}}{\arg\max}\; p\left(\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{\sigma},\boldsymbol{\tau}|\bar{\boldsymbol{N}},\boldsymbol{X}\right)\\
&= \underset{\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{\sigma},\boldsymbol{\tau}}{\arg\max}\; p\left(\bar{\boldsymbol{N}}|\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{X}\right)\times p\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right)\times p(\boldsymbol{\tau})\times p\left(\boldsymbol{U}|\boldsymbol{\sigma}\right)\times p(\boldsymbol{\sigma})\\
&= \underset{\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{\sigma},\boldsymbol{\tau}}{\arg\max}\; \prod_{i=1}^{N}\prod_{t=1}^{T}\text{MNL}_{it}\left(\pi_{it}^{(1)},\ldots,\pi_{it}^{(M_{it})}\right)\\
&\times p\left(\boldsymbol{\beta_a}|\tau_a\right)p\left(\boldsymbol{\beta_b}|\tau_b\right)p\left(\boldsymbol{\beta_p}|\tau_p\right)p\left(\boldsymbol{\beta_q}|\tau_q\right)\times p(\tau_a)p(\tau_b)p(\tau_p)p(\tau_q)\\
&\times\prod_{i=1}^{N}p\left(u_i^{(a)}|\sigma_a\right)p\left(u_i^{(b)}|\sigma_b\right)p\left(u_i^{(p)}|\sigma_p\right)p\left(u_i^{(q)}|\sigma_q\right)\times p(\sigma_a)p(\sigma_b)p(\sigma_p)p(\sigma_q)\;.
\end{aligned}
$$

2.1.2 Without RE

Integrating out all random effects $u_i^{(a)}, u_i^{(b)}, u_i^{(p)}, u_i^{(q)}$ leads

$$
\int_{\mathcal{U}} p\left(\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{\sigma},\boldsymbol{\tau}|\bar{\boldsymbol{N}},\boldsymbol{X}\right)d\boldsymbol{U} = \int_{\mathcal{U}} p\left(\bar{\boldsymbol{N}}|\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{X}\right)p\left(\boldsymbol{U}|\boldsymbol{\sigma}\right)d\boldsymbol{U}\;\;\times p(\boldsymbol{\sigma})p\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right)p(\boldsymbol{\tau})p(\boldsymbol{\sigma})\;,
$$

with product space $\mathcal{U} = \times_{i=1}^{N}\mathcal{U}_i^{\mathcal{A}\times\mathcal{B}\times\mathcal{P}\times\mathcal{Q}}$ and $\boldsymbol{U}_i = \left(u_i^{(1)}, u_i^{(2)}, u_i^{(3)}, u_i^{(4)}\right)$ living in $\mathcal{U}_i^{\mathcal{A}\times\mathcal{B}\times\mathcal{P}\times\mathcal{Q}}$. Hence, the MAP estimator for the structural parameters $\Theta_{\text{MAP}}^{\mathbf{S}}$ only is

$$
\begin{aligned}
\Theta_{\text{MAP}}^{\mathbf{S}} &= \underset{\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau}}{\arg\max}\; p\left(\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau}|\bar{\boldsymbol{N}},\boldsymbol{X}\right)\\
&= \underset{\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau}}{\arg\max}\; \int_{\mathcal{U}} p\left(\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{\sigma},\boldsymbol{\tau}|\bar{\boldsymbol{N}},\boldsymbol{X}\right)d\boldsymbol{U}\\
&= \underset{\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau}}{\arg\max}\; \int_{\mathcal{U}} p\left(\bar{\boldsymbol{N}}|\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{X}\right)p\left(\boldsymbol{U}|\boldsymbol{\sigma}\right)d\boldsymbol{U}\times p(\boldsymbol{\sigma})p\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right)p(\boldsymbol{\tau})\\
&= \underset{\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau}}{\arg\max}\; \prod_{i=1}^{N}\int_{\mathcal{U}_i^{\mathcal{A}\times\mathcal{B}\times\mathcal{P}\times\mathcal{Q}}}\prod_{t=1}^{T}\text{MNL}_{it}\left(\pi_{it}^{(1)},\ldots,\pi_{it}^{(M_{it})}\right)p\left(\boldsymbol{U}_i|\boldsymbol{\sigma}\right)d\boldsymbol{U}_i\\
&\times p(\boldsymbol{\sigma})p\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right)p(\boldsymbol{\tau})\;,
\end{aligned}
$$

The integral to be approximated via quadrature (i.e. to be evaluated numerically) is

$$
\begin{aligned}
\mathcal{I}_{RE} &= \int_{\mathcal{U}_i^{\mathcal{A}\times\mathcal{B}\times\mathcal{P}\times\mathcal{Q}}}\prod_{t=1}^{T}\text{MNL}_{it}\left(\pi_{it}^{(1)},\ldots,\pi_{it}^{(M_{it})}\right)p\left(\boldsymbol{U}_i|\boldsymbol{\sigma}\right)d\boldsymbol{U}_i\;,\\
&= \int_{\mathcal{U}_i^{\mathcal{A}\times\mathcal{B}\times\mathcal{P}\times\mathcal{Q}}}\prod_{t=1}^{T}\text{MNL}_{it}\left(\pi_{it}^{(1)},\ldots,\pi_{it}^{(M_{it})}\right)p\left(u_i^{(a)}|\sigma_a\right)p\left(u_i^{(b)}|\sigma_b\right)p\left(u_i^{(p)}|\sigma_p\right)p\left(u_i^{(q)}|\sigma_q\right)d\boldsymbol{U}_i\;.
\end{aligned}
$$

2.1.3 Deriving Gauss-Hermite quadrature approximation to the RE integral

From the previous section, the integral to be evaluated numerically has the form

$$\mathcal{I}_{RE} = \int_{\mathcal{U}_i^{\mathcal{A}\times\mathcal{B}\times\mathcal{P}\times\mathcal{Q}}} \prod_{t=1}^{T} \mathrm{MNL}_{it}\left(\pi_{it}^{(1)},\ldots,\pi_{it}^{(M_{it})}\right) p\left(u_i^{(a)}|\sigma_a\right) p\left(u_i^{(b)}|\sigma_b\right) p\left(u_i^{(p)}|\sigma_p\right) p\left(u_i^{(q)}|\sigma_q\right) d\boldsymbol{U}_i \ .$$

For simplicity, we first derive the Gauss-hermite approximation for the one-dimensional case and then transfer the result to the 4-dim. case above. The one dimensional case where integration is over one of the four REs requires to evaluate

$$\begin{aligned}\mathcal{I}_{RE-1dim} &= \int_{\mathcal{U}_i} \prod_{t=1}^{T} \mathrm{MNL}_{it}\left(\pi_{it}^{(1)},\ldots,\pi_{it}^{(M_{it})}\right) p\left(u_i^{(q)}|\sigma\right) du_i \\ &= \int_{\mathcal{U}_i} h(u_i)\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(u_i-\mu)^2}{2\sigma^2}\right\} du_i \ , \end{aligned} \tag{1}$$

where $\prod_{t=1}^{T} \mathrm{MNL}_{it}\left(\pi_{it}^{(1)},\ldots,\pi_{it}^{(M_{it})}\right) \equiv h(u_i)$ i.e. (for the moment) we view the product of multinomial likelihoods as a function $h(\cdot)$ of one particular RE u_i holding the other REs fixed. Note that $h(\cdot)$ is a highly nonlinear function of u_i (as well as the other REs). It takes the form

$$\begin{aligned} h(u_i) &= \exp\left\{\log\left\{\prod_{t=1}^{T} \mathrm{MNL}_{it}\left(\pi_{it}^{(1)},\ldots,\pi_{it}^{(M_{it})}\right)\right\}\right\} \\ &= \exp\left\{\sum_{t=1}^{T}\left[\log(\bar{n}_{it}) - \sum_{k=1}^{M_{it}} \log(n_{it}^{(k)}!) + \sum_{k=1}^{M_{it}} n_{it}^{(k)} \log(\pi_{it}^{(k)})\right]\right\} \ , \end{aligned} \tag{2}$$

with $\log(\pi_{it}^{(k)}) = \log\left(F_{\mathrm{GB2}}(z_{it}^{(k)};a_{it},b_{it},p_{it},q_{it}) - F_{\mathrm{GB2}}(z_{it}^{(k-1)};a_{it},b_{it},p_{it},q_{it})\right)$ and thus

$$F_{\mathrm{GB2}}(z_{it}^{(k)};a_{it},b_{it},p_{it},q_{it}) = B(d_{it}^{(k)};p_{it},q_{it}) = \frac{\int_0^{d_{it}^{(k)}} t^{p-1}(1-t)^{q_{it}-1}dt}{\mathrm{B}(p_{it},q_{it})} \ , \ d_{it}^{(k)} = \frac{(z_{it}^{(k)}/b_{it})^{a_{it}}}{1+(z_{it}^{(k)}/b_{it})^{a_{it}}} \ .$$

Gauss hermite quadrature utilizes the following approximation:

$$\int_{-\infty}^{+\infty} f(\nu)d\nu = \int_{-\infty}^{+\infty} w(\nu)g(\nu)d\nu = \int_{-\infty}^{+\infty} \exp\{-\nu^2\}g(\nu)d\nu \approx \sum_{m=1}^{M} w_m g(\nu_m) \ ,$$

where $m = 1,\ldots,M$ is an index for the M different "knots" ν_m each attached with a corresponding weight w_m. The latter are computed from the weight function $w(u) = \exp\{-u^2\}$ evaluated at the corresponding knots. The approximation of the integral as a sum is exact whenever g is at most a polynomial of order $2M-1$. If g is close to a polynomial the approximation still tends to be reasonable since the approximation error diminishes with an increasing number of knots. However, the number of knots has to increase with the dimension of the integral to be approximated, typically at exponential rate. A rule of thumb is to use a total of M^D knots where D is the dimension of the integral and M^1 is the number of knots used for a sufficiently accurate approximation for a one dimensional integral.

Unfortunately, the integrand in (1) does not exactly resemble the integrand $\int_{-\infty}^{+\infty} \exp\{-\nu^2\} g(\nu) d\nu$ required for Hermite-quadrature. Hence, we use integration by substitution to obtain the necessary integrand form. In particular, let $\nu = \frac{u_i - \mu}{\sqrt{2}\sigma} \equiv \phi(u_i)$ implying $u_i = \sqrt{2}\sigma\nu + \mu$, $\frac{du_i}{d\nu} = \sqrt{2}\sigma$. Integration by substitution leads

$$
\begin{aligned}
\mathcal{I}_{RE-1dim} &= \int_{u_i=-\infty}^{u_i=+\infty} \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{2}\sigma} \exp\left\{ -\frac{(u_i-\mu)^2}{2\sigma^2} \right\} h(u_i) du_i \\
&= \frac{1}{\sqrt{\pi}} \int_{\nu=\phi(-\infty)=-\infty}^{\nu=\phi(+\infty)=+\infty} \underbrace{\exp\left\{-\nu^2\right\}}_{w(\nu)} \underbrace{h(\sqrt{2}\sigma\nu + \mu)}_{g(\nu)} d\nu \\
&\approx \frac{1}{\sqrt{\pi}} \sum_{m=1}^{M} w_m h(\sqrt{2}\sigma\nu_m + \mu) \ .
\end{aligned}
$$

As discussed earlier, the above approximation crucially depends on two conditions: Firstly, the number of knots M employed which can be controlled by the researcher and set according to her computational budget. (although the total number of knots required scales exponentially with the dimension of the integrand, as discussed previously, we are interested only in integrals up to a dimension of four) Secondly, on how well $h(\sqrt{2}\sigma\nu + \mu) \equiv g(\nu)$, as a function of ν, is approximated by some polynomial ν of order no larger than $2M - 1$.

It remains to transfer the approximation from the univariate integration problem to the 4-dimensional one over all 4 different random effects attached to $a, b, p-$ and $q-$parameters which we considered at the beginning of this section.

$$
\begin{aligned}
\mathcal{I}_{RE} = &\int_{\mathcal{U}_i^{\mathcal{A}\times\mathcal{B}\times\mathcal{P}\times\mathcal{Q}}} \prod_{t=1}^{T} \mathrm{MNL}_{it}\left(\pi_{it}^{(1)}, \ldots, \pi_{it}^{(M_{it})}\right) p\left(u_i^{(a)}|\sigma_a\right) p\left(u_i^{(b)}|\sigma_b\right) p\left(u_i^{(p)}|\sigma_p\right) p\left(u_i^{(q)}|\sigma_q\right) d\boldsymbol{U}_i \ , \\
&= \int_{\mathcal{U}_i^{\mathcal{A}}} \int_{\mathcal{U}_i^{\mathcal{B}}} \int_{\mathcal{U}_i^{\mathcal{P}}} \int_{\mathcal{U}_i^{\mathcal{Q}}} \prod_{t=1}^{T} h(u_i^{(a)}, u_i^{(b)}, u_i^{(b)}, u_i^{(q)}) p\left(u_i^{(a)}|\sigma_a\right) p\left(u_i^{(b)}|\sigma_b\right) p\left(u_i^{(p)}|\sigma_p\right) p\left(u_i^{(q)}|\sigma_q\right) du_i^{(a)} du_i^{(b)} du \\
&\approx \frac{1}{\pi^2} \sum_{m_a=1}^{M} \sum_{m_b=1}^{M} \sum_{m_p=1}^{M} \sum_{m_q=1}^{M} w_{m_a} w_{m_b} w_{m_p} w_{m_q} \\
&\times h(\sqrt{2}\sigma_a\nu_{m_a} + \mu_a, \sqrt{2}\sigma_b\nu_{m_b} + \mu_b, \sqrt{2}\sigma_p\nu_{m_p} + \mu_p, \sqrt{2}\sigma_q\nu_{m_q} + \mu_q) \ .
\end{aligned}
$$

2.1.4 GH-quadrature Algorithms for MAP estimation

In this section we provide algorithms for implementing MAP estimation i.e. maximizing the posterior kernel, derived in section (2.1.2), where REs are integrated out using HQ-quadrature approximations as derived in section (2.1.3). There are two cases: a one dimensional case with one RE being integrated out and a four dimensional case where all four REs for a, b, p and q are integrated out.

In the 1D case, where we integrate out the REs attached to e.g. the parameter a, the MAP estimator for the structural parameters Θ_{MAP}^{1D} and all RE of the other parameters b, p and q not integrated out is

$$
\begin{aligned}
\Theta_{\text{MAP}}^{1D} &= \underset{\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{U}^{-(a)}}{\arg\max}\, p\left(\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{U}^{-(a)}|\bar{\boldsymbol{N}},\boldsymbol{X}\right) = \underset{\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{U}^{-(a)}}{\arg\max} \int_{\mathcal{U}^{\mathcal{A}}} p\left(\boldsymbol{\beta},\boldsymbol{U}^{(a)},\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{U}^{-(a)}|\bar{\boldsymbol{N}},\boldsymbol{X}\right) d\boldsymbol{U}^{(a)} \\
&= \underset{\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{U}^{-(a)}}{\arg\max} \left[\int_{\mathcal{U}^{\mathcal{A}}} p\left(\bar{\boldsymbol{N}}|\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{X}\right) p\left(\boldsymbol{U}^{(a)}|\boldsymbol{\sigma}\right) d\boldsymbol{U}^{(a)}\right] \times p\left(\boldsymbol{U}^{-(a)}|\boldsymbol{\sigma}^{-(a)}\right) p(\boldsymbol{\sigma}) p\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right) p(\boldsymbol{\tau}) \\
&= \underset{\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{U}^{-(a)}}{\arg\max} \\
&\prod_{i=1}^{N} \left\{ \underbrace{\left[\int_{\mathcal{U}_i^{\mathcal{A}}} \prod_{t=1}^{T} \text{MNL}_{it}\left(\pi_{it}^{(1)},\ldots,\pi_{it}^{(M_{it})}\right) p\left(u_i^{(a)}|\sigma^{(a)}\right) du_i^{(a)}\right]}_{\mathcal{I}_{RE-1dim}} \times p\left(u_i^{(b)}|\sigma_b\right) p\left(u_i^{(p)}|\sigma_p\right) p\left(u_i^{(q)}|\sigma_q\right) \right\} \\
&\times p(\boldsymbol{\sigma}) p\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right) p(\boldsymbol{\tau}) \,, \\
&\approx \underset{\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{U}^{-(a)}}{\arg\max} \\
&\underbrace{\prod_{i=1}^{N} \left\{ \left[\frac{1}{\sqrt{\pi}} \sum_{m_a=1}^{M} w_{m_a} h(\sqrt{2}\sigma_a \nu_{m_a} + \mu_a, u_i^{(b)}, u_i^{(p)}, u_i^{(q)})\right] \times p\left(u_i^{(b)}|\sigma_b\right) p\left(u_i^{(p)}|\sigma_p\right) p\left(u_i^{(q)}|\sigma_q\right) \right\}}_{\text{Part 1: Likelihood evaluation}} \\
&\times \underbrace{p(\boldsymbol{\sigma}) p\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right) p(\boldsymbol{\tau})}_{\text{Part 2: Prior evaluation}} \;.
\end{aligned}
$$

An algorithmic implementation of MAP estimation would maximize the sum of the logarithms of both parts, the likelihood "Part 1" and prior "Part 2".

Likelihood, Part 1:

$$\sum_{i=1}^{N} const_1 + \log\left(\sum_{m_a=1}^{M} w_{m_a} h(\sqrt{2}\sigma_a\nu_{m_a} + \mu_a, u_i^{(b)}, u_i^{(p)}, u_i^{(q)})\right) + const_2 - \left(\log(\sigma_b) + \frac{(u_i^{(b)} - \mu_b)^2}{2\sigma_b^2}\right)$$
$$+ const_3 - \left(\log(\sigma_p) + \frac{(u_i^{(p)} - \mu_p)^2}{2\sigma_p^2}\right) + const_4 - \left(\log(\sigma_q) + \frac{(u_i^{(q)} - \mu_b)^2}{2\sigma_q^2}\right) \;,$$

where $h(\sqrt{2}\sigma_a\nu_{m_a} + \mu_a, u_i^{(b)}, u_i^{(p)}, u_i^{(q)})$ equals

$$\exp\left\{\sum_{t=1}^{T}\left[\log(\bar{n}_{it}) - \sum_{k=1}^{M_{it}}\log(n_{it}^{(k)}!) + \sum_{k=1}^{M_{it}} n_{it}^{(k)}\log\left(B(f_{it}^{(k)}; p_{it}, q_{it}) - B(f_{it}^{(k-1)}; p_{it}, q_{it})\right)\right]\right\} \;,$$

and the knot transformation $h(\sqrt{2}\sigma_a\nu_{m_a} + \mu_a, \ldots)$ enters via

$$B(d_{it}^{(k)}; p_{it}, q_{it}) = \frac{\int_0^{d_{it}^{(k)}} t^{p-1}(1-t)^{q_{it}-1}dt}{\mathrm{B}(p_{it}, q_{it})} \;,\; d_{it}^{(k)} = \frac{(z_{it}^{(k)}/b_{it})^{a_{it}}}{1 + (z_{it}^{(k)}/b_{it})^{a_{it}}} \;,$$
$$a_{it} = \exp\left(\eta_{it}^{(a)}\right) = \exp\left(\left(x_{it}^{(a)}\right)' \beta^{(a)} + u_i^{(a)}\right) \;.$$

Prior, Part 2:

$$\begin{aligned}
&\log\left\{p(\boldsymbol{\sigma})p\left(\boldsymbol{\beta}|\boldsymbol{\tau}\right)p(\boldsymbol{\tau})\right\} \\
=&\log\left\{p(\sigma_a)p(\sigma_b)p(\sigma_p)p(\sigma_q) \times p\left(\boldsymbol{\beta_a}|\tau_a\right)p\left(\boldsymbol{\beta_b}|\tau_b\right)p\left(\boldsymbol{\beta_p}|\tau_p\right)p\left(\boldsymbol{\beta_q}|\tau_q\right) \times p(\tau_a)p(\tau_b)p(\tau_p)p(\tau_q)\right\} \\
=&\log\left(p(\sigma_a)\right) + \log\left(p(\sigma_b)\right) + \log\left(p(\sigma_p)\right) + \log\left(p(\sigma_q)\right) \\
&+\log\left(p\left(\boldsymbol{\beta_a}|\tau_a\right)\right) + \log\left(p\left(\boldsymbol{\beta_b}|\tau_b\right)\right) + \log\left(p\left(\boldsymbol{\beta_p}|\tau_p\right)\right) + \log\left(p\left(\boldsymbol{\beta_q}|\tau_q\right)\right) \\
&+\log\left(p(\tau_a)\right) + \log\left(p(\tau_b)\right) + \log\left(p(\tau_p)\right) + \log\left(p(\tau_q)\right) \\
\propto& -(\alpha_{\sigma_a}+1)\log(\sigma_a^2) - \frac{\beta_{\sigma_a}}{\sigma_a^2} - (\alpha_{\sigma_b}+1)\log(\sigma_b^2) - \frac{\beta_{\sigma_b}}{\sigma_b^2} \\
&-(\alpha_{\sigma_p}+1)\log(\sigma_p^2) - \frac{\beta_{\sigma_p}}{\sigma_p^2} - (\alpha_{\sigma_q}+1)\log(\sigma_q^2) - \frac{\beta_{\sigma_q}}{\sigma_q^2} \\
&-(\alpha_{\tau_a}+1)\log(\tau_a^2) - \frac{\beta_{\tau_a}}{\tau_a^2} - (\alpha_{\tau_b}+1)\log(\tau_b^2) - \frac{\beta_{\tau_b}}{\tau_b^2} \\
&-(\alpha_{\tau_p}+1)\log(\tau_p^2) - \frac{\beta_{\tau_p}}{\tau_p^2} - (\alpha_{\tau_q}+1)\log(\tau_q^2) - \frac{\beta_{\tau_q}}{\tau_q^2} \\
&-\frac{1}{2}d_{\beta_a}\log\left(\tau_a\right) - \frac{1}{2}\boldsymbol{\beta}_a^\top\tau_a^{-2}\boldsymbol{I}_{d_{\beta_a}}\boldsymbol{\beta}_a - \frac{1}{2}d_{\beta_b}\log\left(\tau_b\right) - \frac{1}{2}\boldsymbol{\beta}_b^\top\tau_b^{-2}\boldsymbol{I}_{d_{\beta_b}}\boldsymbol{\beta}_b \\
&-\frac{1}{2}d_{\beta_p}\log\left(\tau_p\right) - \frac{1}{2}\boldsymbol{\beta}_p^\top\tau_p^{-2}\boldsymbol{I}_{d_{\beta_p}}\boldsymbol{\beta}_p - \frac{1}{2}d_{\beta_q}\log\left(\tau_q\right) - \frac{1}{2}\boldsymbol{\beta}_q^\top\tau_q^{-2}\boldsymbol{I}_{d_{\beta_q}}\boldsymbol{\beta}_q
\end{aligned}$$

Algorithm: *1-dimensional integral approximation*

-