bite - an R-Package for Bayesian Inference on Treatment Effects

Master’s Thesis
to obtain the academic degree of
Master of Science
in the Master Program
Statistics
SWORN DECLARATION

I hereby declare under oath that the submitted Master’s Thesis has been written solely by me without any third-party assistance, information other than provided sources or aids have not been used and those used have been fully documented. Sources for literal, paraphrased and cited quotes have been accurately credited.

The submitted document here present is identical to the electronically submitted text document.

Linz, 15.11.2019

Patrick Pfeiffer
Contents

1 Introduction .......................................................... 4

2 Treatment Effects Modeling for Panel Outcomes ............ 6
   2.1 Shared Factor (SF) Approach ................................. 9
   2.2 Switching Regression (SWR) Approach ....................... 10
   2.3 Data Structure Implications ................................. 12
   2.4 Bayesian Inference on Treatment Models ................... 14
   2.5 Posterior Inference ............................................. 16

3 Simulation Studies .................................................... 21
   3.1 Simulation Details ............................................. 21
   3.2 Simulation Results ............................................. 22

4 Application on Economic Data ........................................ 33
   4.1 Effect of Long Maternity Leave on Income of Mothers ........ 33
       4.1.1 Background .................................................. 33
       4.1.2 Modeling results ........................................... 36
   4.2 Effect of High School Athletics Program on Future Income .... 41
       4.2.1 Modeling Results ........................................... 42

5 Computational Aspects of 'bite' ..................................... 49
   5.1 Computations with 'bite' ....................................... 49
       5.1.1 Overall performance ....................................... 53
       5.1.2 Code Profiling of 'bite' .................................... 54

6 Conclusion ............................................................. 56
Chapter 1

Introduction

The main goal of this thesis is the development of an R-package for Bayesian inference on treatment effects. The package 'bite', short for ”Bayesian Inference on Treatment Effects” is currently freely available within the R framework, easily accessible through github (https://github.com/PatrickPfeifferDsc/bite). The main functionality of 'bite' is a treatment effect estimation in a Bayesian paradigm for short or midterm outcomes in a panel format. The work is based on previous findings of Jacobi, Wagner, and Frühwirth-Schnatter [2016]. The thesis gives insight on the theoretical foundations of the implemented models, estimation procedures and computational aspects. To show the functionality real data applications and a simulation study are carried out. The applications are performed on two separate economic data sets. The package can be used in a treatment setting with observed subjects who could choose whether or not to intake treatment at a given point in time. The subsequent outcomes represent the response variables, which can be the realization of some medical issues or a socio-economic indicator like income. In settings where the outcomes are observed as a panel, inter-correlating effects between panel time points have to be taken into account. There is the possibility of confounding factors and dependence between the initial decision to take treatment and the outcome. Certain traits may influence treatment intake as well as the outcomes. The objective is then to identify the isolated effect of the treatment on the outcome sequence. Confounding is a well known phenomenon especially in economics, which has spawned variety of modelling approaches. In a Bayesian approach two models are used, the Shared Factor model and Switching Regression model. The structure of the R-program, how it can be used and some computational aspects are also explained. The estimation results are accomplished by MCMC methods. Additionally,
the popular technique of spike & slab priors is used for variable selection within the Bayesian framework, which can be utilized in the drawing steps of the MCMC sampler. In Chapter 2 the two models are explained in detail and a theoretical background is provided. It contains an exposition of the model assumptions, some aspects of Bayesian statistics and posterior inference, the Markov Chain Monte Carlo methods and the treatment effects estimation. Chapter 3 discusses several simulations under the modeling assumptions and showcases how the models estimate the effects of influential and non-influential variables. Chapter 4 shows the application of the R-package on two economic data sets. The first is an extract from the Austrian Social Security register, which contains information on mothers in maternal leave. The model results show the effect of long maternity leave on mothers incomes when re-entering labor market, based on Jacobi et al. [2016]. Results are then compared to the original paper, as the estimation process of 'bite' works slightly differently. The second data set contains information on participation in schools athletics programs of US high school students and their subsequent incomes after entering the labor market. This data set has been a popular target of analysis, e.g. by Barron et al. [2000] or Chib and Jacobi [2007]. The findings of Chib and Jacobi [2007] serve as comparison from an economic perspective. In Chapter 5 the use of 'bite' and its structure and some of the computational aspects are discussed. A short overview is provided on profiling and benchmarking the program. Chapter 6 summarizes the results and applications. For the interested reader, the official CRAN documentation, as well as side information on the project can be found in the Appendix.
Chapter 2

Treatment Effects Modeling for Panel Outcomes

Formulating a model for treatment effects can be thought of an extension of past models like the one introduced by Lee [1978]. It utilizes a continuous response variable (outcome), a binary variable which indicates treatment and a formulation of marginal models under both treatment states which tie the independent variables and response variable together. For modeling treatment another marginal model is used and unobserved confounders affect both the response and the treatment.

In the following sections the two modeling approaches provided by ‘bite’, the shared factor model and switching regression model, are described in detail and the core differences are outlined. Mainly, the assumed dependence of the panel outcomes and the treatment, modeled by a subject specific utility are handled in different ways. Adjustments for confounding are made as well. Medical trials or experiments can be planned with control mechanisms like randomization or matching to deal with confounding. When looking at already observed data and a treatment variable which depends on an individual’s choice other mechanisms are used. An unobserved subject utility is assumed to enable correlation structures between treatment and outcomes. The utility is then a normally distributed quantity assumed to direct treatment decisions.
Notation

Let $y_i = \{y_{i1}, \ldots, y_{iT_i}\}$, $y \in \mathbb{R}^{T_i}$ be the observed outcomes of subject $i$. Each subject can have an individual panel length of outcomes, i.e. $T_i \neq T_j$ for $i \neq j$ and $T_i \in \mathbb{N}$ being the highest panel time on subject $i$. In the context of only being able to observe one of two outcome vectors per subject, i.e. with or without treatment, the outcomes are called potential outcomes. In general, panel times may differ between subjects. To keep expressions as easily understandable as possible, the maximum panel time of individual $i$ $T_i$ is just called $T$. This still means that individuals can have different panel lengths. For the subjects a vector of features $w_{it}$ for each panel time $t \in 1, \ldots, T$ is observed. Summarizing all feature vectors in matrix form yields for subject $i$ $W_i = (w_{i1}, w_{i2}, \ldots, w_{iT})$ were elements may vary or stay unchanged from one panel time to another, i.e. with $t \neq u$, $t, u \in \{1, \ldots, T\}$ comes $w_{i,t,1} = w_{i,u,1}$ or $w_{i,t,1} \neq w_{i,u,1}$. Furthermore a subject specific vector of baseline features $m_i$ is observed, which are used to model subjects treatment decision. Treatment is denoted by $x_i$, in the binary case $x_i \in \{0, 1\}$, $x_i = 1$ means the subject took the treatment.

ATE

The natural way of calculating the treatment effect with a difference in outcomes, with and without treatment, is not possible since only one outcome series is observed. The potential outcome series of subject $i$ is defined as $y_i = y_{i,x_i=0}(1 - x_i) + y_{i,x_i=1}x_i$ which makes identification of an effect on individual level $y_{i,x_i=1} - y_{i,x_i=0}|W_i$ impossible. Therefore focus is on the average treatment effect

$$ATE(W_i) = E(y_{i,x_i=1}) - E(y_{i,x_i=0})$$

which is the expected difference of outcomes under and without treatment given a certain matrix of subject characteristics $W_i$. The estimation of the average treatment effects on subgroups of the treated and untreated subjects are possible along the procedure of estimating the overall ATE. The treatment effect on the treated (TT) and untreated (TU) are defined as

$$TT(W, M) = E(y_1 - y_0|W, M, x = 1) = E(y_1|W, M, x = 1) - E(y_0|W, M, x = 1)$$
$$TU(W, M) = E(y_1 - y_0|W, M, x = 0) = E(y_1|W, M, x = 0) - E(y_0|W, M, x = 0)$$

They also depend on the baseline covariates $M$. 7
Modeling Mean and General Dependence Structure

It is of interest to examine the joint distribution of subject decisions and the potential outcomes. A key property is the definition of the dependence structure of treatment and outcomes. The subject specific treatment choice is modeled based on baseline features and expressed in a regression equation. The choice- or selection model is specified via the probit model

\[ x_i^* = m_i' \alpha + \eta_i \]  

where \( x_i^* \) is the latent utility assumed for each subject, \( m_i \) is the baseline feature vector of subject \( i \), \( \alpha \) the coefficient vector. The probit assumption yield a normal distribution of the utility \( x_i^* \), the error term is then \( \eta_i \sim N(0, \sigma_i^2) \).

The potential outcome vectors are then assumed to follow the regression equations

\[ y_{0,i} = 1_T \mu + W_i \gamma + \varepsilon_{0,i} \]  
\[ y_{1,i} = 1_T (\mu + \kappa) + W_i (\gamma + \theta) + \varepsilon_{1,i} \]
\[ \varepsilon_{j,i} \sim N(0, \Sigma_j) \]  

This notation captures the overall structural mean with \( \mu \), the weighted influence of covariates on outcomes with \( \gamma \), the general effect of treatment on the outcomes with \( \kappa \) and a heterogeneous effect on treatment with \( \theta \), which is dependent on \( W_i \). The effects \( \kappa \) and \( \theta \) are additive on the base potential outcomes \( y_{0,i} \) without treatment. This formulation allows for all covariate effects to change when switching treatment. For some cases it might be preferable to restrict some covariates to have a common effect on both potential outcomes, i.e. its coefficient value does not depend on treatment. For common effects the corresponding elements in \( \theta \) would be fixed to 0. In general, all parameters can be subject to variable selection, as is later discussed in Section 2.5. Variable selection mechanisms are used to exclude effects from covariates that are negligible. The expression of dependence between utility and the potential outcomes series relies on the joint distribution of the outcome errors \( \varepsilon_{j,i} \) and the selection equation error \( \eta_i \). Depending on the modeling strategy \( \varepsilon_{j,i} \) can be expressed differently and are therefore called composite errors, since they are not iid Normal. The joint distribution of the errors are defined as
When looking at the composite errors marginally, they follow the distribution $\varepsilon_{j,i} \sim \mathcal{N}_T(0, \Sigma_j)$. The covariance matrices $\Sigma_j$ represent the outcome covariance through the panel times under treatment $x_i = j$. As mentioned, only one treatment state is observed at a time for the same subject, so the covariances $\Sigma_{01}$ can never be identified directly. For modeling this specific structure two different methods are used, the Shared Factor approach and the Switching Regression approach.

### 2.1 Shared Factor (SF) Approach

To model the dependence between outcomes and treatment an unobserved factor can be assumed for each subject. This shared factor captures the correlation via its factor loadings. The shared factor as further specification of the errors then yields

$$
\varepsilon_{0i} = \lambda_0 f_i + \epsilon_{0i}, \quad \varepsilon_{0i} \sim \mathcal{N}_T(0, \Sigma_0)
$$

$$
\varepsilon_{1i} = \lambda_1 f_i + \epsilon_{1i}, \quad \varepsilon_{1i} \sim \mathcal{N}_T(0, \Sigma_1)
$$

$$
\eta_i = \lambda_x f_i + \nu_i, \quad \nu_i \sim N(0, 1)
$$

The shared factor $f_i$ is assumed to be $f_i \sim N(0, 1)$ distributed and the outcome errors are $\epsilon_{ji} \sim \mathcal{N}_T(0, \Sigma_j)$. These are assumed to be noise quantities with $\Sigma_j = \text{diag}(\sigma^2_{j1}, ..., \sigma^2_{j,T})$. The outcome errors $\nu_i$ and are mutually independent. To be able to distinguish the source of variance from the component errors, $\text{Var}(f_i) = 1$ and $\text{Var}(\nu_i) = 1$ is assumed, which will not be a relevant restriction. This implies $\text{Var}(\eta_i) = \sigma^2_x = 1 + \lambda^2_x$.

When marginalized over the shared factor $f_i$ the joint distribution of errors has the $2(T + 1)$-variate distribution

$$
\begin{pmatrix}
\varepsilon_{0i} \\
\varepsilon_{1i} \\
\eta_i
\end{pmatrix}
\sim
\mathcal{N}_{2T+1}
\begin{pmatrix}
\Omega_0 + \lambda_0 \lambda_0' \\
\lambda_0 \lambda_1' \\
\omega_0'
\end{pmatrix}
\begin{pmatrix}
\lambda_0 \lambda_0' \\
\Omega_1 + \lambda_1 \lambda_1' \\
\omega_1'
\end{pmatrix}
\begin{pmatrix}
\omega_0 \\
\omega_1 \\
\sigma^2_x
\end{pmatrix}
$$

The dependence between potential outcomes and treatment, represented in the unobserved utility, is given as $\text{Cov}(y_{ji}, x^*_i) = \text{Cov}(\varepsilon_{ji}, \nu_i) = \omega_j = \lambda_x \lambda_j$. 

9
The covariance matrix of the potential outcomes is given as $\Omega_j = \Sigma_j + \lambda_j \lambda_j'$, the correlation in panel outcomes is determined by the $\lambda_j$ loadings in $\varepsilon_{ji}$, which may be interpreted as the influence of an unobserved characteristic on the subject, represented by the shared factor $f_i$. By introducing the error equations above the shared factor approach implies a dependence of the two potential outcome sequences as $\Omega_{01} = \lambda_0 \lambda_1'$. For a standard Normal $f_i$ the likelihood of $(-\lambda_x)(-f_i)$ and $\lambda_x f_i$ will be equal and therefore the signs cannot be identified. This is also the case for $(-\lambda_j)(-f_i)$ and $\lambda_j f_i$.

### 2.2 Switching Regression (SWR) Approach

The switching regression model is based on early formulations of Roy [1951], which has been popularly used in economics. Chib [2007] explained that the specification of the joint distribution of all potential outcomes is not necessary. Chib and Jacobi [2007] and Jacobi et al. [2016] also exploit this to formulate their approach, which is closer described below. The switching regression model also utilizes equations 2.1 and 2.2, but omits the specification of the joint distribution of the outcome errors $\varepsilon_0$ and $\varepsilon_1$ under both treatments, as outcomes on individual level cannot be observed together. It is left to specify $\omega_j$ and $\Omega_j$ to describe the joint distribution of each outcome vector depending on treatment and the latent utility.

First, consider a random intercept term for specification of the error. This approach is further specified as SWR-I. For both treatments $j = 0, 1$ it is specified

$$\varepsilon = 1_T b_{ji} + \varepsilon_{ji}$$

where $b_{ji}$ is the random intercept term, i.e. $b_{ji} \sim N(0, D_j)$ and is independent from the idiosyncratic errors $\varepsilon_{ji} \sim N(0, \Sigma_j)$. $\Sigma_j$ is a $T \times T$ diagonal matrix with the independent panel error variances in the diagonal, i.e. $\text{diag}(\Sigma_j) = (\sigma^2_{j1}, \ldots, \sigma^2_{jT})$. Marginalized over the random intercept $b_{ji}$ the errors $\varepsilon_{ji}$ follow a multivariate Normal distribution with covariance given as
\[
\text{Cov}(\varepsilon_{ji}) = \Sigma_j + D_j \mathbf{1}_T \mathbf{1}_T'
\]

which is given from the independence of the error components. In this model the symmetry structure would suggest a constant dependence of outcomes throughout the panel, which will be rather restrictive for most applications. A more flexible dependence structure can be modeled with different weights or factor loadings on \(b_{ji}\). This will be specified as SWR-F approach. SWR-F will be used for estimation later as SWR-I is just a special decomposition of errors of the more general SWR-F and with that is more restrictive on the dependence. Consider subject and treatment specific factors \(\tilde{b}_{ji} \sim N(0,1)\) with factor loadings \(\lambda_j = (\lambda_{j,1}, \ldots, \lambda_{j,T})\) which yields the equation

\[
\varepsilon_{ji} = \lambda_j \tilde{b}_{ji} + \epsilon_{ji}.
\]

As above, the factors \(\tilde{b}_{ji}\) are assumed to be independent from the errors \(\epsilon_{ji}\). The errors follow the same distribution, i.e. \(\epsilon_{ji} \sim N(0, \Sigma_j)\), but marginalizing over the factor yields a different outcome covariance structure.

\[
\text{Cov}(\varepsilon_{ji}) = \Sigma_j + \lambda_j \lambda_j'
\]

The SWR-F model therefore allows for time-dependent covariances entries, but also has more parameters to estimate. The SWR-I model can be viewed as a special case of the flexible SWR-F, i.e. for \(\lambda = \sqrt{D_j} \mathbf{1}_T\). Chib [2007] has shown that the specification of the joint distribution of just the treatment and one outcome sequence at a time is sufficient, if goal is the identification of the average treatment effect. This leads to the following joint distribution of the pure error and the treatment error:
\[
\begin{pmatrix}
\epsilon_{0i} \\
\eta_{0i}
\end{pmatrix} \sim N_{T_i+1} \left( 0, \begin{pmatrix}
\Omega_0 & \omega_0 \\
\omega_0' & \sigma_x^2
\end{pmatrix} \right)
\]

and
\[
\begin{pmatrix}
\epsilon_{1i} \\
\eta_{1i}
\end{pmatrix} \sim N_{T_i+1} \left( 0, \begin{pmatrix}
\Omega_1 & \omega_1 \\
\omega_1' & \sigma_x^2
\end{pmatrix} \right).
\]

(2.9)

The variance of \(\eta_i\) is restricted to be \(\text{Var}(\eta_i) = \sigma_x^2 = 1\), based on the argument of variance identification. The covariance \(\text{Cov}(\epsilon_{ji}) = \omega_j = \Sigma_j^{1/2} \rho_j\) implies a correlation of \(\rho_j = \text{Cor}(\epsilon_{ji}, \eta_i)\). There are no further explicit structural assumptions on \(\rho_j = (\rho_{j,1}, \ldots, \rho_{j,T})\) except for the positive definiteness of the covariance matrix of the \(T+1\) variate Normal distribution given in (2.9) and (2.10).

The main difference of the two models SWR and SF is therefore the dependence structure of outcomes and treatment. For both models the marginal correlation between \(x^*_i\) and \(y_{j, it}\) can be written as

\[
\text{Cor}(y_{j, it}, x^*_i) = \frac{\text{Cov}(y_{j, it}, x^*_i)}{\text{sd}(x^*_i)\text{sd}(y_{j, it})} = \frac{\omega_{j, t}}{\sigma_x \sqrt{\Omega_{j,[t]}}}
\]

(2.11)

where \([tt]\) denotes the diagonal element at position \(t\) in the covariance matrix of the outcomes \(\Omega_j\).

### 2.3 Data Structure Implications

There are a few implications and assumptions connected to the models. The potential outcome series \(y_{0i}\) is only observed when \(x_i = 0\), which in the utility modeling corresponds to \(x^*_i < 0\) and \(y_{1i}\) is observed only for \(x_i = 1\), which corresponds to \(x^*_i > 0\). The utility is an intricate model element and therefore not observable. This means the observed treatment \(x_i\) restricts the utility to either \(I_0 = (-\infty, 0]\) if \(x_i = 0\) or \(I_1 = (0, \infty)\) if \(x_i = 1\). The joint vector of outcomes and utility then only exist in a subspace of \(\mathbb{R}^{T+1}\), which for a positive utility is \(\mathbb{R} \times I_1\) and for a negative utility is \(\mathbb{R} \times I_0\). The case \(x^*_i = 0\) technically has a 0 probability, but for completion is above included in treatment \(x_i = 0\). The distribution of \(x^*_i\) when conditioned on the potential outcome \(y_{ji}\) is given by \(x^*_i | y_{ji} \sim N(\tilde{\mu}_{ji}, \tilde{\sigma}_j^2)\) with the moments

\[
\tilde{\mu}_{ji} = \mu(x^*_i) + \omega_j \Omega_j^{-1}(y_{ji} - \mu(y_{ji})), \quad \tilde{\sigma}_j^2 = \sigma_x^2 - \omega_j' \Omega_j^{-1} \omega_j,
\]
where $\mu(x^*_i)$ and $\mu(y_{ji})$ are the conditional means given a certain covariate vector. Like above, the mean in utility is then denoted by

$$
\mu(x^*_i) = E(x^*_i|m'_i) = m'_i \alpha
$$

and the mean in outcomes for treatment $x_i = 0$ is given as

$$
\mu(y_{0i}) = E(y_{0i}|W_i) = 1_T \mu + W_i \gamma
$$

and for treatment $x_i = 1$ is given as

$$
\mu(y_{1i}) = E(y_{1i}|W_i) = 1_T(\mu + \kappa) + W_i (\gamma + \theta)
$$

in accordance with the models above.

The joint distributions of the outcomes and the treatment, depending on the treatment observed are

$$
p(y_{0i}, x_i = 0) = p(y_{0i}) \int_{-\infty}^{0} p(x^*_i|y_{0i})dx^*_i = p(y_{0i})(\Phi\left(\frac{x^*_i - \tilde{\mu}_{0i}}{\tilde{\sigma}_{0}}\right))
$$

and

$$
p(y_{1i}, x_i = 1) = p(y_{1i}) \int_{0}^{\infty} p(x^*_i|y_{1i})dx^*_i = p(y_{1i})(1 - \Phi\left(\frac{x^*_i - \tilde{\mu}_{1i}}{\tilde{\sigma}_{1}}\right))
$$

where $\Phi(.)$ denotes the cdf of a standard normal distribution and $y_{ji}$ is marginally distributed as $y_{ji} \sim N_T(\mu(y_{ji}), \Sigma_j)$. With the 2 parts for each realization of the treatment $x_i \in \{0, 1\}$ an overall data-generating distribution for subject $i$ is specified by

$$
p(y_i, x_i) = p(y_{0i}, x_i = 0)I(x_i = 0) + p(y_{0i}, x_i = 1)I(x_i = 1)
$$

The probability distributions of the observed outcome sequences conditioned on the treatment can be specified with the marginal distribution of treatment $p(x_i)$ from the probit specification in (2.1)

$$
p(y_i|x_i) = \frac{p(y_i, x_i)}{p(x_i)} = \begin{cases} p(y_{0i}) \frac{1-\Phi(\tilde{\mu}_{0i}/\tilde{\sigma}_{0})}{1-\Phi(\mu(x^*_i)/\sigma_x)}, & x_i = 0 \\ p(y_{1i}) \frac{\Phi(\tilde{\mu}_{1i}/\tilde{\sigma}_{1})}{\Phi(\mu(x^*_i)/\sigma_x)}, & x_i = 1 \end{cases}
$$

which is a function of the observed outcome sequence $y_i$, when conditioned on $x_i$ and the conditional mean $\tilde{\mu}_{x_i,i}$. 

13
2.4 Bayesian Inference on Treatment Models

For all models a Bayesian modeling framework is applied, model parameters are estimated under certain prior distributions. Bayesian inference is extended to offer variable selection on all parameters in the regression models, which should provide a method to assess their importance. The posterior inference via MCMC algorithm is also given in more detail.

The structural mean in the treatment selection equation is given as \( \mu(x_i^*) = m_i'\alpha \), where \( m_i = (m_{i1}, ..., m_{i\alpha}) \) is the subject specific feature vector, \( \alpha \) is the corresponding coefficient vector of length \( \alpha \).

The structural mean of outcomes with treatment state \( j \) is denoted \( \mu(y_{ji}) = W_{ji}\beta \), where

\[
W_{ji} = \begin{cases} 
(1_T W_i 0_T 0), & \text{for } j = 0, \\
(1_T W_i 1_T W_i), & \text{for } j = 1 
\end{cases}
\]

denotes a covariate matrix which is subject specific and records features in panel periods, i.e. is of dimension \( T \times d_\beta \). Then the subject specific outcome vector \( \beta = (\mu, \gamma, \kappa, \theta) \) is of length \( d_\beta \) the , see (2.9) and (2.10).

Priors

For a full Bayesian model approach all model parameters have to be assigned a prior distribution. First a joint prior for model classes \( M \in \{SWR, SF\} \) is given by \( p^M(\Theta^M) \), where \( \Theta^M \) refers to the set of all parameters of model \( M \).

\[
p^{SWR}(\Theta^{SWR}) = p(\beta)p(\alpha)\prod_{j=0}^{1}p^{SWR}(\sigma_j)p^{SWR}(\lambda_j)p^{SWR}(\rho_j) \quad (2.13)
\]

\[
p^{SF}(\Theta^{SF}) = p(\beta)p(\alpha)p^{SF}(\lambda_x)\prod_{j=0}^{1}p^{SF}(\sigma_j)p^{SF}(\lambda_j) \quad (2.14)
\]

The \( \alpha \) parameters from treatment selection equation and \( \beta \) parameters from the outcome equations are assigned a standard Normal prior. In the shared factor case, with the property that conditioning on the shared factor \( f_i \), \( x_i^* \) and \( y_{ji} \) are independent, a prior can be specified that is conditionally conjugate for all parameters. For the pure error variances \( \text{Var}(\epsilon_{j,t}) = \sigma_{j,t}^2 \) independent inverse Gamma distributions are chosen, such that \( \sigma_{j,t}^2 \sim G^{-1}(s_{0,j,t}, S_{0,j,t}) \) and for the factor loadings a Normal prior is chosen,
such that $\lambda_x \sim N(l_x, L_x)$ and $\lambda_j \sim N_T(l_{j0}, L_{j0})$.

For the SWR model the covariance structure in the joint distribution is specified in (2.9) and (2.10). Calculation of the Cholesky decomposition of the covariance matrix yields the lower triangular matrix

$$G_j = \begin{pmatrix}
\Sigma_j^{1/2} \\
\rho_j' \\
1 - \sum_{t=1}^{T} \rho_{j,t}^2 \end{pmatrix}^{1/2}.
$$

By ensuring positive definiteness of the Cholesky triangular, the positive definiteness of the covariance matrix is also ensured, which is given when $1 - \sum_{t=1}^{T} \rho_{j,t}^2 > 0$.

The $\rho$ parameters are assigned a T-variate Normal prior $N_T(r_0, R_0)$ truncated to the subspace which allows for a positive definite $G_j$. A T-variate Normal prior $N_T(c_0, C_0)$ is used for the log-transformed variances $\ln(\sigma_j)$, which follows Chib and Jacobi [2007].

For the factor loadings a Normal prior is chosen $\lambda_j \sim N_T(l_{j0}, L_{j0})$.

**Spike and Slab Priors**

An important task when building regression models is to only include influential regressors into the final model, especially when a large set of potential covariates are available. Omitting regressors with a true non-zero effect in the regression context will lead to bias in the results, while inclusion of a real zero-effect regressor will introduce larger variance and lead to worse generalizability of the model and to worse predictions on new data. A variable selection mechanism can be incorporated into the estimation process. In a Bayesian context variable selection can be performed by assigning a mixture prior distribution with a spike and a slab component. The slab component is classically a rather flat distribution characterizing the natural space of the coefficient. The spike component is designed to shrink the parameter to 0 if it has no real influence on the dependent variable. In the regression equation context the indicator vectors $\delta = (\delta_1, ..., \delta_d)$ and $\delta = (\delta_1, ..., \delta_d)$ are used for selection and outcome coefficients, where $\delta_d = 1$ if the d\textsuperscript{th} covariate is assigned to the slab component and $\delta_d = 0$ if the d\textsuperscript{th} covariate is assigned to the spike component. The priors of regression coefficients assigned to the spike component are considered mutually independent and to be a priori independent of the priors of regression effects assigned to the slab component. Then the joint prior of the regressors $\alpha = (\alpha_1, ..., \alpha_d)$ can be written as
\[ p(\alpha|\delta) = \prod_{j: \delta_j = 1} p_{\text{slab}}(\alpha_j) \prod_{j: \delta_j = 0} p_{\text{spike}}(\alpha_j) \]

The prior inclusion probabilities follow a hierarchical model

\[ p(\delta_j|\omega_j) = \omega_j, \quad \omega_j \sim \mathcal{B}(a_{\omega_j}, b_{\omega_j}). \]

When conditioned on the prior inclusion probabilities \( \omega \) the \( \delta_j \) are independent, but marginally are dependent. Prior information on single effects could also be included, which would lead to different inclusion probabilities. The extreme case would be to force inclusion of the \( j^\text{th} \) effect, i.e. \( p(\delta_j = 1) = 1 \) and not drawing \( \omega \) from a Beta distribution.

For the regression parameters a Dirac Spike prior is used, i.e. \( p_{\text{spike}}(\alpha_j) = p(\alpha_j|\delta_j = 0) = \Delta_0(\alpha_j) \) which is point mass at 0. There is a variety of continuous functions that could be used for spike priors to employ a shrinking mechanic, some are further examined in e.g. Malsiner-Walli and Wagner [2011].

Posterior inference can be done using MCMC sampling of the model parameters. For the Dirac spike \( \delta_j = 0 \) would imply \( \alpha_j = 0 \) and vice versa. To avoid a reducible Markov chain the marginal posterior \( p(\delta|y) \propto p(y|\delta)p(\delta) \) is used, where effects under selection are integrated out. The likelihood \( p(y|\delta) \) denotes the likelihood of a regression model, where only non-zero effects are included.

## 2.5 Posterior Inference

### MCMC Sampling Scheme

For both models a Gibbs sampling algorithm is employed to accumulate samples from the posterior distribution of the parameters.

### Shared Factor Model

In the shared factor model the error terms of the underlying regressions are independent and therefore the joint likelihood of parameters and including the outcomes and the latent utility can be written as the product
\[
p(x, x^*, y | \Theta^{SF}, f) = \prod_{i=1}^{n} p(x_i, x^*_i | \alpha, \lambda_x, f_i)
\cdot \prod_{i: x_i = 0} p(y_{0i} | \beta, \Sigma_0, f_i) \prod_{i: x_i = 1} p(y_{1i} | \beta, \Sigma_1, \lambda_1, f_i)
\]  
\tag{2.15}
\]

where \( \Theta^{SF} \) denotes the set of all parameters in the shared factor model. When conditioning on the shared factors, the models for potential outcomes and the utility are just regression models with an additional covariate \( f_i \), with \( \lambda_j \) and \( \lambda_x \) effects. Sampling of the parameters \( (\alpha, \lambda_x) \) will be performed in the same block, as well as \( (\beta, \lambda) \). In the beginning, sensible starting points are chosen by using 0.5 for all mixture weight priors, i.e. initial inclusion probabilities. Estimation of the regression coefficients is done via standard probit model for the \( \alpha \)-coefficients of the treatment selection equation and a OLS estimator for the coefficients of the outcome equations.

From then, the sampling routine continues as follows:

1. Sample the variances of idiosyncratic errors \( \text{Var}(\epsilon_j) = \sigma^2_j \), as in equation 2.5, from an inverse Gamma distribution \( G^{-1}(s_{n,jt}, S_{n,jt}) \) where

\[
s_{n,jt} = s_{0,jt} + n_j/2, \quad S_{n,jt} = S_{0,jt} + S\epsilon_{jt}/2
\]

and

\[
S\epsilon_{jt} = \sum_{i: x_i = j} (y_{j,it} - W_{j,it}\beta - f_i \lambda_{jt})^2
\]

\( n_j \) is the number of subjects with \( x_i = j \), \( W_{j,it} \) are the covariate values at panel time \( t \).

2. For \( i = 1, \ldots, n \) sample the shared factor \( f_i \) from the full conditional posterior

\[
p(f_i | \Theta^{SF}, x^*_i, y_{x,i}) \propto p(x^*_i, y_{x,i} | \Theta^{SF}, f_i) p(f_i)
\]

which because of the conjugate nature of prior and likelihood is a Normal distribution \( N(f_{n,i}, F_{n,i}) \) with moments

\[
F_{n,i} = \frac{1}{1 + \lambda_x^2 + \sum_{t=1}^{T} \frac{\lambda_{j,t}^2}{\sigma_{j,t}^2}}
\]

\[
f_{n,i} = F_{n,i} \left( \frac{\lambda_{j,1}}{\sigma_{j,1}}, \ldots, \frac{\lambda_{j,T}}{\sigma_{j,T}} \right) \left( x^*_i - Z_i \alpha \right)
\]  
\tag{2.16}
\]

\[
F_{n,i} = \frac{1}{1 + \lambda_x^2 + \sum_{t=1}^{T} \frac{\lambda_{j,t}^2}{\sigma_{j,t}^2}}
\]

\[
f_{n,i} = F_{n,i} \left( \frac{\lambda_{j,1}}{\sigma_{j,1}}, \ldots, \frac{\lambda_{j,T}}{\sigma_{j,T}} \right) \left( y_{ji} - W_{ji}\beta \right)
\]  
\tag{2.17}
\]

\[17\]
3. For all subjects $i = 1, \ldots, n$ the latent utility $x^*_i$ is sampled from $N(Z_i\alpha + \lambda_x f_i, 1)$ which is truncated to the respective interval $I_{x_i}$, dependent on observed the observed treatment $x_i$.

4. Sampling of the regression parameters $(\alpha, \lambda_x)$ (as $\lambda_x$ is part of the compound error) in the treatment selection model

$$x^*_i = m_i'\alpha + f_i\lambda_x + \nu_i, \quad \nu_i \sim N(0, 1)$$

The regression coefficients $\alpha$ are subject to variable selection, $\lambda_x$ is always included. For selection on $\alpha$ a vector $v$ of indicators is sampled from a Beta distribution.

5. Sampling of the regression parameters $\beta, \lambda_0, \lambda_1$ in the outcome models for observed $y_{x_i,i}, i = 1, \ldots, n$

$$y_{x_i,i} = W_{x_i,i}\beta + f_i\lambda_x + e_{x_i,i}, \quad \sim N(0, \Sigma_{x_i})$$

The regression coefficients are subject to variable selection, sampling of a vector of inclusion indicators $\delta$.

6. Sampling of the sign of shared factor $f_i$ and loadings $\lambda_0, \lambda_1, \lambda_x$ due to unidentifiability by sampling $\xi$, $P(\xi = 1) = P(\xi = -1) = 0.5$, then set $f^\text{new} = \xi f$, $\lambda^\text{new}_x = \xi \lambda_x$, $\lambda^\text{new}_j = \xi \lambda_j$ and use the new values in the Markov chain.

7. Sampling the updated inclusion probabilities $\pi_\alpha$ from $B(1 + k_\alpha, 1 + d_\alpha - k_\alpha)$ and $\pi_\beta$ from $B(1 + k_\beta, 1 + d_\beta - k_\beta)$ where $k_\alpha = \sum \nu_i$ is the number of selected regressors on the latent utility and $k_\beta = \sum \delta_i$ is the number of selected regressors on the outcomes.

**Switching Regression Model**

For the SWR model the MCMC scheme is conducted as follows:

1. First sample the latent utility $x^*_i$ for all subjects $i \in 1, \ldots, n$ from the conditional Normal posterior, truncated to the respective positive or negative interval $I_{x_i}$.

2. Sample the indicator variables and regression effects
(a) Sample \((\mathbf{\nu}, \mathbf{\delta})\) and \((\mathbf{\alpha}, \mathbf{\beta})\) from the respective joint regression posterior under consideration of spike and slab selection, i.e. the respective inclusion probabilities \(\pi_\alpha\) and \(\pi_\beta\).

(b) For all \(n\) subjects sample the latent factor \(\tilde{b}_{ji}\) from the conditional Normal posterior \(N(\tilde{h}_i, \tilde{H}_i)\) with treatment dependent moments:

\[
\tilde{H}_i = (1 + \mathbf{\lambda}_j^\prime (\Sigma_j - \mathbf{\omega}_j^\prime)^{-1} \mathbf{\lambda}_j)^{-1},
\]

\[
\tilde{h}_i = \tilde{H}_i \mathbf{\lambda}_j^\prime (\Sigma_j - \mathbf{\omega}_j^\prime)^{-1} \tilde{y}_i
\]

3. For both treatments \(i = 0, 1\) sample factor loadings \(\mathbf{\lambda}_j\) from the treatment specific multivariate Normal \(N(l_j, L_j)\), where

\[
L_j = \left( \sum_{i:x_i = j} \tilde{b}_{ji}^2 (\Sigma_j - \mathbf{\omega}_j^\prime)^{-1} + L_{j0}^{-1} \right)^{-1},
\]

\[
l_j = L_j \left( \sum_{i:x_i = j} \tilde{b}_{ji} (\Sigma_j - \mathbf{\omega}_j^\prime)^{-1} \tilde{y}_i + L_{j0}^{-1} l_{j0} \right)
\]

Perform a random sign switch on factors \(b\) and loadings \(\lambda\): draw \(\xi\) for \(j = 0, 1\) with \(P(\xi_j = -1) = P(\xi_j = 1) = 0.5\). Then set \(\lambda_j^{\text{new}} = \xi_j \lambda_j\) and \(\tilde{b}_{ji}^{\text{new}} = \tilde{b}_{ji} \xi_j\) for all \(i\), where \(x_i = j\) and use the \(b^{\text{new}}, \lambda_j^{\text{new}}, \lambda_i^{\text{new}}\) as updated values in the Markov chain.

4. Sample the log-transformed variances \(\log(\sigma_{j,t}^2)\) for \(j = 0, 1\) and \(t = 1, \ldots, T\) from a conditional distribution \(p(\log(\sigma_{j,t}^2|\Theta_{SRI}, b, x, x^*, y))\). Updates in MCMC process are done in random order of panel time 1, \(\ldots, T\).

5. Sample the correlation parameter \(\rho_{j,t}\) for \(j = 0, 1\) and \(t = 1, \ldots, T\) from a conditional distribution \(p(\log(\rho_{j,t}|\Theta_{SRI}, b, x, x^*, y))\). Updates in MCMC process are done in random order of panel time 1, \(\ldots, T\).

6. Sample the updated inclusion probabilities \(\pi_\alpha\) from \(B(1 + k_\alpha, 1 + d_\alpha + k_\alpha)\) and \(\pi_\beta\) from \(B(1 + k_\beta, 1 + d_\beta + k_\beta)\), where \(k_\alpha = \sum \nu_l\) is the number of selected regression coefficients in the treatment selection model (utility) and \(k_\beta = \sum \delta_l\) is the number of selected regression coefficients in the outcome models.

For steps 2 and 3 the involved subjects are only of treatment group \(j\). The MH-algorithm employed uses as proposal a \(t\)-distribution with 10 degrees of freedom and
distribution parameters used from a Likelihood maximization routine. For maximization a nonlinear programming algorithm from the package ‘nloptr’ Ypma et al. [2018] is used. Within a reasonable number of iterations a candidate value for the maximum of the likelihood can be obtained. The candidate values are points of high likelihood and not the exact maxima, which would take a long time to compute. For a sufficient number of MCMC iterations (e.g. 10000, like used in the applications) this does not affect the results in a negative way. For the $\mathbf{\rho}$ parameters the proposal distribution is further restricted to a region of stationarity, for purposes of regularity, to grant a positive definite covariance. The distribution is truncated to the interval $(-\sqrt{0.999 - \sum_{t\neq t^*} \rho_{t,j}^2}, \sqrt{0.999 - \sum_{t\neq t^*} \rho_{t,j}^2})$. 


Chapter 3

Simulation Studies

This chapter presents a simulation study to underline the functionality of the sampling algorithm described in Chapter 2. The program should be able to identify model parameters correctly and distinguish influential from uninfluential covariates. It is also examined how a misspecification of the correlation structure impacts the results, e.g. choosing the SF model to estimate the parameters on the SWR data set and vice versa. As mentioned in Chapter 2 the dependence structures assumed by the models are quite different.

3.1 Simulation Details

For the simulation study two different data sets are created, one from the SF model and one from the SWR model. Each of them contains a total of 5000 subjects with 4 panel outcome observations each. The proportion of treated subjects is roughly 50%. For modeling the treatment selection we consider 3 simulated variables, the mean latent utility is then specified as

$$\mu(x_i^t) = \mathbf{m}_i \alpha = \alpha_0 + \alpha_1 z_{i,1} + \alpha_2 z_{i,2} + \alpha_3 z_{i,3}$$

with $z_1$ being standard normally distributed and $z_2$ and $z_3$ being binary variables. For data generation $p(z_2 = 1) = p(z_3 = 1) = 0.5$ and $\alpha = (-0.9, 0.8, 0, 1.5)$ are chosen. Generation of the outcome series is done with $z_1$, $z_2$ and the panel times $t = 2, 3, 4$ as covariates for an effect over time. No influence of $z_3$ is assumed. To define the outcome sequences, as defined in in equations 2.2 and 2.3, the common inter-
cept and covariate effects are \((\mu, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5) = (3.1, 0, 0.1, 0.15, 0.2)\) and the constant and feature dependent (heterogeneous) covariate effects are \((\kappa, \theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (-0.5, 0, 0.2, -0.1, 0, 0.1)\).

The resulting average treatment effects of the 4 panel times of this setting are \((-0.4, -0.5, -0.4, -0.3)\).

In sampling, the variances of pure errors are set to \(\text{Var}(\epsilon_{0,t}) = \sigma_{0,t} = 0.25\) and \(\text{Var}(\epsilon_{1,t}) = \sigma_{1,t} = 1\). For the switching Regression data set a correlation of \(\rho_0 = (0.6, 0.5, 0.4, 0.3)\) and \(\rho_1 = -(0.6, 0.5, 0.4, 0.3)\) is used to imply a weakening dependence over time. The factor loadings are defined to be \(\lambda_0 = (0.4, 0.35, 0.3, 0.25)\) and \(\lambda_1 = (0.7, 0.6, 0.5, 0.4)\) respectively for the SWR data set. For the SF data set the factor loadings are set to \(\lambda_2 = 0.7\) , \(\lambda_0 = (0.6, 0.6, 0.5, 0.5)\) and \(\lambda_0 = (-0.6, -0.6, -0.5, -0.5)\). For the SF model setting the factor loadings implies a full covariance matrix for the potential outcomes and treatment, for the SWR model the additional specification of the correlation parameters is needed (which are independent from the other model parameters). Averaging over the latent factor this gives marginal correlation values of utility and outcomes

\[
\begin{align*}
\text{Cor}(x^*_i, y_{0i})^{SF} &= (0.529, 0.529, 0.513, 0.513) \\
\text{Cor}(x^*_i, y_{1i})^{SF} &= (-0.295, -0.295, -0.256, -0.256) \\
\text{Cor}(x^*_i, y_{0i})^{SWR} &= (0.234, 0.205, 0.171, 0.134) \\
\text{Cor}(x^*_i, y_{1i})^{SWR} &= (-0.492, -0.429, -0.358, -0.279)
\end{align*}
\]

which can be calculated from the correlation equation 2.11.

These specifications are exactly the ones used in Jacobi et al. [2016]. Their findings are used for comparison of the functionality of the models.

### 3.2 Simulation Results

For the simulations overall MCMC iterations are set to 10000, 5000 of which are burn-in phase and therefore posterior sample of size 5000 is used for inference. To reach regions of high posterior density quicker and to lessen computation time, selection on the parameters is avoided until iteration 3000. For the regression coefficients the results are reported in Table 3.1 for the \(\alpha\)-coefficients and in Table 3.2 for the \(\beta\)-coefficients respectively.
The tables show in parallel the true value chosen in the creation of the data sets, and estimation results of both SF and SWR models. The results from the selection model in Table 3.1 show, that both modeling approaches identify the parameters well. Also the misspecification does not impact the results in the selection model severely. It is evident though, that the SWR model performs worse than expected. The process is more unstable and would definitely suffer if too little MCMC iterations are chosen.

This is reflected in the $\beta$-coefficients of the SWR model: in the right-bottom section of Table 3.2 the effect of $t_2$ is only included with 0.71 and 0.62 posterior probability respectively. Also the estimate of the $t_4$ effect is not on point. The slow convergence may result from the chosen optimization algorithm and the resulting sub-optimal maxima suggestions. Even with 10000 iterations the algorithm does not pinpoint the true coefficients, although when repeated it will tend to either under- or overestimation. This leads to the conjecture that the convergence takes unusually long, but no bias is involved. Further aspects to support this are the generally correctly estimated variance parameters and well estimated correlation structures. In general, the models identify the coefficients correctly, also the standard deviations are accurate. The posterior inclusion probabilities and resulting $\delta$-indicators sampled during spike & slab selection indicate importance of variables. The inclusion probabilities are shown in the respective
Table 3.2: Outcome models: β-coefficients of SF and SWR estimation on simulated data sets

<table>
<thead>
<tr>
<th></th>
<th>Real Effects</th>
<th>SF Model</th>
<th>SWR Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(µ, γ) (κ, θ)</td>
<td>(µ, γ)(π) (κ, θ)(π)</td>
<td>(µ, γ)(π) (κ, θ)(π)</td>
</tr>
<tr>
<td><strong>SF data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>µ, κ</td>
<td>3.00 -0.50</td>
<td>2.981(n.a.) -0.503(1.00)</td>
<td>2.937(n.a.) -0.476 (1.00)</td>
</tr>
<tr>
<td>z₁</td>
<td>1.00 0.00</td>
<td>0.997(1.00) 0.001(0.04)</td>
<td>0.984(1.00) 0.013 (0.24)</td>
</tr>
<tr>
<td>z₂</td>
<td>0.00 0.20</td>
<td>0.000(0.03) 0.205(1.00)</td>
<td>0.001(0.04) 0.206 (1.00)</td>
</tr>
<tr>
<td>t₂</td>
<td>0.10 -0.10</td>
<td>0.124(1.00) -0.126(0.99)</td>
<td>0.110(1.00) -0.096 (0.65)</td>
</tr>
<tr>
<td>t₃</td>
<td>0.15 0.00</td>
<td>0.169(1.00) 0.000(0.06)</td>
<td>0.172(1.00) 0.002 (0.08)</td>
</tr>
<tr>
<td>t₄</td>
<td>0.20 0.10</td>
<td>0.205(1.00) 0.104(0.93)</td>
<td>0.203(1.00) 0.096 (0.65)</td>
</tr>
<tr>
<td><strong>SWR data</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>µ, κ</td>
<td>3.00 -0.50</td>
<td>3.027(n.a.) -0.559(1.00)</td>
<td>3.017(n.a.) -0.527 (1.00)</td>
</tr>
<tr>
<td>z₁</td>
<td>1.00 0.00</td>
<td>1.013(1.00) 0.001(0.05)</td>
<td>0.999(1.00) 0.001(0.04)</td>
</tr>
<tr>
<td>z₂</td>
<td>0.00 0.20</td>
<td>0.000(0.01) 0.191(1.00)</td>
<td>-0.000(0.01) 0.193 (1.00)</td>
</tr>
<tr>
<td>t₂</td>
<td>0.10 -0.10</td>
<td>0.064(0.99) 0.000(0.07)</td>
<td>0.040(0.71) -0.035 (0.62)</td>
</tr>
<tr>
<td>t₃</td>
<td>0.15 0.00</td>
<td>0.153(1.00) 0.001(0.03)</td>
<td>0.145(1.00) 0.001 (0.04)</td>
</tr>
<tr>
<td>t₄</td>
<td>0.20 0.10</td>
<td>0.197(1.00) 0.105(0.90)</td>
<td>0.153(1.00) 0.192 (1.00)</td>
</tr>
</tbody>
</table>

The overall intercepts are marked with n.a., since they are not subject to selection and therefore always included. The α-coefficients reported in the SWR estimation of SF data set are the actual returned values from the function. These are different, because of the different assumptions on variance/covariance structures. In terms of the SF model parameters, the SWR model estimates with Var(η) = 1, but estimates a marginal correlation with the outcomes which explains the portion not catched by the resulting coefficients.

For the dependence structure, obviously using the wrong model yields results which are not well comparable. Table 3.3 and 3.4 show only the correct models and the deviations of estimates vs. true values for correlation and factor loadings λ. For the SF model factor loading on the treatment was set to λₓ = 0.7 and was estimated 0.710.

In Table 3.5 the posterior mean of the variance parameters for the panel outcome are shown.
<table>
<thead>
<tr>
<th>t</th>
<th>true $\rho_0$</th>
<th>$\rho_0$ (sd)</th>
<th>true $\rho_1$</th>
<th>$\rho_1$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6</td>
<td>0.660 (0.044)</td>
<td>-0.6</td>
<td>-0.578 (0.032)</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>0.456 (0.047)</td>
<td>-0.5</td>
<td>-0.501 (0.034)</td>
</tr>
<tr>
<td>3</td>
<td>0.4</td>
<td>0.434 (0.036)</td>
<td>-0.4</td>
<td>-0.415 (0.034)</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>0.205 (0.045)</td>
<td>-0.3</td>
<td>-0.359 (0.040)</td>
</tr>
</tbody>
</table>

Table 3.3: Estimated correlation parameters $\rho$ from SWR model on SWR data set

<table>
<thead>
<tr>
<th>t</th>
<th>true $\lambda_0$</th>
<th>$\lambda_0$ (sd)</th>
<th>true $\lambda_1$</th>
<th>$\lambda_1$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4</td>
<td>0.413 (0.014)</td>
<td>0.7</td>
<td>0.678 (0.034)</td>
</tr>
<tr>
<td>2</td>
<td>0.35</td>
<td>0.347 (0.015)</td>
<td>0.6</td>
<td>0.628 (0.033)</td>
</tr>
<tr>
<td>3</td>
<td>0.3</td>
<td>0.296 (0.012)</td>
<td>0.5</td>
<td>0.511 (0.030)</td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
<td>0.250 (0.014)</td>
<td>0.4</td>
<td>0.343 (0.033)</td>
</tr>
</tbody>
</table>

Table 3.4: Estimated factor loadings $\lambda$ from SF model on SF data set
Table 3.5: Estimated variance parameters of panel outcomes from SWR and SF model

<table>
<thead>
<tr>
<th>SF data</th>
<th>true SF model (sd)</th>
<th>SWR model (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{01}$</td>
<td>0.25</td>
<td>0.251 (0.011)</td>
</tr>
<tr>
<td>$\sigma_{02}$</td>
<td>0.25</td>
<td>0.245 (0.010)</td>
</tr>
<tr>
<td>$\sigma_{03}$</td>
<td>0.25</td>
<td>0.272 (0.010)</td>
</tr>
<tr>
<td>$\sigma_{04}$</td>
<td>0.25</td>
<td>0.253 (0.010)</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>1</td>
<td>0.952 (0.042)</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>1</td>
<td>0.984 (0.041)</td>
</tr>
<tr>
<td>$\sigma_{13}$</td>
<td>1</td>
<td>1.048 (0.038)</td>
</tr>
<tr>
<td>$\sigma_{14}$</td>
<td>1</td>
<td>1.024 (0.039)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SWR data</th>
<th>true SF model (sd)</th>
<th>SWR model (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_{01}$</td>
<td>0.25</td>
<td>0.244 (0.011)</td>
</tr>
<tr>
<td>$\sigma_{02}$</td>
<td>0.25</td>
<td>0.251 (0.010)</td>
</tr>
<tr>
<td>$\sigma_{03}$</td>
<td>0.25</td>
<td>0.244 (0.008)</td>
</tr>
<tr>
<td>$\sigma_{04}$</td>
<td>0.25</td>
<td>0.253 (0.008)</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>1</td>
<td>1.010 (0.048)</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>1</td>
<td>1.014 (0.045)</td>
</tr>
<tr>
<td>$\sigma_{13}$</td>
<td>1</td>
<td>0.962 (0.035)</td>
</tr>
<tr>
<td>$\sigma_{14}$</td>
<td>1</td>
<td>1.079 (0.036)</td>
</tr>
</tbody>
</table>
The resulting average treatment effects are shown in Figures 3.1 and 3.2. The results are to be expected and show that the models identify their parameters rather accurate. The vertical lines depict the credibility intervals, the black horizontal line represents the true ATE values. When using the wrong model the results suffer bias, due to the misspecification of the dependence structure. For the SF model 3 of 4 credibility intervals do not overlap the true value, the SWR model has 2 of 4 credibility intervals not overlapping. The SWR model needs more MCMC iterations to converge.

![Average Treatment Effects of SWR Model on Simulated Data](image)

Figure 3.1: ATEs of SWR model on both data sets

In Figure 3.3 and 3.4 a 400 iteration window of the MCMC process around the starting point of variable selection (iteration 3000) is shown. Selection on the regression parameters should force uninfluential coefficients to 0 while sampling the other coefficients around their true value. The graphs reflect the expected results, $\gamma_2$ and $\theta_1, \theta_{13}$ did not have an effect when creating the outcome and are forced to 0 immediately.
When looking at Figures 3.5 and 3.6 the beta coefficients of SWR model on the SWR simulated data set can be seen. The selection mechanism false reduces the effects of panel period \( t_2 \) to 0 more extensively than expected. Table 3.2 shows this in form of smaller posterior inclusion probabilities.
Figure 3.3: Extract of MCMC paths for SF data estimated via SF model, $\beta_0 = (\mu, \gamma)$. Variable selection starts at iteration 3000.
Figure 3.4: Extract of MCMC paths for SF data estimated via SF model, $\beta_1 = (\kappa, \theta)$. Variable selection starts at iteration 3000.
Figure 3.5: Extract of MCMC paths for SWR data estimated via SWR model, $\beta_0 = (\kappa, \gamma)$. Variable selection starts at iteration 3000.
Figure 3.6: Extract of MCMC paths for SWR data estimated via SWR model, $\beta_1 = (\kappa, \theta)$. Variable selection starts at iteration 3000.
Chapter 4

Application on Economic Data

In this chapter an application on two different data sets is done to illustrate the functionalities of the package. The first application will be a re-analysis of the data that Jacobi et al. [2016] used in their treatment modeling. It serves as comparison and reassurance that there are no unexpected results. The second application is an analysis of the NLSY data set, which has been a popular research target, for instance by Chib and Jacobi [2007]. The main point of interest is to find a connection of participation in high school athletics program and subsequent labor market earnings.

4.1 Effect of Long Maternity Leave on Income of Mothers

4.1.1 Background

The data set used stems from the Austrian social security register and contains socio-economic information on 31051 mothers. Background was a policy change in 2001 in Austria which gave mothers a benefit of extended job protection of 2 years and extended parental leave payment of 30 months after birth, as opposed to 18 months until then. Since the policy change was openly announced in August 2001, but included all mothers that gave birth from July 2000, the sample was chosen to consist of mothers around this time. This minimizes effects of mothers which would have delayed the birth to benefit from the policy change and the pregnancy decision can be considered exogenous. The treatment is then defined as taking the long leave, over 18 months, the short leave are
mothers taking a maternal leave of less than 18 months. The goal is to identify the effect of the treatment on the subsequent earnings when re-entering the labor market with methods described before. In Figure 4.1 and 4.2 the distribution of the mothers is exemplary described. Figure 4.1 shows just mothers under treatment with 1, 3, and 6 panel times, which are less frequently observed. Figure 4.2 shows the overall mean log-income for mothers with long and short leave. In Table 4.1 the covariates used are listed, together with the rough distributional description.

![Mean Log-Income on Panel Times](image)

*Figure 4.1: Grouped means of log-income of mothers in all panel periods*
<table>
<thead>
<tr>
<th>outcome model covariates $W$</th>
<th>selection model covariates $m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of children $(1, 2, 3+)$</td>
<td></td>
</tr>
<tr>
<td>job experience $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>white/blue collar $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>interaction blue collar * experience $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>base income above 25% quantile $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>base income above 50% quantile $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>base income above 75% quantile $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>return to same employer $(0,1)$</td>
<td>policy change $(0,1)$</td>
</tr>
<tr>
<td>$t = 2$ $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>$t = 3$ $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>$t = 4$ $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>$t = 5$ $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>$t = 6$ $(0,1)$</td>
<td></td>
</tr>
<tr>
<td>years since 1999 $(\text{integer})$</td>
<td></td>
</tr>
<tr>
<td>years since 1999 squared $(\text{integer})$</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.1: Used covariates in selection and outcome models
4.1.2 Modeling results

In Table 4.2 a descriptive overview of the data is given, reporting the discrepancy in the mean and standard deviation in treated and untreated groups, as well as overall mean and standard deviation. For the earnings a small decrease in the mean with treatment is observed. 80% of mothers in short leave return to the old employer, while only 69% of long leave mothers do the same. In the first year returning to the labor market there is a big decline in income when taking a long leave. The other factors are not severely different between treatment groups, with exception of the leave duration and z, which are directly related to treatment.

To analyse the data set both models are used. MCMC iterations are set to 10000,
with a burnin of 5000 and starting the variable selection on regression coefficients from 2500 onward. In Table 4.3 the estimations of the \( \alpha \)-coefficients under SWR and SF model regimes are presented. Calculations of posterior means are based on 5000 posterior samples, in parenthesis the according standard deviations are given and the posterior inclusion probabilities are in the right-hand column. Truly influential parameters would be considered to have high inclusion probabilities. The effects in the treatment selection model are well identified and comparable to the results of Jacobi et al. [2016]. Estimates under the SF model are more similar to the results of Jacobi et al. [2016] than the respective SWR results. One factor might be the used optimization algorithm.

In Table 4.4 the estimated coefficients in the outcome models are given. The columns ”+ trt 1” refer to the additional effect of the coefficients in case of treatment switch \((\kappa, \theta_1, \theta_2)\), which were described in Equation 2.3. The results are similar to Jacobi et al. [2016]. The SWR model gives slightly different results by estimation of a smaller overall mean and larger coefficients of the panel times. Also the common effect of year is estimated to be lower.

In Figure 4.3 the resulting treatment effects are depicted. Results match the find-
<table>
<thead>
<tr>
<th>covariates</th>
<th></th>
<th>SF model</th>
<th></th>
<th>SWR model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>p.mean (sd)</td>
<td>p.incl.prob.</td>
<td>p.mean (sd)</td>
<td>p.incl.prob.</td>
</tr>
<tr>
<td>intercept</td>
<td>-1.597 (0.033)</td>
<td>n.a.</td>
<td></td>
<td>-1.522 (0.029)</td>
<td>n.a.</td>
</tr>
<tr>
<td>policy</td>
<td>2.899 (0.024)</td>
<td>1.000</td>
<td></td>
<td>2.894 (0.023)</td>
<td>1.000</td>
</tr>
<tr>
<td>nchild:2</td>
<td>0.040 (0.037)</td>
<td>0.591</td>
<td></td>
<td>0.020 (0.030)</td>
<td>0.346</td>
</tr>
<tr>
<td>nchild:3+</td>
<td>-0.010 (0.030)</td>
<td>0.136</td>
<td></td>
<td>-0.009 (0.026)</td>
<td>0.136</td>
</tr>
<tr>
<td>experience</td>
<td>0.102 (0.025)</td>
<td>0.999</td>
<td></td>
<td>0.082 (0.028)</td>
<td>0.967</td>
</tr>
<tr>
<td>blue collar</td>
<td>-0.064 (0.044)</td>
<td>0.347</td>
<td></td>
<td>-0.063 (0.042)</td>
<td>0.747</td>
</tr>
<tr>
<td>inter exp-blue</td>
<td>-0.012 (0.034)</td>
<td>0.138</td>
<td></td>
<td>-0.009 (0.031)</td>
<td>0.107</td>
</tr>
<tr>
<td>base-earn Q2</td>
<td>0.002 (0.010)</td>
<td>0.048</td>
<td></td>
<td>0.001 (0.008)</td>
<td>0.030</td>
</tr>
<tr>
<td>base-earn Q3</td>
<td>-0.001 (0.006)</td>
<td>0.028</td>
<td></td>
<td>-0.000 (0.004)</td>
<td>0.016</td>
</tr>
<tr>
<td>base-earn Q4</td>
<td>-0.150 (0.027)</td>
<td>1.000</td>
<td></td>
<td>-0.160 (0.027)</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 4.3: Treatment selection regression parameters for both models

The resulting treatment effects of the SWR model are surprisingly not far from their result, although the outcome model coefficients differ slightly. The treatment effect in panel time 6 is around 0 and not clearly identified whether positive or negative. In the Appendix in Tables 6.1, 6.2 and 6.3 the variance, covariance and factor loading parameters can be found.
Table 4.4: Results in outcome models: posterior means, standard deviation, posterior inclusion probabilities > 0.5 are marked with a *

<table>
<thead>
<tr>
<th>covariates</th>
<th>SF model</th>
<th>SWR model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>trt 0 pmean (sd)  + trt 1 pmean (sd)</td>
<td>trt 0 pmean (sd)  + trt 1 pmean (sd)</td>
</tr>
<tr>
<td>intercept</td>
<td>9.389 (0.011) -0.009 (0.014)</td>
<td>9.233 (0.012) -0.126* (0.012)</td>
</tr>
<tr>
<td>nchild:2</td>
<td>0.000 (0.001) 0.000 (0.000)</td>
<td>-0.000 (0.001) -0.000 (0.001)</td>
</tr>
<tr>
<td>nchild:3+</td>
<td>0.000 (0.001) 0.001 (0.004)</td>
<td>0.000 (0.001) 0.000 (0.002)</td>
</tr>
<tr>
<td>experience</td>
<td>-0.087* (0.005) 0.000 (0.001)</td>
<td>-0.088* (0.008) 0.005 (0.010)</td>
</tr>
<tr>
<td>blue collar</td>
<td>-0.101* (0.005) 0.000 (0.001)</td>
<td>-0.103* (0.006) 0.000 (0.002)</td>
</tr>
<tr>
<td>inter exp-blue</td>
<td>0.001 (0.006) 0.001 (0.006)</td>
<td>0.001 (0.004) 0.007 (0.013)</td>
</tr>
<tr>
<td>base-earn Q2</td>
<td>0.069* (0.006) 0.000 (0.002)</td>
<td>0.069* (0.006) 0.000 (0.002)</td>
</tr>
<tr>
<td>base-earn Q3</td>
<td>0.297* (0.010) -0.055* (0.012)</td>
<td>0.293* (0.010) -0.049* (0.012)</td>
</tr>
<tr>
<td>base-earn Q4</td>
<td>0.620* (0.009) -0.111* (0.011)</td>
<td>0.615* (0.010) -0.117* (0.012)</td>
</tr>
<tr>
<td>same employer</td>
<td>0.054* (0.005) 0.000 (0.000)</td>
<td>0.051* (0.005) 0.001 (0.003)</td>
</tr>
<tr>
<td>panel time 2</td>
<td>0.099* (0.005) 0.068* (0.005)</td>
<td>0.070* (0.006) 0.095* (0.007)</td>
</tr>
<tr>
<td>panel time 3</td>
<td>0.117* (0.007) 0.106* (0.005)</td>
<td>0.115* (0.007) 0.119* (0.006)</td>
</tr>
<tr>
<td>panel time 4</td>
<td>0.161* (0.010) 0.119* (0.005)</td>
<td>0.242* (0.009) 0.139* (0.007)</td>
</tr>
<tr>
<td>panel time 5</td>
<td>0.215* (0.013) 0.123* (0.006)</td>
<td>0.320* (0.012) 0.143* (0.007)</td>
</tr>
<tr>
<td>panel time 6</td>
<td>0.266* (0.015) 0.140* (0.007)</td>
<td>0.393* (0.014) 0.170* (0.008)</td>
</tr>
</tbody>
</table>

(year − 1999)  0.037 (0.004)  0.019 (0.004)
(year − 1999)^2 -0.004(0.0002) -0.004 (0.002)

Table 4.4: Results in outcome models: posterior means, standard deviation, posterior inclusion probabilities > 0.5 are marked with a *
Figure 4.3:
4.2 Effect of High School Athletics Program on Future Income

On the data set of Austrian mothers the SF model performed as expected, the SWR model seems a bit more unstable and results differ slightly from the original findings of Jacobi et al. [2016]. The National Longitudinal Survey of Youth (NLSY) data is concerned with US high school graduates, previous publications like Barron et al. [2000] and Chib and Jacobi [2007] already showed diverse results on the data set. The package 'bite' may offer a new way to look on the NLSY data. The data set contains 2113 subjects which are examined for the effect of the participation of high school athletics on subsequent labor market earnings. The panel is chosen to be 1989 to 1992, giving \( t = 4 \) panel year earnings for each subject. Higher degree of family education seems to encourage participation, which can be an influential factor on treatment choice and earnings. Confounders identified by Barron are low returns of human capital, general high capability and a lower preference for idleness. Average weekly incomes range from USD 383.75 in 1989 to USD 468.72 in 1992.

<table>
<thead>
<tr>
<th>outcome covariates ( W )</th>
<th>selection covariates ( m )</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender: male (binary)</td>
<td>enrollment (numeric)</td>
</tr>
<tr>
<td>ethnicity: black (binary)</td>
<td>bad health (binary)</td>
</tr>
<tr>
<td>parents sec. school (binary)</td>
<td>height (numeric)</td>
</tr>
<tr>
<td>parents high school (binary)</td>
<td>weight (numeric)</td>
</tr>
<tr>
<td>parents college (binary)</td>
<td></td>
</tr>
<tr>
<td>parents grad school (binary)</td>
<td></td>
</tr>
<tr>
<td>age (integer)</td>
<td></td>
</tr>
<tr>
<td>education (integer)</td>
<td></td>
</tr>
<tr>
<td>job tenure (numeric)</td>
<td></td>
</tr>
<tr>
<td>married (binary)</td>
<td></td>
</tr>
<tr>
<td>year 1990 (binary)</td>
<td></td>
</tr>
<tr>
<td>year 1991 (binary)</td>
<td></td>
</tr>
<tr>
<td>year 1992 (binary)</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.5: Used covariates in selection and outcome models

In Table 4.5 a overview of variables used in the models is given with their data type.
Independent variables include gender, ethnicity, parents indicators of highest graduation, age (in whole years - 17), education (in total school years), job tenure (in years), marital status, a panel year indicator, enrollment (in student enrollment of school/100), indication of bad health, weight and height.

<table>
<thead>
<tr>
<th>variable</th>
<th>untreated mean</th>
<th>sd</th>
<th>treated mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>log-income</td>
<td>5.92</td>
<td>0.67</td>
<td>6.14</td>
<td>0.74</td>
</tr>
<tr>
<td>age</td>
<td>12.62</td>
<td>2.42</td>
<td>12.52</td>
<td>2.39</td>
</tr>
<tr>
<td>education</td>
<td>13.53</td>
<td>2.00</td>
<td>14.38</td>
<td>2.15</td>
</tr>
<tr>
<td>job tenure</td>
<td>4.23</td>
<td>3.71</td>
<td>3.99</td>
<td>3.50</td>
</tr>
<tr>
<td>gender male</td>
<td>0.45</td>
<td></td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>enrollment</td>
<td>14.60</td>
<td>7.08</td>
<td>12.55</td>
<td>7.12</td>
</tr>
<tr>
<td>height</td>
<td>57.35</td>
<td>3.60</td>
<td>58.50</td>
<td>4.65</td>
</tr>
<tr>
<td>weight</td>
<td>14.38</td>
<td>2.93</td>
<td>15.18</td>
<td>3.16</td>
</tr>
</tbody>
</table>

Table 4.6: Mean and standard deviation of selected variables, grouped by treatment

In Table 4.6 the discrepancy between treated and untreated subjects in a few selected variables is characterized with mean and median. People in treated group are on average slightly younger, do have more educational years, their school’s enrollment size is smaller. Also the male proportion is higher in the treated group with 58% being male and only 45% of the untreated being male.

4.2.1 Modeling Results

The results of SF and SWR modeling will be compared with Chib and Jacobi [2007]. They used a similarly structured model like the SWR model used in 'bite', a main difference is that now spike & slab variable selection is used. They also used a switching regression type modeling approach with exploitation of conditional moments, when marginalizing over the latent factor \( \tilde{b}_{ji} \) as explained in Chapter 2. They used a restricted and an unrestricted approach for estimation, where the unrestricted is similar to the standard SWR approach utilized in 'bite’. Their restricted approach set outcome covariance parameters to a fixed value, as described in the random intercept modeling in SWR, see Section 2.2. The method to compare will be the more flexible approach
Table 4.7: Treatment selection ($\alpha$) regression parameters for both models in upper part with and in lower part without performing variable selection with factor loadings. Also, they only chose 1000 burnin iterations, which worked fine in their simulations, the estimation used with 'bite' will stay with 3000 burnin iterations, followed by 7000 posterior sample draws. In Table 4.7 the coefficients from treatment equations are given. Interestingly, there seems to be more discrepancy in the coefficients when comparing the 2 methods. The parameters still agree in terms of the direction (sign), but the magnitude is estimated differently. For instance male gender falls out of the SWR model completely, while the SF model accepts it with a posterior inclusion probability of nearly 1. The SF model rejects the covariate parents grade of education: some college years (parents scol) as an important variable, while the SWR model barely includes it with probability 0.592. Both methods regard covariates enrollment, and parents graduation status of college and university graduation as important.

In Table 4.8 the results on the $\beta$-coefficients are displayed together with the results of Chib and Jacobi [2007]. When comparing column pairs (2,3) and (6,7) the coefficients are rather similar. The SWR model reduced smaller coefficients to zero, which reflects heavily on the coefficients under treatment. Note, that panel time effects are not explicitly reported in Chib and Jacobi [2007]. There is again a notable difference between SF and SWR model coefficients. The
panel time effects are reversed, SWR giving a negative base effect for non treated and a strong positive effect in earnings when switching to treated. SF identifies a weak overall positive effect over panel times, indifferent when treated. The SWR model reduces almost all coefficients under treatment to near 0, except for intercept and panel time effects. The SF model on the other hand yields no intercept effect under treatment, but an additional positive effect of age under treatment and a negative effect of being male under treatment. Effects that are estimated in consent are education as positive effect, job tenure as positive, being male as positive and being black as negative. Tables 6.4, 6.5 and 6.6 in the Appendix show further parameters of variance, correlations and factor loadings.
### β-coefficients with variable selection

<table>
<thead>
<tr>
<th>covariates</th>
<th>SWR model trt 0 pmean (sd)</th>
<th>+ trt 1 pmean (sd)</th>
<th>SF model trt 0 pmean (sd)</th>
<th>+ trt 1 pmean (sd)</th>
<th>Chib, Jacobi (2007) trt 0 pmean (sd)</th>
<th>+ trt 1 pmean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>4.697 (0.086)</td>
<td>-0.240* (0.058)</td>
<td>4.738 (0.082)</td>
<td>0.011 (0.022)</td>
<td>4.466 (0.131)</td>
<td>-0.073 (0.138)</td>
</tr>
<tr>
<td>age</td>
<td>0.000 (0.001)</td>
<td>0.000 (0.002)</td>
<td>0.000 (0.000)</td>
<td>0.019* (0.005)</td>
<td>0.004 (0.007)</td>
<td>0.012 (0.007)</td>
</tr>
<tr>
<td>education</td>
<td>0.089* (0.005)</td>
<td>-0.000 (0.002)</td>
<td>0.091* (0.005)</td>
<td>-0.000 (0.001)</td>
<td>0.083 (0.008)</td>
<td>0.002 (0.007)</td>
</tr>
<tr>
<td>job tenure</td>
<td>0.034* (0.002)</td>
<td>0.000 (0.000)</td>
<td>0.031* (0.003)</td>
<td>0.000 (0.000)</td>
<td>0.024 (0.003)</td>
<td>-0.003 (0.003)</td>
</tr>
<tr>
<td>married</td>
<td>-0.000 (0.001)</td>
<td>0.000 (0.002)</td>
<td>0.000 (0.002)</td>
<td>-0.000 (0.002)</td>
<td>0.004 (0.018)</td>
<td>0.003 (0.018)</td>
</tr>
<tr>
<td>gender: male</td>
<td>0.370* (0.022)</td>
<td>-0.000 (0.006)</td>
<td>0.463* (0.044)</td>
<td>-0.208* (0.081)</td>
<td>0.361 (0.032)</td>
<td>-0.005 (0.032)</td>
</tr>
<tr>
<td>race: black</td>
<td>-0.158* (0.030)</td>
<td>0.007 (0.028)</td>
<td>-0.176* (0.034)</td>
<td>0.016 (0.047)</td>
<td>-0.170 (0.038)</td>
<td>0.070 (0.040)</td>
</tr>
<tr>
<td>parents hs</td>
<td>0.001 (0.005)</td>
<td>0.001 (0.006)</td>
<td>0.000 (0.004)</td>
<td>0.001 (0.007)</td>
<td>0.070 (0.037)</td>
<td>-0.004 (0.043)</td>
</tr>
<tr>
<td>parents scol</td>
<td>-0.000 (0.004)</td>
<td>-0.001 (0.010)</td>
<td>0.000 (0.003)</td>
<td>-0.000 (0.004)</td>
<td>0.061 (0.053)</td>
<td>-0.043 (0.057)</td>
</tr>
<tr>
<td>parents col</td>
<td>0.012 (0.034)</td>
<td>0.001 (0.014)</td>
<td>0.033 (0.079)</td>
<td>-0.036 (0.100)</td>
<td>0.153 (0.062)</td>
<td>-0.035 (0.057)</td>
</tr>
<tr>
<td>parents grad</td>
<td>-0.000 (0.008)</td>
<td>0.001 (0.013)</td>
<td>-0.000 (0.004)</td>
<td>-0.001 (0.009)</td>
<td>0.000 (0.082)</td>
<td>0.085 (0.065)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>year</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>-0.276* (0.036)</td>
<td>0.700* (0.042)</td>
<td>0.091* (0.017)</td>
<td>0.000 (0.003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>-0.144* (0.053)</td>
<td>0.305* (0.057)</td>
<td>0.149* (0.019)</td>
<td>0.002 (0.012)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td>-0.158* (0.054)</td>
<td>0.327* (0.056)</td>
<td>0.121* (0.019)</td>
<td>-0.000 (0.005)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.8: Results in outcome models ($\beta$): posterior means, standard deviation, posterior inclusion probabilities > 0.5 are marked with *
These results are interesting for two reasons. First, the SWR model showed some problems with slow convergence in the simulation studies, still the procedure identifies $\beta$ effects similar to Chib, Jacobi. Second, the two models seem to estimate different effects, which can be seen when calculating the ATEs. Figure 4.4 shows the average treatment effects on the log-income of SWR and SF models. The SWR model estimates a slightly negative effect on the first panel period, the biggest effect on the second period followed by weaker positive effects in third and fourth periods. The ATEs based on the SF model are rather strong positive effects, but with little panel variation. Treatment has the strongest effect in the first panel period, the weakest in second and from then increasing in third and fourth. The SF results are similar to what Chib and Jacobi [2007] reported. In Figure 4.5 their results are shown. They did use a different approach of predictive ATEs on newly sampled data and reporting the distribution of the predictive ATEs, compared to the in-sample ATEs which 'bite' calculates on the same sample. Also ATEs where calculated on a weekly income scale in USD, where 'bite' estimated the effect on annual log-earnings. The discrepancy of SWR and Chib, Jacobi is surprising and may result from different estimated correlation structure of treatment and outcomes and weaker, penalized covariates from spike & slab sampling.
Figure 4.4: In sample calculated ATEs on annual log-income
Figure 4.5: Points represent ATE median, lines are covering the 5% and 95% quantiles in the predictive ATEs sample
Chapter 5
Computational Aspects of ’bite’

Motivation

This chapter contains a brief demonstration of the R code of ’bite’ and some issues of computational efficiency and run times. R is a popular framework nowadays for statistical computing, which is the reason that ’bite’ is written in R. R has more intuitive properties than other languages like C++ or Java, it offers a live run time environment and sacrifices speed for usability. A part of its slow computations is tied to the implementation of the R language and how code is interpreted. This part can be optimized.

In this chapter an overview of the ’bite’ interface and usage is given. It relates to the ’bite’ package implementation and shows which contact points are made through the program. There are different strategies, depending on the method or matter of computation, like MCMC sampling, regular computations, matrix multiplications and matrix updates (indexing issues).

5.1 Computations with ’bite’

In general, for modeling treatment effects with the ’bite’ package there are a few requirements on the data set used.

- The function call \texttt{bayesTrtEffects()} is the only user interface to ’bite’. That means all relevant parameters to control the process and data are passed on here.
- \texttt{bayesTrtEffects(base\_mat, panel\_mat)} is the minimal call. \texttt{base\_mat} is the baseline data frame where each row represent one subject with unique ID, treatment...
indicator, number of panel observations and any number of covariates used in the treatment model equations (see 2.1). \textit{panel\_mat} is a data frame with the panel information on subjects. Each row contains one panel vector of a subject, including at least ID, an outcome (target) value, the panel time and any variables that are included in the outcome equations (see 2.2, 2.3).

- Further arguments can be given: \texttt{bayesTrtEffects(base\_mat, panel\_mat, type = ”SF”, covars = NULL, mcmc\_control = list(burnin = 1000, select = 500, M = 1000), prior\_control = list(var\_sel = 5, var\_fix = 0.1), control = list(fix\_alpha = FALSE, fix\_beta = FALSE, fix\_sigma = FALSE, fix\_f = FALSE, sort\_data = TRUE, test = FALSE), model\_name = ”model1”, data\_name = ”treatment data set 1”)

- \texttt{type} controls the used modeling type, either ”SF” or ”SWR”

- \texttt{covars} is a list object which can be used to fix covariates, not making them subject to selection, or define common variables which are estimated without separate effects under and without treatment. To give an example, in the application of Austrian mothers it had the structure \texttt{covars = list(x\_fix = rep(0,9), y\_fix = c(rep(0,14), 1, 1), y\_common = c(rep(0,14), 1, 1), y\_nb = 9, y\_np = 5)} where \texttt{x\_fix} and \texttt{y\_fix} define the covariates not under selection. \texttt{y\_common} defines the common covariates.

- \texttt{mcmc\_control} controls the number of MCMC iterations and length of burnin period.

- \texttt{prior\_control} gives starting values to the priors of inclusion probabilities.

- \texttt{control} is used for testing, users would not want to change it.

- \texttt{model\_name} and \texttt{data\_name} are also internal strings used to name the output.

In Figures 5.1 and 5.2 example pictures are shown how the baseline and panel data frame should be structure when used. The column names are important in older versions of the program, the newer version takes the first column as subject ID and the second and third as seen in the figures.
Figure 5.1: Structure of the baseline data frame

<table>
<thead>
<tr>
<th>ID</th>
<th>Ti</th>
<th>trt</th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>-2.69465500</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>-3.03325706</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>-2.12402536</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>-1.14260418</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>-1.92611410</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>-1.37073047</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>4</td>
<td>0</td>
<td>-0.84613595</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>-1.35649924</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>4</td>
<td>0</td>
<td>-1.66161893</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>4</td>
<td>0</td>
<td>-2.89241003</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>4</td>
<td>0</td>
<td>-2.10374746</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>-3.06634923</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>4</td>
<td>0</td>
<td>-1.84671152</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>4</td>
<td>0</td>
<td>-1.19752521</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>4</td>
<td>0</td>
<td>-1.73538503</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>4</td>
<td>0</td>
<td>-1.86086637</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>17</td>
<td>4</td>
<td>0</td>
<td>-0.67826199</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>18</td>
<td>4</td>
<td>0</td>
<td>-1.78938486</td>
<td>0</td>
</tr>
<tr>
<td>ID</td>
<td>panelT</td>
<td>y</td>
<td>V1</td>
<td>V2</td>
<td>t2</td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>-------</td>
<td>--------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-0.696520931</td>
<td>-2.6946550</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-0.327784273</td>
<td>-2.6946550</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-0.500029457</td>
<td>-2.6946550</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>-0.765029482</td>
<td>-2.6946550</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>-0.41507237</td>
<td>-3.0332571</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>-0.164324189</td>
<td>-3.0332571</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>-0.470670852</td>
<td>-3.0332571</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0.406928347</td>
<td>-3.0332571</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>0.877949558</td>
<td>-2.1240254</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>1.099665698</td>
<td>-2.1240254</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>0.246109730</td>
<td>-2.1240254</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>0.507845997</td>
<td>-2.1240254</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>0.337356458</td>
<td>-1.1426042</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>0.577382272</td>
<td>-1.1426042</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>1.036294334</td>
<td>-1.1426042</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>1.790683651</td>
<td>-1.1426042</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>5</td>
<td>-0.182724731</td>
<td>-1.9261141</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>0.608922747</td>
<td>-1.9261141</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 5.2: Structure of the panel data frame
5.1.1 Overall performance

A computational comparison of the Matlab program and the R program is given. The two engines work differently in background, given that R is mostly developed in C and a portion FORTRAN, Matlab is written in Java and C++, but contains many third party contributions and optimization, hence the commercial character of Matlab (for information on Matlab see Natick [2019]). Matlab is highly column oriented and column/matrix computationally optimized, to best emulate a mathematical way of operating. Different tools are used to evaluate the run time for the two statistical softwares. For Matlab timeit() delivers run times for arbitrary functions f(), but not for open code segment and also brings the most accurate results.

For 'bite' the R package microbenchmark (Ulrich [2019]) is used which returns, depending on the operating system, values on nanosecond basis. It automatically repeats the process for a fixed number of runthroughs. The package profvis (Winston Chang [2019]) provides an advanced way of code profiling to identify segments that are inefficient in either run time or use of memory. Code profiling is not 100% accurate, since some factors are accounted to take 0 time, like + operator. They are also inaccurate below a scale of 0.005 seconds. To optimize code and run time within R, in which 'bite' is written, a few steps have been implemented.

- Matrix multiplications are rarely done via the %*% operator, but via crossprod(), which may take more than 10% less time.
- For-loops have been avoided if possible and been replaced by use of R generic apply functions, or beforehand indexing of vectors and matrices, so that a repetitive call of functions or loop counting is not necessary. It is still necessary in some places to iterate through a loop.
- Big Matrices are divided into sub-matrices to decrease the number of rows and columns for indexing and operations. Indexing matrix rows and columns, then overwriting the entries is slow in R.

Since Matlab and R have very similar appearance, syntax and functionalities it is easy to compare them. In Table 5.1 a comparison of the run times in R and Matlab can be found. The measurements were done for 100 MCMC iterations and divided by 10 to give a robust duration of 10 MCMC iterations and to minimize impact of system
Matlab
normal selection

R
normal selection

<table>
<thead>
<tr>
<th></th>
<th>Matlab</th>
<th></th>
<th>R</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SF</td>
<td>0.121</td>
<td>0.193</td>
<td>0.311</td>
<td>1.01</td>
</tr>
<tr>
<td>SWR</td>
<td>6.18</td>
<td>6.31</td>
<td>44.18</td>
<td>44.98</td>
</tr>
</tbody>
</table>

Table 5.1: Mean run times for 10 MCMC iterations in seconds

delay times which may happen in the beginning of the procedure. The procedures where done on the simulated data set of SF model, which is also appended to the 'bite' package. The Matlab code runs faster than the R coding based 'bite' in both modeling approaches. The SF method of the program is optimized to a good point. For SF estimation it is about 2.5 times slower when calculating normal MCMC iterations and about 5 times slower when calculating MCMC iterations with spike & slab selection. The SWR estimation routine of 'bite' is rather slow, making it hard to recommend for daily use at the moment. Both parts can be optimized by implementing techniques like Rcpp.

5.1.2 Code Profiling of 'bite'

The package 'profvis' samples code to benchmarks function calls and also highlights memory allocation as well as de-allocation from one segment of code to the other. The memory allocation can be confusing, the bars shown are not the difference of memory usage at that very line of code. It can also accommodate changes in the previous lines of code, depending on the nature of the call. Still it delivers a neatly arranged overview. A few examples are shown to illustrate the development of the 'bite' code. The versions are the 0.1.0.0 for the first released version on Github and 0.3.0.0 which is the version to this date. This is also the version used to generate the results of previous chapters. The data of mothers parental leave was used to create the profiles below, four MCMC iterations were calculated: two burnin and two saved MCMC iterations with variable selection. Figure 5.3 and 5.4 illustrate how the core function of bayesTrtEffects changed from version 0.1.0.0 to 0.3.0.0. The columns to the right show memory usage (Mb) and run time (ms). The refinement of 'bite' brought a very strong run time improvement, going from about 3 minutes for only 4 iterations to about 4-5 seconds. Memory allocation was also improved, the program needs less input parameters and is
more efficient in storing intermediate results. It also de-allocates less memory, as seen in the negative number at memory. This is a consequence of using less and redundant objects in the environment.
Chapter 6

Conclusion

In this thesis a general approach for modelling of panel treatment effect settings is provided. The R-package 'bite' provides methods of estimating coefficients and treatment effects with the shared factor (SF) model and switching regression (SWR) model. The core difference lies in the assumptions of the dependence structure of treatment and outcomes. For SF model the dependence is directly given through the factor loadings, in the SWR model the co-dependence of the potential outcomes can only be modeled through a marginal correlation, independent of other parameters. It has been shown, that the package is capable of identifying simulation parameters correctly for the SF model. The use of the SWR method suffers from the long computation time and tends to converge less reliably than SF. On the low dimensional simulated data set this is no big concern, but for the economic data applications the results have not been as convincing, though the SWR computations are feasible. On the mothers parental leave data the results are similar to the findings of Jacobi et al. [2016], which is reassuring that the MCMC mechanics used for generation of posterior samples and subsequent inference in 'bite' are working as expected. The SWR method seems to have problems with converging, as well as showing too aggressive behavior in the variable selection. For the NLSY data, the results are counter intuitive, since SWR yields other coefficients in sign and magnitude than Chib and Jacobi [2007]. The SF approach has very dampened effect, not as varying through panel times. This may be a result of the differently specified dependence structure and of the shrinkage mechanics of spike & slab variable selection. Some of the original effects, like being male, black, etc. are preserved in both models. The resulting effects of the SF model are much weaker than the findings of Chib, Jacobi, while the SWR model offers other parameters and treatment
effects altogether. Overall the SF model runs more stable, but may offer a restrictive
dependence structure for some data, like NLSY.
To conclude, a first step of panel treatment effect estimation has been done, the com-
putational details of 'bite' leave room for improvement: The code is currently written
exclusively in R, which can be optimized by translating some scripts to C++. The
optimization steps of the SWR model are unstable, the used algorithm can be replaced
with a more efficient one in future work. The package could also be extended in many
ways, the restriction to binary treatment for example could be lifted.
Appendix

System Basis

Things like computing time and storage efficiency are not only dependent on programming, but will also be affected by the computer hardware in use. This is especially true for the effective core speed of the CPU that runs the operations. The results in Section 5 were achieved on the following build.

- Windows 10-x64, Version 1809
- Intel Core i5 3570k, running at a minimum of 3.5 GHz
- 16 GB DDR3-1600 Ram, running at constantly 1600 MHz
- R version 3.5.3 "Great Truth"
- R Studio version 1.2.1335

Further Results on 4.2 Effect of Long Maternity Leave on Income of Mothers
### Table 6.1: Estimated variance parameters of panel outcomes from SF and SWR model without treatment (0) and with treatment (1) on mothers data

<table>
<thead>
<tr>
<th>t</th>
<th>$\sigma_{0}^{SF}$</th>
<th>$\sigma_{1}^{SF}$</th>
<th>$\sigma_{0}^{SWR}$</th>
<th>$\sigma_{1}^{SWR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.123</td>
<td>0.102</td>
<td>0.145</td>
<td>0.108</td>
</tr>
<tr>
<td>2</td>
<td>0.072</td>
<td>0.043</td>
<td>0.078</td>
<td>0.047</td>
</tr>
<tr>
<td>3</td>
<td>0.044</td>
<td>0.019</td>
<td>0.052</td>
<td>0.033</td>
</tr>
<tr>
<td>4</td>
<td>0.026</td>
<td>0.029</td>
<td>0.041</td>
<td>0.049</td>
</tr>
<tr>
<td>5</td>
<td>0.028</td>
<td>0.045</td>
<td>0.031</td>
<td>0.064</td>
</tr>
<tr>
<td>6</td>
<td>0.046</td>
<td>0.058</td>
<td>0.055</td>
<td>0.068</td>
</tr>
</tbody>
</table>

### Table 6.2: Estimated correlation parameters $\rho$ and factor loadings $\lambda$ from SWR model on mothers data

<table>
<thead>
<tr>
<th>t</th>
<th>$\rho_{0}$ (sd)</th>
<th>$\rho_{1}$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.162 (0.016)</td>
<td>0.309 (0.031)</td>
</tr>
<tr>
<td>2</td>
<td>-0.269 (0.016)</td>
<td>0.031 (0.057)</td>
</tr>
<tr>
<td>3</td>
<td>-0.401 (0.016)</td>
<td>0.429 (0.068)</td>
</tr>
<tr>
<td>4</td>
<td>-0.552 (0.018)</td>
<td>0.245 (0.063)</td>
</tr>
<tr>
<td>5</td>
<td>-0.524 (0.018)</td>
<td>0.260 (0.050)</td>
</tr>
<tr>
<td>6</td>
<td>-0.376 (0.019)</td>
<td>0.129 (0.044)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t</th>
<th>$\lambda_{0}$ (sd)</th>
<th>true $\lambda_{1}$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.340 (0.004)</td>
<td>0.290 (0.003)</td>
</tr>
<tr>
<td>2</td>
<td>0.380 (0.003)</td>
<td>0.353 (0.003)</td>
</tr>
<tr>
<td>3</td>
<td>0.411 (0.003)</td>
<td>0.385 (0.003)</td>
</tr>
<tr>
<td>4</td>
<td>0.427 (0.003)</td>
<td>0.384 (0.003)</td>
</tr>
<tr>
<td>5</td>
<td>0.413 (0.003)</td>
<td>0.366 (0.003)</td>
</tr>
<tr>
<td>6</td>
<td>0.395 (0.003)</td>
<td>0.356 (0.004)</td>
</tr>
</tbody>
</table>

Table 6.2: Estimated correlation parameters $\rho$ and factor loadings $\lambda$ from SWR model on mothers data
The table shows the estimated marginal correlations between utility and outcome series for SF and SWR model on mothers data. The correlations are given for different time points, along with the standard deviations. The table compares the results with those of Jacobi et al. [2016].
Further Results on 4.3 Effect of High School Athletics Program on Future Income

<table>
<thead>
<tr>
<th>t</th>
<th>$\sigma_0^{SF}$</th>
<th>$\sigma_1^{SF}$</th>
<th>$\sigma_0^{SWR}$</th>
<th>$\sigma_1^{SWR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.151</td>
<td>0.135</td>
<td>0.233</td>
<td>0.152</td>
</tr>
<tr>
<td>2</td>
<td>0.114</td>
<td>0.128</td>
<td>0.142</td>
<td>0.262</td>
</tr>
<tr>
<td>3</td>
<td>0.105</td>
<td>0.277</td>
<td>0.088</td>
<td>0.251</td>
</tr>
<tr>
<td>4</td>
<td>0.202</td>
<td>0.144</td>
<td>0.194</td>
<td>0.130</td>
</tr>
</tbody>
</table>

Table 6.4: Estimated variance parameters of potential outcomes from SWR and SF model on NLSY data
Table 6.5: Estimated correlation parameters $\rho$ and factor loadings $\lambda$ from SWR model on mothers data)

<table>
<thead>
<tr>
<th>t</th>
<th>$\rho_0$ (sd)</th>
<th>$\rho_1$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.162 (0.016)</td>
<td>0.309 (0.031)</td>
</tr>
<tr>
<td>2</td>
<td>-0.269 (0.016)</td>
<td>0.031 (0.057)</td>
</tr>
<tr>
<td>3</td>
<td>-0.401 (0.016)</td>
<td>0.429 (0.068)</td>
</tr>
<tr>
<td>4</td>
<td>-0.552 (0.018)</td>
<td>0.245 (0.063)</td>
</tr>
<tr>
<td>5</td>
<td>-0.524 (0.018)</td>
<td>0.260 (0.050)</td>
</tr>
<tr>
<td>6</td>
<td>-0.376 (0.019)</td>
<td>0.129 (0.044)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t</th>
<th>$\lambda_0$ (sd)</th>
<th>true $\lambda_1$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.340 (0.004)</td>
<td>0.290 (0.003)</td>
</tr>
<tr>
<td>2</td>
<td>0.380 (0.003)</td>
<td>0.353 (0.003)</td>
</tr>
<tr>
<td>3</td>
<td>0.411 (0.003)</td>
<td>0.385 (0.003)</td>
</tr>
<tr>
<td>4</td>
<td>0.427 (0.003)</td>
<td>0.384 (0.003)</td>
</tr>
<tr>
<td>5</td>
<td>0.413 (0.003)</td>
<td>0.366 (0.003)</td>
</tr>
<tr>
<td>6</td>
<td>0.395 (0.003)</td>
<td>0.356 (0.004)</td>
</tr>
</tbody>
</table>

Table 6.6: Estimated marginal correlations between utility and outcome series for SF and SWR model on mothers data)

<table>
<thead>
<tr>
<th>t</th>
<th>$SF_0$ (sd)</th>
<th>$SF_1$ (sd)</th>
<th>$SWR_0$ (sd)</th>
<th>$SWR_1$ (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.174 (0.008)</td>
<td>0.169 (0.008)</td>
<td>-0.116 (0.012)</td>
<td>0.229 (0.023)</td>
</tr>
<tr>
<td>2</td>
<td>-0.205 (0.009)</td>
<td>0.216 (0.010)</td>
<td>-0.155 (0.009)</td>
<td>0.016 (0.029)</td>
</tr>
<tr>
<td>3</td>
<td>-0.223 (0.010)</td>
<td>0.236 (0.011)</td>
<td>-0.182 (0.008)</td>
<td>0.146 (0.023)</td>
</tr>
<tr>
<td>4</td>
<td>-0.235 (0.011)</td>
<td>0.230 (0.011)</td>
<td>-0.195 (0.006)</td>
<td>0.098 (0.025)</td>
</tr>
<tr>
<td>5</td>
<td>-0.232 (0.011)</td>
<td>0.217 (0.010)</td>
<td>-0.198 (0.007)</td>
<td>0.131 (0.025)</td>
</tr>
<tr>
<td>6</td>
<td>-0.220 (0.010)</td>
<td>0.208 (0.010)</td>
<td>-0.180 (0.009)</td>
<td>0.072 (0.025)</td>
</tr>
</tbody>
</table>
Package ‘bite’

November 11, 2019

Type Package

Title Bayesian Inference On Treatment Effects For Panel Data

Version 0.1.0

Date 2019-08-18

Maintainer Patrick Pfeiffer <patrick.pfeiffer.dsc@gmail.com>

Description The bite package offers two Bayesian modeling strategies for treatment effects estimation on panel structured outcomes. The package contains a shared factor and a switching regression approach which estimate regression type equations under full posterior inference. The models are helpful to deal with confounding factors of the subjects as well as correlation structures between treatment and panel outcome series.

Imports numbers, Rsolnp, nloptr, truncnorm, Matrix

Suggests ggplot2

Depends R(>= 3.5.2)

URL http://github.com/PatrickPfeifferDsc/bite

BugReports http://github.com/PatrickPfeifferDsc/bite/issues

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.1

R topics documented:

bayesTrtEffects ......................................................... 2
comp_logml ................................................................. 4
comp_trt_effects .......................................................... 4
condmoments_util ....................................................... 5
drawAlphaIndicSF ....................................................... 5
drawBetaLambdaIndicesSF ............................................ 6
Bayesian Estimation of Treatment Effects in Panel Setting

Description

Performs Bayesian estimation of regression coefficients and treatment effects on panel structured data on a outcome of interest. The dataset may be unbalanced, but needs to fulfill some conditions to get proper results. A baseline matrix of characteristics before treatment choice of the subject is needed. A panel matrix is needed which contains the panel structured outcomes. The function offers 2 model choices, shared factor and switching regression which determine the framework in which the parameters are estimated.

Usage

bayesTrtEffects(base_mat, panel_mat, type = "SF", covars = NULL,
mcmc_control = list(burnin = 1000, select = 500, M = 1000),
prior_control = list(var_sel = 5, var_fix = 0.1),
control = list(fix_alpha = FALSE, fix_beta = FALSE, fix_sigma = FALSE,
fix_f = FALSE, sort_data = TRUE, test = FALSE), model_name = "model1",
data_name = "treatment dataset 1")
Arguments

**base_mat**
a data frame or matrix object, containing columns with subject id (first column),
maximum panel time $T_i$ on each subject (second column), the treatment variable,
binary (third column) and the baseline features (the following column), which
will all be used for constructing the model matrix.

**panel_mat**
a data frame of matrix object, containing id (first column), panel time indicator
per subject (second column), the panel outcomes (third column) and all features
relevant for regression on the outcomes.

**type**
contains one of two possible strings ("SF" - Shared Factor, "SWR" - Switching
Regression) In general "SF" will take much less computation time, while "SWR"
offers slightly more flexible dependence structure.

**covars**
By default is NULL. May Contain the following elements of a list, to further
specify fixed variables (that are not subject to variable selection) or common
effects (variables get the same effect, independent of treatment). MUST con-
tain the following elements, if not all variables are automatically subjected to
selection, and assumed to be able to have different effects on the target under
treatment.

- 'x_fix' vector of 0s and 1s with length of the number of variables in treat-
  ment selection model. Any 1 will not be target of variable selection. De-
defaults to all 0s
- 'y_fix' vector of 0s and 1s with length of the number of variables in the
  outcome models. Any 1 will not be target of variable selection. Defaults to
  all 0s
- 'y_common' a vector of 0s and 1s with length equal to number of variables
  in the outcome models. If a 1 occurs, the corresponding variable will be
  estimated with indifference to treatment and only have a common effect. Is
  automatically excluded from variable selection
- 'y_nb' number of base parameters
- 'y_np' number of parameters in the selection model

**mcmc_control**
a list of items controlling the mcmc parameters, see examples

**control**
a list with some general control parameters, the most important is control$sort_data
which sorts the data for ids, treatment and panel times.

**model_name**
String which names the model, defaults to "model1"

**data_name**
a character vector to store in the result object, for an automated creation of many
objects in sequence, based on different data.

**sort_data**
boolean which indicates whether the baseline data and panel data should be
organised by treatment and panel time. If the dataset is not yet sorted this option
may sort data in correct format.

**cov_y_common**
by default is NULL. If not NULL, has to be a vector of 0s and 1s with length
matching the number of covariates in panel matrix. For the covariates indexed
with 1, the model does not estimate different effects w.r.t. treatment, but assumes
the covariate has an overall effect on the different panel outcomes.

**model_name**
a character vector to store in the result object, for an automated creation of many
objects in sequence, based on different models.

Value

For adequate input files, bayesTrtEffects returns an output list object 2 lists `mcmc` which holds
the Markov Chains of all parameters and `effects` containing the average treatment effects and their
standard deviations.
comp_logml  
Compute Marginal log-likelihood

Description
Internal. comp_logml computes the marginal log-likelihood for \( y = X \ast \alpha + \epsilon, \epsilon \sim N(0, S), \alpha \sim N(0, A_0^{-1}) \)

Usage
comp_logml(ySy, XSy, invBN, indic, iB0)

Arguments
- ySy: quadratic form of y multiplied with standard deviation matrix
- XSy: matrix product of data, sd and outcome
- invBN: inverse beta matrix
- indic: indices for regression effects, their likelihood to be accessed
- iB0: kxk prior precision matrix

Value
Returns an object containing the numeric value of the marginal likelihood and \( \beta \)-parameters of the regression equations.

comp_trt_effects  
Compute treatment effects

Description
Internal. computes treatment effects from the coefficients sampled in the MCMC process.

Usage
comp_trt_effects(mcmc_select, thin = 1)
condmoms_util

Conditional Moments Computation

Description

Internal. `condmoms_util` computes the conditional moments of the utility from treatment equations.

Usage

```r
condmoms_util(mu_xst, resy, start_x, nx, start_y, sigma, rho,
               lambda = NULL)
```

Arguments

- `mu_xst`: mean vector of utility values
- `resy`: residual vector on y
- `start_x`: matrix of \(T \times x_2\) holding information at which indicator value observations with panel time \(t\) and treatment \(j\) start
- `nx`: \(T \times x_2\) matrix containing number of observations for each combination of \(t\) and \(j\)
- `start_y`: analogous to `start_x` for outcome values with indicator values for panel observations
- `sigma`: variance matrix
- `rho`: covariance matrix
- `lambda`: (optional) reserved parameter for further programming

Value

Returns a list of moments of centered \(x\):

1. conditional mean
2. conditional variance

drawAlphaIndicSF

Draw indicators and regression coefficients

Description

Internal. `drawAlphaIndicSF` draws indicator values \(\delta\) and regression coefficients \(\alpha\) for the selection model (probit model of choosing treatment)

Usage

```r
drawAlphaIndicSF(y, X, delta, deltafix, omega, invA0, isel,
                 fix.alpha = FALSE)
```
**Arguments**

- **y**  
  outcome vector (binary)
- **X**  
  model matrix
- **delta**  
  delta index of selection
- **deltafix**  
  fixed delta indices, is 1 for indices not subject to selection
- **omega**  
  correlation parameter
- **invA0**  
  inverse alpha entries matrix
- **isel**  
  indicator whether selection is performed

**Value**

\[
\delta = k \times 1 \text{ indicators}
\]

\[
\alpha = k \times 1 \text{ regression effects}
\]

---

**drawBetaLambdaIndicesSF**

*Draw Indices and Coefficients of Beta and Lambda*

**Description**

Internal. `draw_indic_beta_lambda` draws beta coefficients for the outcome models (`y`) and lambda parameters for shared factor model, as well as their delta indicator.

**Usage**

```r
drawBetaLambdaIndicesSF(y, X, sigma2, delta, deltafix, omega1, omega2, dy,
                        invB0, isel, fix.beta = FALSE)
```

**Arguments**

- **y**  
  outcome vector
- **X**  
  model matrix
- **sigma2**  
  variance parameters
- **delta**  
  vector of indices which model parameters are exposed to the spike and slab variable selection process
- **deltafix**  
  vector indicating which variables are fixed
- **omega1**  
  omega parameter of beta
- **omega2**  
  omega parameter of lambda
- **dy**  
  dimension of outcome
- **invB0**  
  startvalue matrix of beta coefficients
- **isel**  
  indicator, is 1 when spike and slab variable selection is performed

**Value**

A list of sampled \( \delta \) and \( \beta \) parameters
drawErrorVariance

Description
Internal.

Usage
drawErrorVariance(eps2, indt, indy0, sn, prior, fix.sigma = FALSE)

Arguments
eps2 vector of squared residuals
indt indicators where panel times start
indy0 index vector y0
sn previous sd matrix
prior contains S_0 prior standard deviation

drawSharedFactor

Description
Internal. draw_sharedfactor draws the latent factors f, for SF model.

Usage
drawSharedFactor(epsx, lambdax, epsy, sigma2, lambda, start_x, nx, start_y, 
Tn, fix.f = FALSE)

Arguments
epsx residual vector on x
lambdax lambda parameter of selection model x*
epsy outcome residuals
sigma2 outcome variance
lambda Λ parameters on outcome for each panel time and treatment
start_x indicator matrix on subject level
nx number of observations
start_y indicator matrix on panel level
Tn number of observation times panel t

Value
a vector of shared factors f of length n
**drawUtility**

**Draw Utilities**

**Description**
Internal. Draw latent utility values of treatment selection equations. Latent utilities $x^*$ are sampled from truncated Normal to (0,Inf) for treatment intake and (-Inf,0) for no treatment.

**Usage**

```r
drawUtility(x, mxst, sxst)
```

**Arguments**

- `x`: vector of subject features
- `mxst`: vector of means of utilities $x^*$
- `sxst`: vector of standard deviations of utilities $x^*$

**Value**

vector of length n $x^*$

---

**draw_cov_parms**

**Draw covariance parameters for SWR model**

**Description**
Internal. Includes a maximization step to get plausible values of the likelihood for variance and correlation parameters.

**Usage**

```r
draw_cov_parms(epsy, mu_xst, start_x, nx, start_y, sigma_old, rho_old, pri)
```

**Arguments**

- `epsy`: residuals of y from
- `mu_xst`: structural mean of utility
- `start_x`: matrix of $T \times max \times 2$ holding information at which indicator value observations with panel time t and treatment j start
- `nx`: $T \times x \times 2$ matrix containing number of observations for each combination of t and j
- `start_y`: analogous to start_x for outcome values with indicator values for panel observations
- `sigma_old`: previous variance value in the MC
- `rho_old`: previous correlation value in the MC
- `pri`: list of prior values

**Value**

Returns the new sampled values for sigma and rho.
**draw_factor_switchreg**  
*Draw latent factors for SWR model*

**Description**
Internal. Samples the subject specific latent factors in each MCMC iteration.

**Usage**
```
draw_factor_switchreg(resy, resx, start_x, nx, start_y, sgma, rho, lambda, D)
```

**Arguments**
- `start_x` matrix of $T_{max}x 2$ holding information at which indicator value observations with panel time $t$ and treatment $j$ start
- `nx` $T x 2$ matrix containing number of observations for each combination of $t$ and $j$
- `start_y` analogous to `start_x` for outcome values with indicator values for panel observations
- `epsy` residuals of $y$ from $\mu_xst$
- `mu_xst` structural mean of utility
- `sgma_old` previous variance value in the MC
- `rho_old` previous correlation value in the MC
- `pri` list of prior values

**Value**
returns vector of factor values for each subject

---

**draw_indic_coeff_switchreg**  
*Draw indices for selection upon sigma and rho parameters for SWR model*

**Description**
Internal.

**Usage**
```
draw_indic_coeff_switchreg(x, y, Wx, Wy, start_x, nx, start_y, sgma, rho, lambda, D, delta, deltafix, omega1, omega2, pri_invB0, isel)
```
draw_loadings  

*Helpfunction for SWR Model for sampling the Latent Factor*

**Description**

Internal.

**Usage**

```
draw_loadings(resy, Xf, resx, start_x, nx, start_y, sgmaj, rho, prior)
```

ineff_factor  

*Inefficiency Factor*

**Description**

Internal. `ineff_factor` calculates inefficiency factors for time series with autocorrelation, see e.g. Geyer (1992).

**Usage**

```
ineff_factor(x)
```

**Arguments**

- **x**
  - a time series of length k

**Value**

2 numeric values, the empirical inefficiency factor and m

lcondx  

*Compute Negative Log-Likelihood*

**Description**

Internal. Computes normal, conditional negative log-likelihood for all subject observations with treatment j.

**Usage**

```
lcondx(jtrt, epsy, mu_xst, start_x, nx, start_y, sgmaj, rhoj)
```
Arguments

- jtrt: treatment j
- epsy: residual vector of y i.e. $\epsilon = y - \mu_y$
- mu_xst: mean vector of $x^*$
- start_x: $T_{max} \times trt$ matrix, containing indicators for panel time - treatment combinations in data vector
- nx: number of observations
- start_y: $T_{max} \times trt$ matrix, containing indicators for panel time - treatment combinations in outcome vector
- sgmaj: variance under treatment j
- rhoj: correlation coefficient under treatment j

Value

The log-Likelihood value for given point

See Also

Other loglikelihoods: llikj, lmvnorm, lnorm, negllik_jt_rho, negllik_jt

llikj 11

Compute log-Likelihood

Description

Internal. Computes the log-likelihood for all observations with treatment j.

Usage

llikj(sgmaj, rhoj, j, epsy, mu_xst, start_x, nx, start_y)

Arguments

- sgmaj, rhoj: vector of $2 \times T_{max}$ parameter of the variance/covariance matrix
- j: treatment
- epsy: residuals on y
- mu_xst: mean vector of $x^*$
- start_x: $T_{max}x2$ matrix, containing indicators for panel time - treatment combinations in data vector
- nx: number of observations
- start_y: $T_{max}x2$ matrix, containing indicators for panel time - treatment combinations in outcome vector

Value

The log-likelihood value corresponding to input values.

See Also

Other loglikelihoods: lcondx, lmvnorm, lnorm, negllik_jt_rho, negllik_jt
**lmvnorm**  
Compute log-Likelihood of multivariate Normal

**Description**  
Internal help function.

**Usage**  
```r
lmvnorm(y, mu, Sigmainv)
```

**Arguments**  
- `y`  
  outcome vector
- `mu`  
  vector of means
- `Sigmainv`  
  inverse of variance matrix

**Value**  
log-Likelihood value under multivariate Normal assumption

**See Also**  
Other loglikelihoods: `lcondx, llikj, lnorm, negllik_jt_rho, negllik_jt`

---

**lnorm**  
Compute Normal log-Likelihood

**Description**  
Internal. Computes log-likelihood of a Normal distributed sample

**Usage**  
```r
lnorm(x, mu, varinv)
```

**Arguments**  
- `x`  
  sample
- `mu, varinv`  
  mean, variance parameters

**Value**  
The log-Likelihood of given parameters.

**See Also**  
Other loglikelihoods: `lcondx, llikj, lmvnorm, negllik_jt_rho, negllik_jt`
**loglik_mvnorm_balpan**  
*Compute log-Likelihood of Balanced Panels*

**Description**  
Internal. `loglik_mvnorm_balpan` computes for balanced panel data $dim(y) = t \times N$ with equal $\Sigma$ across observations.

**Usage**  
```r
loglik_mvnorm_balpan(y, mu = matrix(0, dim(y)[1], dim(y)[2]), Sigmainv)
```

**Arguments**  
- `y` : data
- `mu` : mean matrix
- `Sigmainv` : inverse variance matrix

**Value**  
A scalar value, the log-likelihood for given data, mu and Sigma.

---

**make_regmat_unbal**  
*Transcribes Basic Object to Internal Matrix*

**Description**  
Internal help function. Takes the dataframes after read-in of base_mat and panel_mat and constructs the according model matrices.

**Usage**  
```r
make_regmat_unbal(data, model)
```

**Arguments**  
- `data` : list item with subject and panel features
- `model` : information on the specified model

**Value**  
A list object with model matrices $W$ and $W_x$. 
### mhStep

*Computes a Metropolis Hastings Step*

**Description**

Internal help function. Performs an accept/reject step with proposal value.

**Usage**

```r
mhStep(theta_old, theta_new, lpost_old, lpost_new, lq_new, lq_old)
```

**Arguments**

- `theta_old`: previous parameter value
- `theta_new`: new proposal parameter value
- `lpost_old`: previous log-posterior
- `lpost_new`: new log posterior
- `lq_new`: \( \log q(\theta_{new}|\theta_{old}) \) the proposal log-densities
- `lq_old`: \( \log q(\theta_{old}|\theta_{new}) \) the proposal log-densities

**Value**

A vector of the new \( \theta \) values

---

### negllik_jt

*Compute Negative log-Likelihood*

**Description**

Internal. Help function to compute negative log-likelihood

**Usage**

```r
negllik_jt(theta, j, t, epsy, mu_xst, start_x, nx, start_y, sgmaj, rhoj)
```

**Arguments**

- `theta`: parameter of variance/covariance matrix of dimension \( 2T_{max}x1 \)
- `j`: treatment
- `t`: panel time
- `epsy`: residuals of y
- `mu_xst`: mean vector of x*
- `start_x`: \( T_{max}x2 \) matrix, containing indicators for panel time - treatment combinations in data vector
- `nx`: dimension of x
- `start_y`: \( T_{max}x2 \) matrix, containing indicators for panel time - treatment combinations in outcome vector
- `sgmaj`: variance for treatment \( j \)
- `rhoj`: correlation for treatment \( j \)
Value

A scalar which is the negative log-Likelihood at the given point.

See Also

Other loglikelihoods: lcondx, llikj, lmvnorm, lnorm, negllik_jt_rho

---

negllik_jt_rho

Computes negative log-Likelihood

Description

Internal. Help function to compute log-likelihood, see negllik_jt.

Usage

negllik_jt_rho(theta, j, t, epsy, mu_xst, start_x, nx, start_y, sgmaj, rhoj)

Arguments

theta parameter of variance/covariance matrix of dimension 2Tmax x 1
j treatment
t panel time
epsy residuals of y
mu_xst mean vector of x*
start_x Tmaxx2 matrix, containing indicators for panel time - treatment combinations in data vector
nx dimension of x
start_y Tmaxx2 matrix, containing indicators for panel time - treatment combinations in outcome vector
sgmaj variance for treatment j
rhoj correlation for treatment j

Value

A scalar which is the negative log-Likelihood at the given point.

See Also

Other loglikelihoods: lcondx, llikj, lmvnorm, lnorm, negllik_jt
postmom_help_reg method

Description
Internal. Computes help quantities for posterior moments of regression coefficients.

Usage
postmom_help_reg(x, y, Wx, Wy, start_x, nx, start_y, sigma, rho, lambda, D)

Arguments
x    data
y    outcome
Wx   design matrix for selection model
Wy   design matrix for outcome data model
start_x matrix containing indices where certain observations begin in the vector sorted
        by treatment and panel times
nx   dimension of x
start_y matrix containing indices where certain outcomes begin in the vector sorted by
treatment and panel times
sigma variance parameter matrix
rho  correlation parameter matrix
lambda matrix of factor loadings
D    initial inclusion for treatment j

Value
List of Matrices used for computation of posterior moments.

readPanelUb Read Data file and transform

Description
Internal. Help function to read data and apply certain structures, for specifics see bayesTrtEffects().

Usage
readPanelUb(base_mat, panel_mat, type, covars, name, sort_data, control_test, data_name)
selectTreatmentSF

Arguments

base_mat, panel_mat
matrices or dataframes with base and panel structure
type
either "SF" or "SWR"
covars
contains additional information about covariate names (optional)
name
contains name of the model, which will be saved
sort_data
control parameter to sort data by treatment, panel time and id

Value

A data object for further computation

selectTreatmentSF Modelling Treatment Effects via Shared Factor Model

Description

Internal. Main function call by bayesTrtEffects to start MCMC routine under shared factor model. Passes parameters on to computational help functions.

Usage

selectTreatmentSF(data, model, prior, mcmc, control)

Arguments

data
takes data object from the function bayesTrtEffects
model
takes model parameters from function bayesTrtEffects
prior
specified prior parameters
mcmc
mcmc parameters specified by bayesTrtEffects, potentially subject to change

Value

Returns a list object containing the Markov Chain of estimated coefficients and other parameters.
selectTreatmentSWR  Modelling Treatment Effects via Switching Regression Model

Description
Internal. Second main function call by bayesTrtEffects to start MCMC routine under switching regression model. Passes parameters on to computational help functions.

Usage
selectTreatmentSWR(data, model, prior, mcmc)

Arguments
- data: takes data object from the function bayesTrtEffects
- model: takes model parameters from function bayesTrtEffects
- prior: specified prior parameters
- mcmc: mcmc parameters specified by bayesTrtEffects, potentially subject to change

Value
Returns a list object containing the Markov Chain of estimated coefficients and other parameters.

startingValues  Computes starting values for parameter search

Description
Internal. Help function to compute starting values for alpha and beta parameters. Alpha parameters are estimated by a GLM fit, beta parameters by OLS fit.

Usage
startingValues(model, x, y, Tn)

Arguments
- model: Contains model parameters
- x: contains data
- y: contains outcome (target) information
- Tn: panel times dimension

Value
Returns a list of
1. starting values for $\alpha$
2. starting values for $\beta$
3. residual variance
**Description**

A simulated dataset from the framework of the shared factor model to demonstrate the use of the SR-model. The file contains 5000 subjects with their respective feature information at baseline time. Consider these features to be relevant for confounding and influencing the treatment intake, not the outcomes. First of two data tables given to function `bayesTrtEffects`.

**Usage**

```
trt_baseline
```

**Format**

A data frame with subject related feature columns:

- **ID** the column which uniquely identifies each subject
- **Ti** panel times, the number of subsequent measures for a subject
- **trt** treatment, usually coded 0 or 1
- **V1** variable 1, normally distributed
- **V2** variable 2, 0/1 distributed
- **V3** variable 3, 0/1 distributed

**Source**

This dataset stems from a simulation process and represents fictive data.

---

**Description**

A simulated dataset from the framework of the shared factor model to demonstrate the use of the SR model in package 'bite'. The file contains 5000 subjects, each of which has 4 panel observations on simulated variables. The second of 2 argument datasets for `bayesTrtEffects`.

**Usage**

```
trt_paneldata
```
Format

A data frame with subject related feature columns:

- **ID**  ID of subject/item
- **panelT**  timepoint t of measured features and dependent variable
- **y**  dependent variable, outcome
- **V1**  variable 1, equivalent to V1 in baseline
- **V2**  variable 2, equivalent to V2 in baseline
- **t2**  dummy for panel time = 2
- **t3**  dummy for panel time = 3
- **t4**  dummy for panel time = 4

Source

This dataset stems from a simulation process and represents fictive data.
Index

# Topic datasets
  trt_baseline, 19
  trt_paneldata, 19

bayesTrtEffects, 2, 19

comp_logml, 4
comp_trt_effects, 4
condmoms_util, 5

draw_cov_parms, 8
draw_factor_switchreg, 9
draw_indic_coeff_switchreg, 9
draw_loadings, 10
drawAlphaIndicSF, 5
drawBetaLambdaIndicesSF, 6
drawErrorVariance, 7
drawSharedFactor, 7
drawUtility, 8

ineff_factor, 10

lcondx, 10, 11, 12, 15
llikj, 11, 12, 15
lmvnorm, 11, 12, 12, 15
lnorm, 11, 12, 12, 15
loglik_mvnorm_balpan, 13

make_regmat_unbal, 13
mhStep, 14

negllik_jt, 11, 12, 14, 15
negllik_jt_rho, 11, 12, 15, 15

postmom_help_reg, 16

readPanelUb, 16

selectTreatmentSF, 17
selectTreatmentSWR, 18
startingValues, 18

trt_baseline, 19
trt_paneldata, 19
Bibliography


