Flexible Bayesian treatment effects models for panel outcomes

Helga Wagner

based on joint work with Sylvia Frühwirth Schnatter and Liana Jacobi

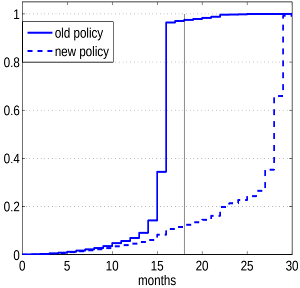
WU Vienna, April 2024

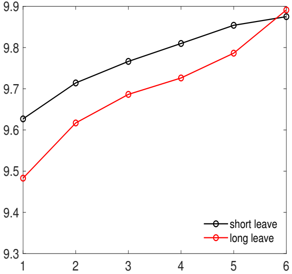
Motivation: motherhood gap

- mothers earn less than women without children
- maternity leave after childbirth causes break in mother's employment history
  (loss of human capital, loss of rents from good job matches)
- large variation in maternity leave duration due to observed and unobserved factors
- different parental leave policies (job protection, financial benefit)

Maternity Leave Policy and Earnings in Austria

- policy change in Austria (July 1, 2000)
  - extension of financial benefit period from 18 to 30 months
  - job protection 24 months (not changed)





Left: proportion of mothers returning to the labour market
right: mean income for mothers with short and long maternity leave

What is the effect of extending maternity leave beyond 18 months on income?

Analysis of Long Leave Effects

- Goal: Estimate the effects of a long maternity leave on earnings after return to labour market
  - earnings of each mother observed only for her choice on length of the maternity leave $\Longrightarrow$ typical situation in treatment effects models
  - individual decision on length of maternal leave $\Longrightarrow$ control for endogeneity
  - earnings are observed over several years $\Longrightarrow$ panel data
- modelling approach
  - capture dependencies
    - between treatment choice and panel earnings outcomes
    - across panel outcomes under each treatment
  - allow for time-varying effects of covariates on earnings

Modelling Treatment Effects: Potential Outcomes

potential outcomes framework (Rubin, 2005)

- binary treatment indicator

$$X_i = \begin{cases} 1 & \text{treatment} \\ 0 & \text{control} \end{cases}$$

- outcome of interest is described by two random variables

Y_{0i}	potential outcome under control conditions
Y_{1i}	potential outcome under treatment

Modelling Treatment Effects: SUTVA

- SUTVA (stable unit treatment value assumption)
  - non-interference among units: potential outcomes of unit *i* are unaffected by treatment assignment on unit *j*

$$Y_{x_i,i}|(X_1 = x_1, \ldots, X_N = x_N) = Y_{x_i,i}$$

  - no hidden variations in treatment
- the potential outcomes model and SUTVA allow to define treatment effects of interest (Li, Ding and Mealli, 2023)

Treatment effects

treatment effects are defined based on the individual outcome differences $Y_{1i} - Y_{0i}$

- individual treatment effect (ITE)

$$\tau_i = \mathrm{E}(Y_{1i} - Y_{0i})$$

- sample average treatment effect (SATE) (i.e. average gain/loss from treatment in the sample)

$$\tau^S = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(Y_{1i} - Y_{0i})$$

- population average treatment effect (PATE) (i.e. average gain/loss from treatment in the population)

$$\tau^P = \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}(Y_{1i} - Y_{0i})$$

where N is the size of the population

Inference on treatment effects

- fundamental problem for inference on treatment effects
  - for each subject only one of the two potential outcomes is observed
  - the outcome difference is never observed

Treatment	Y_{i1}	Y_{i0}
$X_i = 1$	$Y_{i1} \| (X_i = 1)$	$Y_{i0} \| (X_i = 1)$
$X_i = 0$	$Y_{i1} \| (X_i = 0)$	$Y_{i0} \| (X_i = 0)$

- the observed outcome depends on the value of X_i

$$Y_i = \begin{cases} Y_{0i} & \text{if } x_i = 0 \\ Y_{1i} & \text{if } x_i = 1 \end{cases}$$

Confounding

- in randomized trials
  - treatment and realized outcome are independent
  - difference of mean outcomes is unbiased for τ^S
- in observational studies
  - individuals choose treatment/no treatment based on expectations
  - simple estimates are biased due to confounding
- treatment selection
  - on observables
    - potential outcomes are independent of treatment selection conditional on observed covariates
    - flexible models for outcomes e.g. via BART (Hill, 2011; Hahn et al., 2021)
  - on unobservables
    dependence between selection into treatment and potential outcomes after conditioning on covariates

Data

- for $i = 1, \ldots, n$ subjects
  - binary treatment x_i
  - covariates $\mathbf{v}_i$ at treatment
  - observed outcome y_i after treatment
  - covariates $\mathbf{w}_i$ for outcome
- observed data for treated and untreated

$$(x_1 = 0, y_{01}, \mathbf{v}_1, \mathbf{w}_1), \ldots, (x_{n_0} = 0, y_{0,n_0}, \mathbf{v}_{n_0}, \mathbf{w}_{n_0})$$
$$(x_{n_0+1} = 1, y_{1,n_0+1}\mathbf{v}_{n_0+1}, \mathbf{w}_{n_0+1}), \ldots, (x_n = 1, y_{1,n}\mathbf{v}_n, \mathbf{w}_n)$$

$x = 0,\ y_0$	$\mathbf{v},\ \mathbf{w}$	
	$\mathbf{v},\ \mathbf{w}$	$x = 1,\ y_1$

- Relation between observed treatment x_i and outcome $\mathbf{y}_i$:

$$y_i = \begin{cases} y_{0i} & \text{if } x_i = 0 \\ y_{1i} & \text{if } x_i = 1 \end{cases}$$

Joint model of treatment selection and potential outcomes

- Probit model for binary treatment x_i at baseline

$$x_i^* = \mathbf{v}_i'\boldsymbol{\alpha} + \varepsilon_{xi} \qquad \varepsilon_{xi} \sim \mathcal{N}\left(0, \sigma_x^2\right)$$
$$x_i = I_{[0,\infty)}(x_i^*)$$

- Regression model for the two potential outcomes y_{ji}

$$y_{0i} = \gamma_0 + \mathbf{w}_i'\gamma + \varepsilon_{0i}$$
$$y_{1i} = (\gamma_0 + \kappa_0) + \mathbf{w}_i'(\gamma + \boldsymbol{\kappa}) + \varepsilon_{1i}$$

- interest is in the treatment effect conditional on covariate values **w** (CATE)

$$\tau(\mathbf{w}) = \mathrm{E}(Y_{1i} - Y_{0i}|\mathbf{w}) = \kappa_0 + \mathbf{w}'\boldsymbol{\kappa}$$

Dependence between treatment and outcome

- regression model for the observed outcomes

$$y_i = \gamma_0 + \mathbf{w}_i'\gamma + x_i(\kappa_0 + \mathbf{w}_i'\boldsymbol{\kappa}) + \varepsilon_{x_i,i}$$

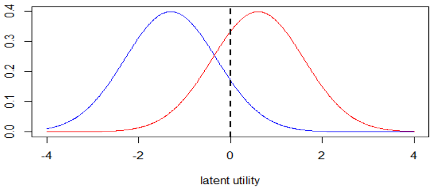
- endogeneity: errors of the observed outcome depend on x_i
- specification of a joint Normal distribution of all error terms

$$\begin{pmatrix} \varepsilon_{xi} \\ \varepsilon_{0i} \\ \varepsilon_{1i} \end{pmatrix} \sim \mathcal{N}_3 \left(\mathbf{0}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_x^2 & \sigma_{x0} & \sigma_{x1} \\ \sigma_{x0} & \sigma_0^2 & \sigma_{01} \\ \sigma_{x1} & \sigma_{01} & \sigma_1^2 \end{pmatrix} \right).$$

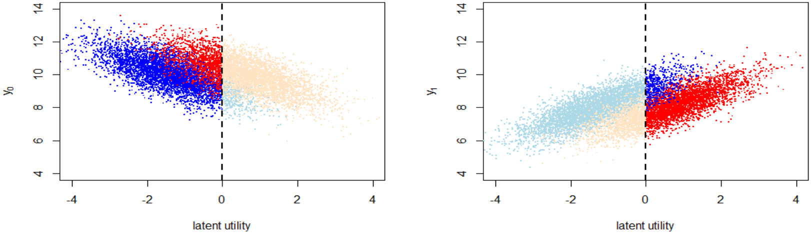
- y_{0i} observed if $x_i^* < 0$; y_{1i} observed if $x_i^* > 0$ but never together
- observed data allow identification of the CATE but not of σ_{01}

Latent utilities and observed outcomes

- latent utility model: one binary covariate



- potential outcomes: different correlation with latent utilty ($\rho_{x0} = -0.7$, $\rho_{x1} = 0.8$)



Panel outcome data

Data structure: information for $i = 1, \ldots, n$ subjects

- at treatment
  - binary variable indicating treatment x_i (0=shorter / 1= longer leave)
  - covariates $\mathbf{v}_i$
- after treatment
  - panel data on outcome $\mathbf{y}_i = \{y_{i1}, y_{i2}, ..., y_{iT_i}\}$ at T_i time points after treatment
  - covariates $\mathbf{W}_i = \{\mathbf{w}_{i1}, \mathbf{w}_{i2}, ..., \mathbf{w}_{iT_i}\}$ at different time points
- interest is in the longitudinal conditional treatment effect

$$\tau(\mathbf{W}) = \mathrm{E}(\mathbf{Y}_{1i} - \mathbf{Y}_{0i} | \mathbf{W})$$

Dependence between treatment and outcome

- outcomes sequences of length T, i.e. joint distribution of dimension (2T+1)
- joint Normal distribution of all error terms

$$\begin{pmatrix} \varepsilon_{0i} \\ \varepsilon_{1i} \\ \varepsilon_{xi} \end{pmatrix} \sim \mathcal{N}_{2T+1}\left(\mathbf{0}, \begin{pmatrix} \boldsymbol{\Sigma}_0 & \Sigma_{01} & \boldsymbol{\sigma}_0 \\ \Sigma_{01} & \boldsymbol{\Sigma}_1 & \boldsymbol{\sigma}_{x1} \\ \boldsymbol{\sigma}'_{x0} & \boldsymbol{\sigma}'_{x1} & \sigma_x^2 \end{pmatrix}\right).$$

Σ_{01} cannot be identified from the observed data

- under longitudinal dependence $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ are not diagonal

Modelling Dependence

- Shared factor model (SF)
  (Carneiro etal., 2003; Jacobi et al. 2016)
  - specification of the joint $2T + 1$ variate distribution of $(\varepsilon_{0i}, \varepsilon_{1i}, \eta_i)$
  - 1 latent factor captures all dependencies
  - dependence between latent utility and each potential outcome induces dependence between the two potential outcomes

- Switching regression model (SR)
  (Chib, 2007; Chib and Jacobi, 2007; Jacobi et al. 2016)
  - model only $\begin{pmatrix} \mathbf{\Sigma}_j & \boldsymbol{\sigma}_{xj} \\ \boldsymbol{\sigma}'_{xj} & \sigma^2_x \end{pmatrix}$; no specification of $\mathbf{\Sigma}_{01}$
  - sufficient for point estimates of treatment effects
  - implicit restrictions for joint Normal distribution

- for a joint multivariate Normal distribution both models imply restrictions that can result in biased treatment effects estimates

The Bifactor model

- Holzinger and Swineford (1937): two or more orthogonal factors
  - one common (or general) factor shared by all responses
  - one or more further group (or specific) factors model the additional correlation among clusters of responses
- application to treatment effects models: subject specific factors
  - 1 common factor f_{ci}
  - two 2 specific factors f_{ji} - one for each potential outcomes sequence

$$\begin{pmatrix} \bullet & & \\ \bullet & \bullet & \\ \bullet & \bullet & \\ \bullet & \bullet & \\ \bullet & & \bullet \\ \bullet & & \bullet \\ \bullet & & \bullet \end{pmatrix}$$

The bifactor treatment effects model

- bifactor treatment effects model

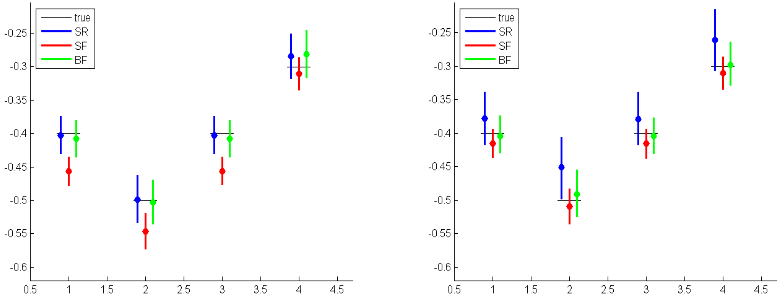
$$\varepsilon_{xi} = \lambda_x f_{ci} + \epsilon_{xi}, \tag{1}$$
$$\boldsymbol{\varepsilon}_{0i} = \boldsymbol{\lambda}_0 f_{ci} + \boldsymbol{\zeta}_0 f_{0i} + \boldsymbol{\epsilon}_{0i} \tag{2}$$
$$\boldsymbol{\varepsilon}_{1i} = \boldsymbol{\lambda}_1 f_{ci} + \boldsymbol{\zeta}_1 f_{1i} + \boldsymbol{\epsilon}_{1i} \tag{3}$$

  - common factor f_{ci} shared by the latent utility x^* and both potential outcomes sequences $\mathbf{y}_0$ and $\mathbf{y}_1$
  - two group factors f_{0i}, f_{1i} for the potential outcomes sequences
- more general than both the SR and SF model
- identification of factor loadings from $\boldsymbol{\sigma}_{xj}$ and $\boldsymbol{\Sigma}_j$ $(j = 0, 1)$ requires outcome panels of length $T \geq 4$

Simulation results

Data generated from

- SR model violating restrictions of SF model
- SF model violating restrictions of SR model



True and estimated average treatment effects
(left: true model SR, right: true model SF)

Application: Effect of maternity leave on income after return

- Data from the Austrian Social Security Register (ASSD is an administrative individual register that collects information for old-age security benefits)
- unbalanced sample of $n = 31015$ mothers
  - birth of last child between July 1998 and June 2002
  - observed 4- 6 panel periods after returning to the labor market
  - employed in the private sector before child birth
  - strong attachment to the labour market
    - ★ employed within 30 days after end of maternity leave
    - ★ earnings $>$ 1100 Euros in subsequent years after reentry

$\Longrightarrow$ 190969 earnings observations

Modelling the mother data

- binary treatment x_i defined based on maternity leave duration m_i

$$x_i = \begin{cases} 0, & \text{if } m_i \leq 18, \\ 1, & \text{if } m_i > 18. \end{cases}$$

- the binary instrument z_i defined as

$$z_i = \begin{cases} 0 & \text{child born before June 30, 2000} \\ 1 & \text{child born after June 30, 2000} \end{cases}$$

- covariates in the selection equation
  - number of children
  - white/blue collar
  - working experience before maternity leave
  - baseline earnings before child
- labor market outcomes (log real earnings per year) $\mathbf{y}_i = \{y_{i1}, y_{i2}, ..., y_{iT_i}\}$ observed for each mother after the end of the maternity leave
- covariates for the outcome model as in the selection model and additionally dummy variable for return to the same employer

Modelling the mother data

- structural model

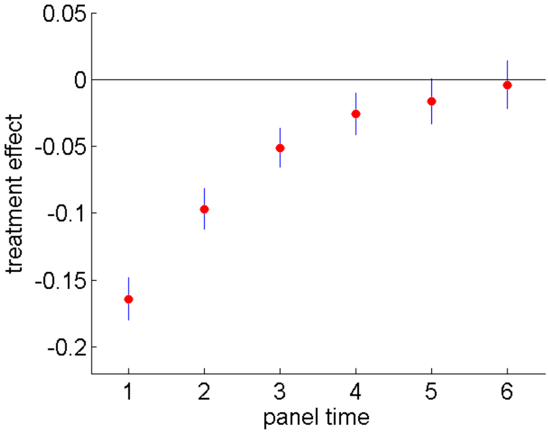
$$
\begin{aligned}
\mathrm{E}(x_i^*) &= \mathbf{v}_i'\boldsymbol{\alpha} \\
\mathrm{E}(y_{0,it}) &= \gamma_{0t} + \mathbf{w}_i'\gamma \\
\mathrm{E}(y_{1,it}) &= (\gamma_{0t} + \kappa_{0t}) + \mathbf{w}_i'(\gamma + \boldsymbol{\kappa})
\end{aligned}
$$

- the model might be overspecified
  - covariate with no effect on selection or outcome
  - no baseline treatment effect κ_0
  - no heterogeneity of treatment effects with respect to a covariate
- enforce sparsity by spike and slab priors on regression effects

Results: Outcome model

	18 months or less		+ longer leave	
	mean(sd)	p	mean(sd)	p_{incl}
intercept	9.309 (0.014)	-	-0.130 (0.011)	1.00
2 children	0.000 (0.001)	0.01	0.000 (0.001)	0.01
$\geq$3 children	0.000 (0.001)	0.01	0.000 (0.002)	0.02
more exp.	-0.092 (0.010)	1.00	0.011 (0.015)	0.39
blue collar	-0.108 (0.006)	1.00	0.000 (0.002)	0.02
more exp., blue	0.001 (0.004)	0.04	0.005 (0.012)	0.17
base earn Q2	0.066 (0.006)	1.00	0.000 (0.002)	0.02
base earn Q3	0.286 (0.011)	1.00	-0.047 (0.014)	1.00
base earn Q4	0.606 (0.010)	1.00	-0.116 (0.013)	1.00
equal employer	0.049 (0.005)	1.00	0.000 (0.003)	0.03
panel T=2	0.066 (0.004)	1.00	0.068 (0.004)	1.00
panel T=3	0.106 (0.006)	1.00	0.115 (0.006)	1.00
panel T=4	0.149 (0.009)	1.00	0.141 (0.008)	1.00
panel T=5	0.201 (0.011)	1.00	0.151 (0.009)	1.00
panel T=6	0.252 (0.013)	1.00	0.163 (0.010)	1.00

Longitudinal treatment effects



sample treatment effect

- negative short-term effect
- no long-term impact

Modelling the mother data

- potential outcomes model

$$
\begin{aligned}
y_{0,it} &= \gamma_{0t} + \mathbf{w}_i'\gamma + \varepsilon_{0,it} \\
y_{1,it} &= (\gamma_{0t} + \kappa_{0t}) + \mathbf{w}_i'(\gamma + \boldsymbol{\kappa}) + \varepsilon_{1,it}
\end{aligned}
$$

  - intercept and treatment effects time-specific (unstructured)
  - time-constant effects of covariates

- more flexible model with time-varying parameter model (TVP)

$$
\begin{aligned}
y_{0,it} &= \gamma_{0t} + \mathbf{w}_i'\gamma_t + \varepsilon_{0,it} \\
y_{1,it} &= (\gamma_{0t} + \kappa_{0t}) + \mathbf{w}_i'(\gamma_t + \boldsymbol{\kappa}_t) + \varepsilon_{1,it}
\end{aligned}
$$

- TVP models usually used for time series data with small n, large T – we have panel data with large n, small T

Time-varying effects

- Notation: $d \times 1$ vector of regression effects $\boldsymbol{\beta}_t = (\gamma_{0t}, \boldsymbol{\gamma}_t, \kappa_{0t}, \boldsymbol{\kappa}_t)'$ $\Longrightarrow T \times d$ regression effects (including the intercept)
- modelling of the development of the regression effects over time
  - allows to borrow information across time
  - a simple model for the dynamics over time is the Normal random walk

    $$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$$

    with $\mathbf{Q} = \mathrm{diag}(\theta_1^2, \ldots, \theta_d^2)$ and starting values

    $$\boldsymbol{\beta}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_0)$$

  - p starting values+ p process variances $\Longrightarrow$ 2d effects

Non-centered parameterisation

- regression effects in non-centered parameterization

$$\boldsymbol{\beta}_t = \boldsymbol{\beta} + \boldsymbol{\Theta}\tilde{\boldsymbol{\beta}}_t$$
$$\tilde{\boldsymbol{\beta}}_t = \tilde{\boldsymbol{\beta}}_{t-1} + \tilde{\boldsymbol{\omega}}_t, \qquad \tilde{\boldsymbol{\omega}}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

  where $\boldsymbol{\Theta} = \text{diag}(\theta_1, \ldots, \theta_d)$ and $\tilde{\beta}_0 \sim \mathcal{N}(0, c\mathbf{I})$
- the model is then given as

$$y_{j,it} = \tilde{\mathbf{w}}_{it}\boldsymbol{\beta} + \tilde{\mathbf{w}}_{it}\boldsymbol{\Theta}\tilde{\boldsymbol{\beta}}_t + \varepsilon_{j,it}, \qquad j = 0, 1$$

- the model is overspecified if elements of $\boldsymbol{\beta}$ and/or $\boldsymbol{\Theta}$ are 0
  - time-constant effect: $\theta_j = 0$
  - no effect: $\theta_j = 0$ and $\beta_j = 0$

Shrinkage priors

- advantages in inference: effects need not be assigned to one of the components as in spike and slab priors
- specified hierarchically with hyperpriors
- triple Gamma priors (Cadonna et al., 2020) on the process variances θ_j^2

$$\theta_j^2|\xi_j^2 \sim \mathcal{G}\left(\frac{1}{2}, \frac{1}{2\xi_j^2}\right),$$

$$\xi_j^2|a_\xi, \kappa_{j(\xi)}^2 \sim \mathcal{G}\left(a_\xi, \frac{a_\xi \kappa_{j(\xi)}^2}{2}\right),$$

$$\kappa_{j(\xi)}^2|c_\xi, \kappa_{B(\xi)}^2 \sim \mathcal{G}\left(c_\xi, \frac{c_\xi}{\kappa_{B(\xi)}^2}\right)$$

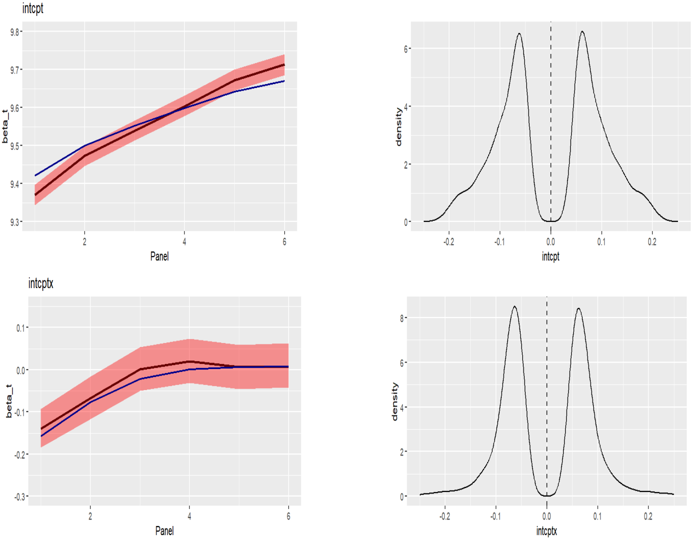
  - Normal-Gamma-Gamma priors (NGG) on the signed process standard deviations $\pm\theta_j$
  - good shrinkage properties in time-series

Mother's Earnings

- data: balanced panel of 18846 mothers with 6 outcome observations (total: 113076 outcome observations)
- intercepts and all effects on potential outcomes time-varying
- priors
  - NGG prior on effects α_j in selection model
  - NGG prior on initial effects β_j^2 in outcome model
  - triple Gamma Prior on process variances θ_j^2

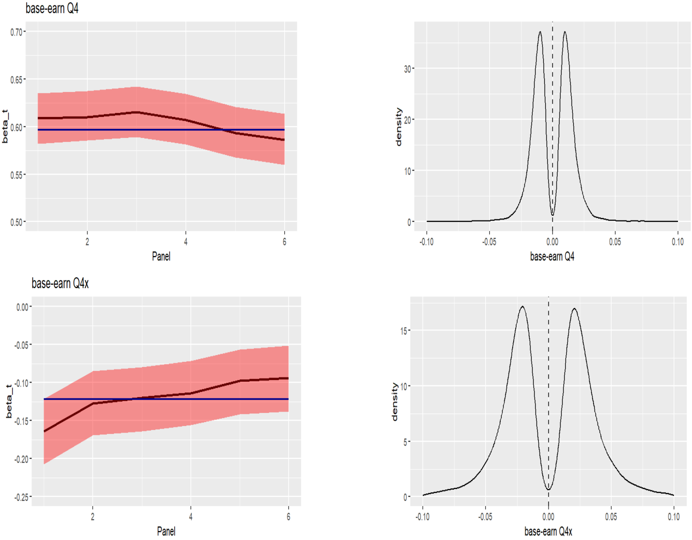
  with fixed default hyper-parameters (a=1/7, c=1/7,$\kappa_B = 1$)
- comparison to a model with unstructured effects of panel time and interaction of treatment and panel time

Mother’s Earnings: Time varying effects

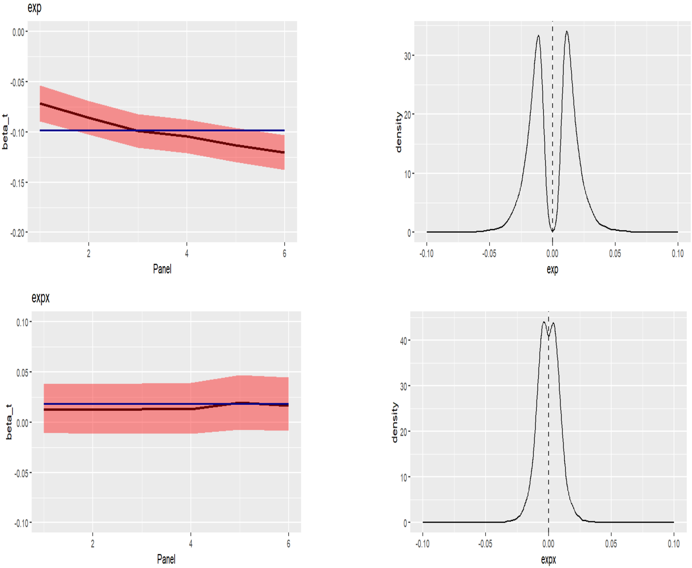


Intercept and baseline treatment effect

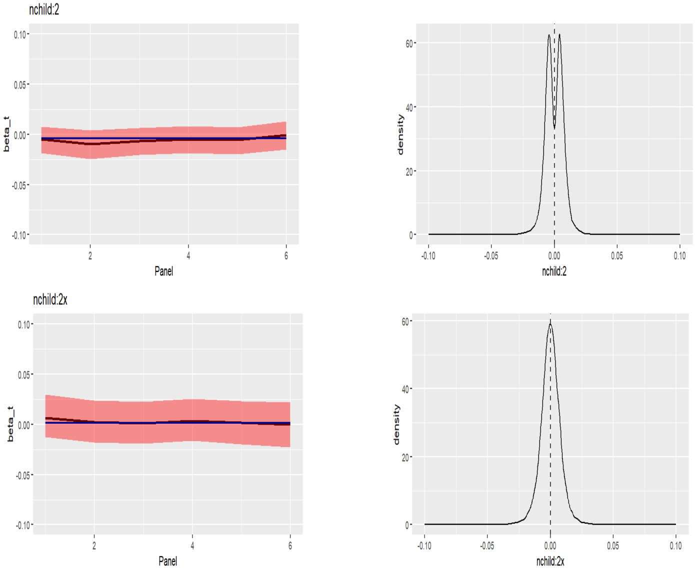
Mother's Earnings: baseline earnings in quartile 4



Mother's Earnings: more than median experience



Mother’s Earnings: two children



Discussion

- estimated longitudinal treatment effects consistent with literature (Lalive and Zweimüller, 2009; Lalive et al. 2014):
  - negative short-term effects (reduction in earnings)
  - no long-term impacts

  for baseline mother but heterogeneity of treatment effects
- estimated factor loadings confirm endogeneity in leave decision
- two factors are supported by data
  - unobserved confounders to explain correlation across time in potential earnings (sign depends on treatment state, i.e. the general factor)
  - additional unobserved non-confounding factors (specific factor)
- many effects vary over time - good shrinking properties of triple Gamma prior also for panel data

Open issues

- prior
  - choice of hyper-parameters of the shrinkage priors
  - shrinkage prior on factor loadings?
- MCMC
  - implementation for unbalanced panel
  - handle intermittent missing values
  - high autocorrelation in MCMC draws of factor loadings
  - extension to other outcome types (tobit, skew-normal,. . .)

References

- Anderson T. W. and Rubin H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, V: 111–150.
- Bitto, A. and Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics* 210, 75-97.
- Cadonna, A., Frühwirth-Schnatter, S. and Knaus, P. (2020). Triple the Gamma - A unifying shrinkage prior for variance and variable selection in sparse state space and TVP models. *Econometrics*, 8(2), 1-36.
- Carneiro P. , Hansen K.-T. and Heckman J. J. (2003). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty of college choice. *International Economic Review*, 44: 361–422.
- Chib, S. and Hamilton B.H. (2000). Bayesian analysis of cross-section and clustered data treatment models," *Journal of Econometrics*, 97(1), 25-50.
- Chib, S. and Hamilton B.H. (2002). Semiparametric Bayes Analysis of Longitudinal Data Treatment Models, *Journal of Econometrics*, 110, 67-89.
- Chib S. (2007). Analysis of treatment response data without the joint distribution of potential outcomes. *Journal of Econometrics*, 140: 401–412.
- Chib S. and Jacobi L. (2007). Modeling and calculating the effect of treatment at baseline from panel outcome. *Journal of Econometrics*, 140: 781–801.
- Hahn R. P., Murray J. and Carvalho C. M. (to appear). Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects.

References

- Heckman J.J., Lopes H. F. and Piatek R. (2014). Treatment effects: A Bayesian perspective *Econometric Reviews*, 33(1–4):36–67.
- Hill, J. L. (2011). Bayesian Nonparametric Modeling for Causal Inference, *Journal of Computational and Graphical Statistics*, 20, 217–240.
- Li, F., Ding P. and Mealli, F. (2023). Bayesian causal inference: a critical review
- Li, M., Tobias, J. L. (2008). Bayesian analysis of treatment effects in an ordered potential outcomes model. In: Fomby, T., Hill, R. C., Millimet, D. L., Smith, J. A., Vytlacil, E. (ed.): Modelling and Evaluating Treatment Effects-
- Holzinger K. and Swineford F. (1937). The Bi-factor method. *Psychometrika*, 2: 41–54 .
- Jacobi L., Wagner H., Frühwirth-Schnatter, S. (2016). Bayesian treatment effects models with variable selection for panel outcomes with an application to earnings effects of maternity leave. *Journal of Econometrics*, 193: 234–250.
- Koop G. and Poirier , D. J. (1997). Learning about the across-regime correlation in switching regression models. Journal of Econometrics 78:217-227.
- Li, M., Poirier, D. J., Tobias, J. L. (2004). Do dropouts suffer from dropping out? Estimation and prediction of outcome gains in generalized selection models. *Journal of Applied Econometrics* 19:203–225.
- Poirier D. J. (1998). Revising beliefs in nonidentified models. *Econometric Theory* 14:483–509.

References

- Poirier D. J. and Tobias J. L. (2003). On the predictive distributions of outcome gains in the presence of an unidentified parameter. *Journal of Business and Economic Statistics, 21:258–268.*
- Rubin D. B. (2005). Bayesian inference for causual effects. In D. Dey and C. Rao (Eds.), Bayesian Thinking, Modelling and Computation, Handbook of Statistics, Vol. 25, Chapter 1, pp. 1-16.
- Vijverberg W. P. M. (1993). Measuring the unidentified parameter of the extended Roy model of selectivity. *Journal of Econometrics* 57, 69–89.

Inference on treatment effects in randomized trials

- random assignment of treatment: treatment and potential outcomes are independent

$$X \perp\!\!\!\perp (Y_1, Y_0)$$

- consequences for the potential outcomes

$$p(Y_0, Y_1 | X = 1) = p(Y_0, Y_1 | X = 0) = p(Y_1, Y_0)$$

  and therefore

$$E(Y_0) = E(Y_0 | X = 0)$$
$$E(Y_1) = E(Y_1 | X = 1)$$

- τ^S can be estimated unbiasedly by the observed outcome difference

$$\widehat{\tau^S} = \frac{1}{n_1} \sum_{i:x_i=1} y_{1i} - \frac{1}{n_0} \sum_{i:x_i=0} y_{0i}$$

Sparse Bayesian modelling

in a Bayesian approach sparsity can be achieved by appropriate prior distributions

- spike and slab priors
  - two component mixture (spike/slab)
  - allow classification of effects as relevant or not
- shrinkage priors
  - mode at zero and fat tails
  - shrink irrelevant effects to zero