Supplementary materials for "Bayesian Nonparametric Instrumental Variables Regression based on Penalized Splines and Dirichlet Process Mixtures"

Manuel Wiesenfarth¹ Carlos Matías Hisgen² Thomas Kneib³

Carmen Cadarso-Suarez⁴

1 Full Conditionals

In the following, full conditionals for the parameters in the r -th equation, i.e. $r = 1$ for equation (4) and $r = 2$ for equation (3), are given.

Nonparametric effects The full conditionals for the regression coefficients of the smooth functions are Gaussian

$$\beta_{r\ell}|\cdot \sim N(\mu_{\beta_{r\ell}}, P_{\beta_{r\ell}}^{-1})$$

¹University of Mannheim, Department of Economics, L7 3-5, 68131 Mannheim, Germany

²Department of Economics, Universidad Nacional del Nordeste, Argentina

³Department of Statistics and Econometrics, Georg-August-Universität Göttingen, Germany

⁴Department of Biostatistics, School of Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain

with precision matrix

$$P_{\beta_{r\ell}} = X_{r\ell}^t \Sigma_{r|-r}^{-1} X_{r\ell} + \frac{\Delta_{r\ell}}{\tau_{r\ell}^2}$$

where $\Delta_{r\ell}$ is the penalty matrix of nonparametric effect ($r\ell$) based on a random walk prior and mean

$$\mu_{\beta_{r\ell}} = P_{\beta_{r\ell}}^{-1} X_{r\ell}^t \Sigma_{r|-r}^{-1} (y_r - \tilde{\eta}_r - E(\varepsilon_r | \varepsilon_{-r}))$$

where $\tilde{\eta}_r = \eta_r - f_{r\ell}$ when $f_{r\ell}$ is to be estimated. Further, $E(\varepsilon_r | \varepsilon_{-r})$ with $\varepsilon_r = (\varepsilon_{r11}, \dots, \varepsilon_{rnn_n})^t$ is the conditional mean of the error terms with

$$E(\varepsilon_{rij} | \varepsilon_{-r,ij}) = \mu_{rij} + \frac{\sigma_{12,ij}}{\sigma_{-r,ij}^2} (y_{-r,ij} - \mu_{-r,ij} - \eta_{-r,ij})$$

and $\Sigma_{r|-r}$ is the conditional covariance matrix with $\Sigma_{r|-r} = \text{diag}(\sigma_{(r|-r),11}^2, \dots, \sigma_{(r|-r),nn_n}^2)$ and

$$\sigma_{(r|-r),ij}^2 = \text{Var}(\varepsilon_{rij} | \varepsilon_{-r,ij}) = \sigma_{rij}^2 - \frac{\sigma_{12,ij}^2}{\sigma_{-r,ij}^2}.$$

Note that the posterior mean of some function $f_{r\ell}$ is given by (subject to centering constraints)

$$f_{r\ell}(\cdot) = (X_{r\ell}^t \Sigma_{r|-r}^{-1} X_{r\ell} + \frac{1}{\tau_{r\ell}^2} \Delta_{r\ell})^{-1} X_{r\ell}^t \Sigma_{r|-r}^{-1} (y_r - \tilde{\eta}_r - E(\varepsilon_r | \varepsilon_{-r})).$$

Here, it can be easily seen that the DPM prior induces different variances and therefore $\Sigma_{r|-r}$ weighs observations accordingly just as in the case of heteroscedasticity.

The full conditionals for the smoothing variance parameters $\tau_{r\ell}^2$, $\ell = 1, \dots, p_r$, $r = 1, 2$ follow inverse Gamma distributions

$$\tau_{r\ell}^2 | \cdot \sim IG(a'_{\tau_{r\ell}}, b'_{\tau_{r\ell}})$$

with parameters

$$a'_{\tau_{r\ell}} = a_{\tau_{r\ell}} + \frac{\text{rank}(\Delta_{r\ell})}{2}, \quad b'_{\tau_{r\ell}} = b_{\tau_{r\ell}} + \frac{1}{2}\beta_{r\ell}^t \Delta_{r\ell} \beta_{r\ell}.$$

Parametric effects The full conditionals for the coefficients γ_r of parametric effects are Gaussian

$$\gamma_r | \cdot \sim N(\mu_{\gamma_r}, P_{\gamma_r}^{-1})$$

with precision matrix $P_{\gamma_r} = V_r^t \Sigma_{r|-r}^{-1} V_r$

and mean $\mu_{\gamma_r} = P_{\gamma_r}^{-1} V_r^t \Sigma_{r|-r}^{-1} (y_r - \tilde{\eta}_r - E(\varepsilon_r | \varepsilon_{-r}))$

where $\tilde{\eta}_r = \eta_r - V_r \gamma_r$.

Components of the error distribution In our default implementation, we make use of R function DPdensity (Jara et al., 2011) for error density estimation adopting algorithm 8 of Neal (2000) with one temporarily existing auxiliary parameter. In the following, the full conditionals are summarized, for more details see Neal (2000).

- Let $c_i \in \{1, \dots, K^*\}, i = 1, \dots, n$ indicate the cluster observation i belongs to.

For $i = 1, \dots, n$:

- If $c_i = c_h$ for some $h \neq i$, create auxiliary component c^* with $(\mu_{c^*}, \Sigma_{c^*})$ drawn from G_0 .
- If $c_i \neq c_h$ for all $h \neq i$, let $c^* = c_i$ with $(\mu_{c^*}, \Sigma_{c^*}) = (\mu_{c_i}, \Sigma_{c_i})$.

- Draw a new value for c_i using

$$c_i | c_{-i}, y_{1i}, y_{2i}, \mu_1, \Sigma_1, \dots, \mu_{K^*}, \Sigma_{K^*}, \mu_{c^*}, \Sigma_{c^*} \sim b \sum_{l=1}^{k^-} \frac{n_l^{-i}}{n-1+\alpha} F((\varepsilon_{1i}, \varepsilon_{2i}), \mu_l, \Sigma_l) + b \frac{\alpha}{n-1+\alpha} F((\varepsilon_{1i}, \varepsilon_{2i}), \mu_{c^*}, \Sigma_{c^*})$$

where k^- is the number of distinct c_h for $h \neq i$, n_l^{-i} is the number of c_h for $h \neq i$ that are equal to l , b is a normalizing constant and $F((\varepsilon_{1i}, \varepsilon_{2i}), \mu_l, \Sigma_l)$ the likelihood for observation i .

- Discard those μ_l, Σ_l that are not associated with one or more observations.
- For all $l \in \{c_1, \dots, c_n\}$: Update μ_l and Σ_l using $\mu_l | \cdot \sim N(m_{\mu_l}, P_{\mu_l}^{-1})$ and $\Sigma_l | \cdot \sim \text{IW}(s'_\Sigma, S'_\Sigma)$ with

$$\begin{aligned} m_{\mu_l} &= (\tau_\Sigma + 1)^{-1} \left(\tau_\Sigma \mu_0 + \sum_{i:c_i=l} ((y_{1i}, y_{2i}) - (\eta_{1i}, \eta_{2i}))^t \right) \\ P_{\mu_l}^{-1} &= \frac{\tau_\Sigma^{-1}}{1 + \tau_\Sigma^{-1}} \Sigma_l / n_l = (\tau_\Sigma + 1)^{-1} \Sigma_l / n_l, \\ s'_\Sigma &= s_\Sigma + \frac{n_l}{2} \\ S'_\Sigma &= S_\Sigma + \frac{1}{2} \frac{1}{1 + \tau_\Sigma^{-1}} \sum_{i:c_i=l} ((y_{1i}, y_{2i}) - (\eta_{1i}, \eta_{2i}) - \mu_0)^t ((y_{1i}, y_{2i}) - (\eta_{1i}, \eta_{2i}) - \mu_0) \end{aligned}$$

- In case τ_Σ is not fixed, the full conditionals of τ_Σ are

$$\tau_\Sigma \sim \text{Ga} \left(\frac{a_\Sigma + K^*}{2}, \frac{1}{2} \left(b_\Sigma + \sum_{l=1}^{K^*} \Sigma_l^{-1} (\mu_l - \mu_0)^2 \right) \right)$$

- The concentration parameter α in case of a gamma prior is drawn from a mixture

of two gamma distributions

$$\alpha|\cdot \sim \frac{a_\alpha + K^* - 1}{n(b_\alpha - \log \omega)} \text{Ga}(a_\alpha + K^*, b_\alpha - \log \omega) + \left(1 - \frac{a_\alpha + K^* - 1}{n(b_\alpha - \log \omega)}\right) \text{Ga}(a_\alpha + K^* - 1, b_\alpha - \log \omega)$$

where ω is a latent variable sampled from a beta distribution $\omega \sim \text{Be}(\alpha + 1, n)$.

In case of a discrete prior for α as in Conley et al. (2008), α is drawn from a multinomial distribution. See Conley et al. (2008) for details.

References

- Conley, T., Hansen, C., McCulloch, R., and Rossi, P. (2008). A semi-parametric Bayesian approach to the instrumental variable problem. *Journal of Econometrics*, 144(1):276–305.
- Jara, A., Hanson, T., Quintana, F., Müller, P., and Rosner, G. (2011). DPpackage: Bayesian semi- and nonparametric modeling in R. *Journal of Statistical Software*, 40(5):1–30.
- Neal, R. (2000). Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265.