Energy Mix

Nadja Klein, Ilya Zarubin, Koautor

1 General Model Specifiction

response: for $\boldsymbol{y} = (y_1, \ldots, y_D)^\top$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D) \in \mathbb{R}^D_{>0}$, the data is dirichlet distributed:
$p(y_1, \ldots, y_D) = \frac{1}{B(\alpha)} \prod_{d=1}^D y_d^{\alpha_d - 1}$ i.e. $\boldsymbol{y} \sim \mathbf{Dir}(\boldsymbol{\alpha})$,
where $D \geq 2$, $\sum_{d=1}^D y_d = 1$, $y_d \geq 0$, and $B(\boldsymbol{\alpha}) = \frac{\prod_{d=1}^D \Gamma(\alpha_d)}{\Gamma(\sum_{d=1}^D \alpha_d)}$.
Moreover, $\mathbb{E}(y_d) = \frac{\alpha_d}{\alpha_0}$, $\alpha_0 = \sum_{d=1}^D \alpha_d$.

link: $\log(\alpha_d) = \eta_d$ linking the d'th parameter to some semi-parametric predictor.

predictor: $\eta_d = \sum_{j=1}^{J_d} f_{j,d}(\boldsymbol{\nu}) = \sum_{j=1}^{J_d} \boldsymbol{Z}_{j,d} \boldsymbol{\beta}_{j,d}$, where $J_d = \#$ effects for d'th parameter

design matrix: depending on the particular covariate effect under consideration, $\boldsymbol{Z}_{j,d}$ is ...

I. linear effects: ... the standard regressor matrix with the first column as a column of 1's

II. non-linear effects: ... composed of S B-spline basis functions evaluated at observed covariates to approximate potentially non-linear effects

III. spatial effects: ... the $N \times S$ design matrix is an incidence matrix, i.e. $\boldsymbol{Z}_{j,d}[n, s]$ equals 1 if observation n belongs to location s and 0 otherwise. (assuming that the location or spatial grouping does not change over time t)

IV. random effects: ... an incidence matrix with zeros and ones, linking individual observations or clusters, e.g $\boldsymbol{Z}_{j,d}[n, g] = 1$ if observation n is in group g and zero otherwise. Here, the assumption that the grouping is stable is more restrictive and might have to be relaxed if the clustering cannot be assumed to be constant over time t.

prior: $p\left(\boldsymbol{\beta}_{j,d} | \tau_{j,d}\right) \propto \left\{\frac{1}{(\tau_{j,d})^2}\right\}^{\mathbf{rk}(\boldsymbol{K}_{j,d})/2} \exp\left\{-\frac{1}{2(\tau_{j,d})^2} \left(\boldsymbol{\beta}_{j,d}\right)^\top \boldsymbol{K}_{j,d} \boldsymbol{\beta}_{j,d}\right\}$,

where $\boldsymbol{K}_{j,d}$: prior precision matrix, $(\tau_{j,d})^2$: hyperprior smoothing variance

I. linear effects: for noninformative priors: $\boldsymbol{K}_{j,d} = \mathbf{0}$ or $\boldsymbol{K}_{j,d} = \boldsymbol{I}$ if $\dim(\boldsymbol{\beta}_{j,d})$ is large

II. non-linear effects: $\boldsymbol{K}_{j,d} = D^\top D$, where D is a difference matrix of appropriate order

III. spatial effects: To enforce spatial smoothness, various neighbourhood sets $(\partial_1, \ldots, \partial_S)$ are defined for regions that share a common border i.e. for any two regions r and s that are neighbours we write $r \sim s$, and $r \nsim s$ otherwise. This implies an adjacency matrix indicating which regions are neighbours

$$\boldsymbol{K}_{j,d} = \begin{cases} -1\ , & s \neq r, r \in \partial_s\ , \\ 0\ , & s \neq r, r \notin \partial_s\ , \\ N_s\ , & s = r\ . \end{cases}$$

IV. random effects: assuming that coefficients of different groups are i.i.d. distributed $\boldsymbol{K}_{j,d} = \boldsymbol{I}$

2 Notation

Panel setting: cross section and time indices are $n = 1, \ldots, N$ and $t = 1, \ldots, T$.

Let $\boldsymbol{Y}_{\text{mat}} = \begin{pmatrix} \boldsymbol{y}_{11} & \cdots & \cdots & \cdots & \boldsymbol{y}_{1T} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \boldsymbol{y}_{nt} & & \vdots \\ \vdots & & & \ddots & \vdots \\ \boldsymbol{y}_{N1} & \cdots & \cdots & \cdots & \boldsymbol{y}_{NT} \end{pmatrix}$, and let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{\alpha}_{11} & \cdots & \cdots & \cdots & \boldsymbol{\alpha}_{1T} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \boldsymbol{\alpha}_{nt} & & \vdots \\ \vdots & & & \ddots & \vdots \\ \boldsymbol{\alpha}_{N1} & \cdots & \cdots & \cdots & \boldsymbol{\alpha}_{NT} \end{pmatrix}$.

where $\boldsymbol{y}_{nt} = (y_{nt,1}, \ldots, y_{nt,d} \ldots, y_{nt,D})$ and $\boldsymbol{\alpha}_{nt} = (\alpha_{nt,1}, \ldots, \alpha_{nt,d} \ldots, \alpha_{nt,D})$.

Then, for any $d = 1, \ldots, D$: $\eta_{nt,d} = \log \alpha_{nt,d}$ and $\eta_{nt,d} = \sum_{j=1}^{J_d} f_{j,d}(\boldsymbol{\nu}_{nt})$, where $\boldsymbol{\nu}_{nt}$ comprises regressor information about possible linear, nonlinear, random and/or spatial effects. When smooth terms are represented as basis functions, the predictors can be written more compactly in matrix form aggregating over $t-$ or $n-$indeces:

$$\boldsymbol{\eta}_{t,d} = \sum_{j=1}^{J_d} \boldsymbol{Z}^t_{j,d} \boldsymbol{\beta}_{j,d} \ , \quad \text{or} \quad \boldsymbol{\eta}_{n,d} = \sum_{j=1}^{J_d} \boldsymbol{Z}^n_{j,d} \boldsymbol{\beta}_{j,d} \ ,$$

where $\boldsymbol{Z}^t_{j,d} = \begin{pmatrix} z_{11,d} & \cdots & \cdots & \cdots & z_{1N_j,d} \\ \vdots & \ddots & & & \vdots \\ \vdots & & z_{nn_j,d} & & \vdots \\ \vdots & & & \ddots & \vdots \\ z_{N1,d} & \cdots & \cdots & \cdots & z_{NN_j,d} \end{pmatrix}$, or $\boldsymbol{Z}^n_{j,d} = \begin{pmatrix} z_{11,d} & \cdots & \cdots & \cdots & z_{1N_j,d} \\ \vdots & \ddots & & & \vdots \\ \vdots & & z_{tn_j,d} & & \vdots \\ \vdots & & & \ddots & \vdots \\ z_{T1,d} & \cdots & \cdots & \cdots & z_{TN_j,d} \end{pmatrix}$,

with corresponding dimensions of the basis coefficient vectors being $\dim(\boldsymbol{\beta}_{j,d}) = N_j$ i.e. $n_j = 1, \ldots, N_j$, and the log-link $\exp(\boldsymbol{\eta}_{t,d}) = \boldsymbol{\alpha}_{t,d} = (\alpha_{1t,d}, \ldots, \alpha_{Nt,d})$, and $\exp(\boldsymbol{\eta}_{n,d}) = \boldsymbol{\alpha}_{n,d} = (\alpha_{n1,d}, \ldots, \alpha_{nT,d})$. The n'th row of $\boldsymbol{Z}^t_{j,d}$ and the t'th row of $\boldsymbol{Z}^n_{j,d}$ are given, respectively, as

$$\boldsymbol{z}^t_{n,j,d} = [z_{n1,d} \ldots, z_{nN_j,d}] \ , \quad \text{and} \quad \boldsymbol{z}^n_{t,j,d} = [z_{t1,d}, \ldots, z_{tN_j,d}] \ .$$

Finally, we write the $\boldsymbol{\beta}$'s, $\boldsymbol{\tau}$'s and $\boldsymbol{Z}$'s compactly as

$$\begin{aligned} \boldsymbol{\theta} &= \left(\boldsymbol{\beta}_{1,1}, \ldots, \boldsymbol{\beta}_{J_1,1}, \ldots, \boldsymbol{\beta}_{1,D}, \ldots, \boldsymbol{\beta}_{J_D,D}, \tau_{1,d}, \ldots, \tau_{J_1,1}, \ldots, \tau_{1,D}, \ldots, \tau_{J_D,D}\right) \ , \\ \boldsymbol{Z}_{n,t} &= \left(\boldsymbol{z}^n_{t,1,1}, \ldots, \boldsymbol{z}^n_{t,J_1,1}, \ldots, \boldsymbol{z}^n_{t,1,D}, \ldots, \boldsymbol{z}^n_{t,J_D,D}\right) \\ &= \left(\boldsymbol{z}^t_{n,1,1}, \ldots, \boldsymbol{z}^t_{n,J_1,1}, \ldots, \boldsymbol{z}^t_{n,1,D}, \ldots, \boldsymbol{z}^t_{n,J_D,D}\right) \ . \end{aligned}$$

3 Model I: independence case

assumptions: The first model version assumes the data to be independently distributed as

$$\boldsymbol{y}_{nt}|\boldsymbol{\theta}, \boldsymbol{Z}_{n,t} \overset{ind.}{\underset{n,t}{\sim}} \mathbf{Dir}(\boldsymbol{\alpha}_{nt}) \ .$$

where $\boldsymbol{\alpha}_{nt} = \left(\alpha_{nt}^1, \ldots, \alpha_{nt}^D\right)$. Note that $\boldsymbol{Z}_{n,t}$ might contain lagged variables/covariates.

likelihood: For notational convenience let $\boldsymbol{Y} := \text{vec}(\boldsymbol{Y}_{\text{mat}}) = (\boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{nt}, \ldots, \boldsymbol{y}_{NT})$ and $\boldsymbol{Z} = \left(\boldsymbol{Z}_{j,d}^1, \ldots, \boldsymbol{Z}_{j,d}^T\right) = \left(\boldsymbol{Z}_{j,d}^1, \ldots, \boldsymbol{Z}_{j,d}^N\right)$. Under independence the likelihood for $\boldsymbol{Y}$ factors along its (multivariate) marginal densities according to

$$p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{Z}) = \prod_{n=1}^{N}\prod_{t=1}^{T} p(\boldsymbol{y}_{nt}|\boldsymbol{\theta}, \boldsymbol{Z}_{n,t}) = \prod_{n=1}^{N}\prod_{t=1}^{T}\left[B(\boldsymbol{\alpha}_{nt})^{-1} \times \prod_{d=1}^{D}(y_{nt,d})^{(\alpha_{nt,d}-1)}\right] \ ,$$

with $B(\boldsymbol{\alpha}_{nt}) = \frac{\prod_{d=1}^{D}\Gamma(\alpha_{nt,d})}{\Gamma\left(\sum_{d=1}^{D}\alpha_{nt,d}\right)}$.

prior: $\boldsymbol{\beta}_{j,d}|\tau_{j,d} \sim \mathcal{N}(0, \boldsymbol{K}_{j,d})$ & $\tau_{j,d} \sim \mathcal{IG}(a, b)$, $a = b = 0.001$ i.e. uninformative $\forall j, d$. $\boldsymbol{K}_{j,d}$ is set according to the effect type (linear, non-linear, spatial, RE etc.).

posterior: The posterior is given as

$$\begin{aligned}
&p(\boldsymbol{\theta}|\boldsymbol{Y}, \boldsymbol{Z}) \propto p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{Z})p(\boldsymbol{\theta}) \\
&= \prod_{n=1}^{N}\prod_{t=1}^{T} B(\boldsymbol{\alpha}_{nt})^{-1} \times \prod_{d=1}^{D}\left[(y_{nt,d})^{(\alpha_{nt,d}-1)} \times \prod_{j=1}^{J_1} p\left(\boldsymbol{\beta}_{j,d}|\tau_{j,d}\right) p\left(\tau_{j,d}\right) \ldots \times \prod_{j=1}^{J_D} p\left(\boldsymbol{\beta}_{j,d}|\tau_{j,d}\right) p\left(\tau_{j,d}\right)\right] .
\end{aligned}$$

Consequently, posterior sub-blocks for each $\boldsymbol{\beta}_{j,d}$ and $\tau_{j,d}$ are given as

$$\begin{aligned}
&p\left(\boldsymbol{\beta}_{j,d}|\boldsymbol{\theta}_{-\boldsymbol{\beta}_{j,d}}\right) \propto (y_{nt,d})^{(\alpha_{nt,d}-1)} \times p\left(\boldsymbol{\beta}_{j,d}|\tau_{j,d}\right) \times \prod_{n=1}^{N}\prod_{t=1}^{T} B(\boldsymbol{\alpha}_{nt})^{-1} \ , \\
&(\tau_{j,d})^2|\boldsymbol{\theta}_{-\tau_{j,d}} \sim \mathcal{IG}(\tilde{a}_j^d, \tilde{a}_j^d) \ ,
\end{aligned}$$

with $\boldsymbol{\theta}_{-\boldsymbol{\beta}_{j,d}} = \boldsymbol{\theta}\backslash\{\boldsymbol{\beta}_{j,d}\}$, $\boldsymbol{\theta}_{-\tau_{j,d}} = \boldsymbol{\theta}\backslash\{\tau_{j,d}\}$, $\tilde{a}_j^d = \frac{\mathbf{rk}(\boldsymbol{K}_{j,d})}{2} + a$, and $\tilde{b}_j^d = (\boldsymbol{\beta}_{j,d})^\top \frac{\mathbf{rk}(\boldsymbol{K}_{j,d})}{2}(\boldsymbol{\beta}_{j,d}) + b$.

Model II: autoregressive (lagged dependent) case

When lagged dependent variables are included there are two ways to deal with the initial conditions. The conditional approach assumes that particular *initial conditions* are known. For a lag order of e.g. $p = 1$ one initial condition denoted as $\boldsymbol{y}_{10}, \ldots, \boldsymbol{y}_{N0}$ is required. Consequently, for p lags we need the corresponding number of p initial conditions. Again, let $\boldsymbol{Y} := \text{vec}(\boldsymbol{Y}_{\text{mat}}) = \big(\boldsymbol{y}_{1(1-p)}, \ldots, \boldsymbol{y}_{N(1-p)}, \ldots, \boldsymbol{y}_{10}, \ldots, \boldsymbol{y}_{N0}, \boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{nt}, \ldots, \boldsymbol{y}_{NT}\big)$. The lagged dependent variable enters via the structured additive predictor: for lag order $p = 1, \ldots, P_d$ at any parameter $d = 1, \ldots, D$ we have

$$\log \alpha_{nt,d} = \eta_{nt,d} = \sum_{j=1}^{J_d} \boldsymbol{z}_{n,j,d}^{t} \boldsymbol{\beta}_{j,d} + \sum_{p=1}^{P_d} \boldsymbol{y}_{n(t-p)} \boldsymbol{\beta}_{p,d}^{lag} ,$$

where $\dim(\boldsymbol{\beta}_{p,d}^{lag}) = D \times 1$. *This allows a dependence between* $\log \alpha_{nt,d}$ *(and hence the expectation of* $y_{nt,d}$*) and lags of previous* $\boldsymbol{y}_{n(t-p)}$*'s i.e.* <u>*the lag of one particular, current energy type may depend on all previous energy types.*</u> The likelihood factorizes along its (multivariate) marginal densities, with conditioning additionally on the lagged dependent variables and corresponding parameters, according to

$$\begin{aligned} p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{Z}) &= \prod_{n=1}^{N} \prod_{t=1}^{T} p(\boldsymbol{y}_{nt}|\boldsymbol{\theta}, \boldsymbol{Z}_{n,t}, \boldsymbol{y}_{n(t-1)}, \ldots, \boldsymbol{y}_{n(t-p)}, \boldsymbol{\beta}_{1,1}^{lag}, \ldots, \boldsymbol{\beta}_{P_1,1}^{lag}, \ldots, \boldsymbol{\beta}_{1,D}^{lag}, \ldots, \boldsymbol{\beta}_{P_D,D}^{lag}) \\ &= \prod_{n=1}^{N} \prod_{t=1}^{T} \left[B(\boldsymbol{\alpha}_{nt})^{-1} \times \prod_{d=1}^{D} (y_{nt,d})^{(\alpha_{nt,d}-1)} \right] , \end{aligned}$$

with $B(\boldsymbol{\alpha}_{nt}) = \frac{\prod_{d=1}^{D} \Gamma(\alpha_{nt,d})}{\Gamma\left(\sum_{d=1}^{D} \alpha_{nt,d}\right)}$.

The prior and posterior specifications for $(\boldsymbol{\beta}_{1,1}^{lag}, \ldots, \boldsymbol{\beta}_{P_1,1}^{lag}, \ldots, \boldsymbol{\beta}_{1,D}^{lag}, \ldots, \boldsymbol{\beta}_{P_D,D}^{lag})$ are exactly as in the previous section with the effect type being modelled linearly.

4 Simulation Results

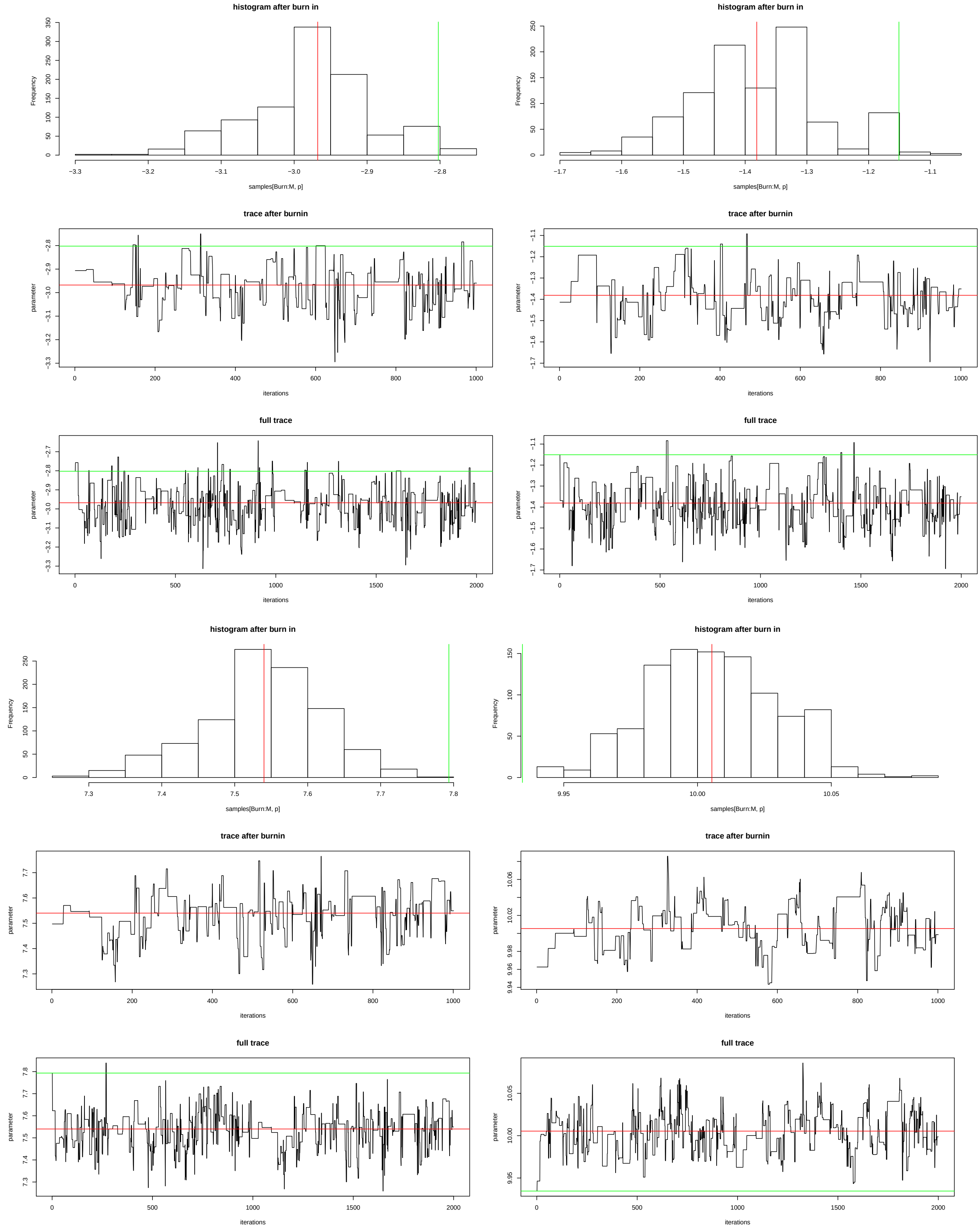
Model II is estimated with MCMC (using IWLS MH-proposals) as well as MLE (for comparison). The number of cross sections is $N = 30$ for $T = 10$ time periods. For $D = 5$ and each α_d, $d = 1, \ldots, 5$ there are 3 linear effects $\boldsymbol{\beta}_d = (\beta_{d,1}, \ldots, \beta_{d,3})$ and one lagged effect $\beta_{d,lag}$ at the one-period lagged dependent variable $y_{nt,d}$. Thus, the modelling equations for the parameters can be written as

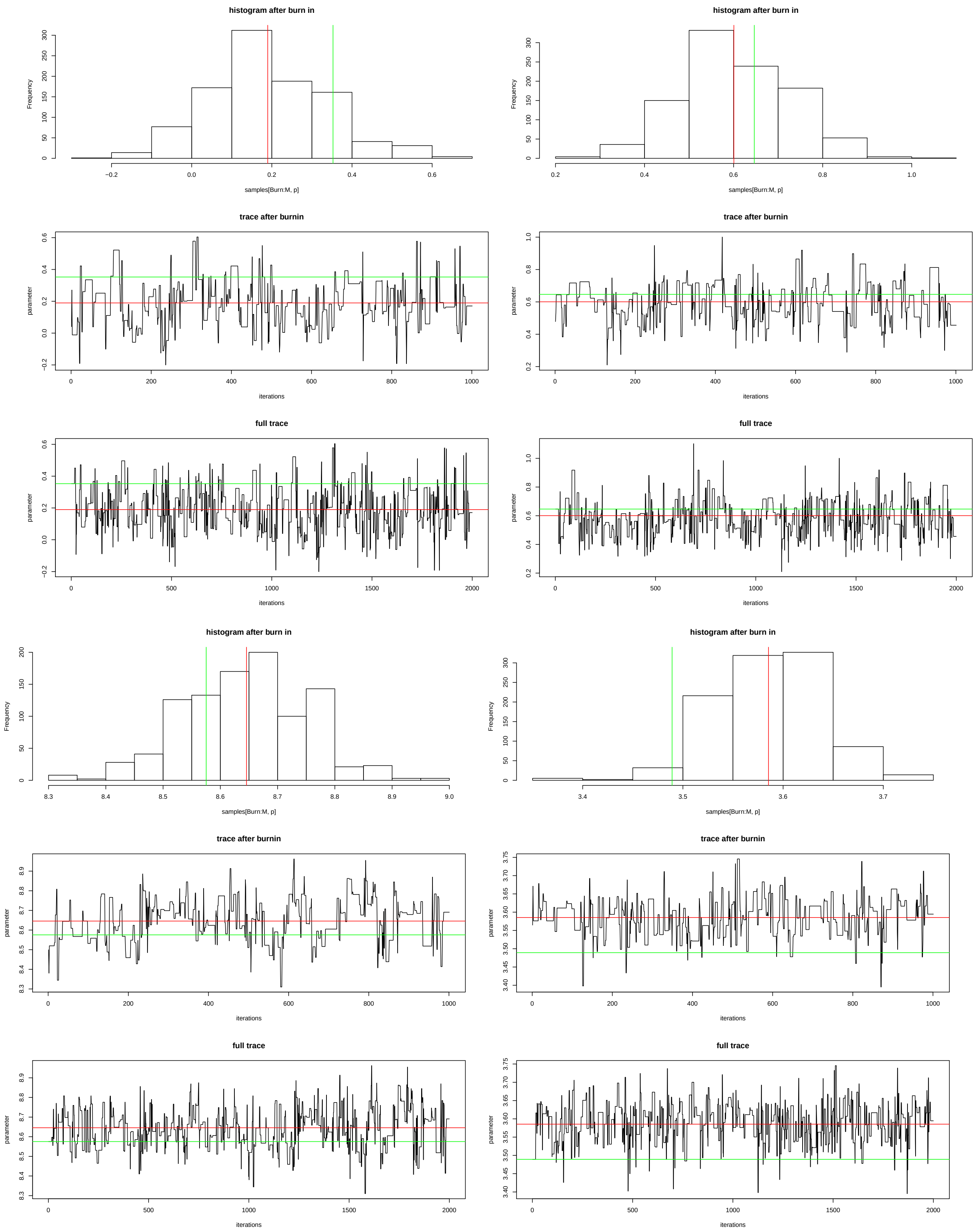
$$\log \alpha_{nt,d} = \eta_{nt,d} = \boldsymbol{x}_{nt,d}\boldsymbol{\beta}_d + y_{n(t-1),d}\beta_{d,lag} \ .$$

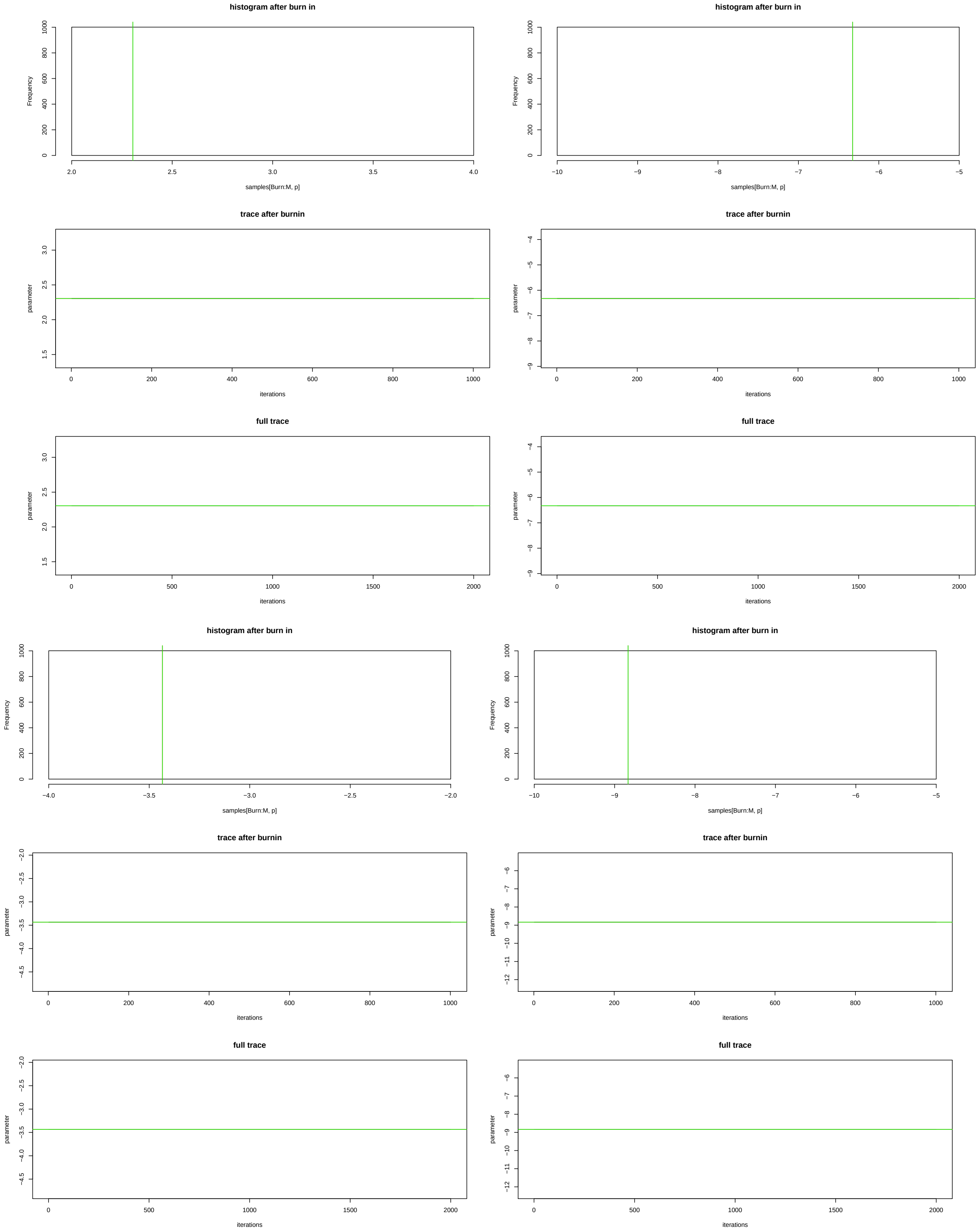
The table summarizes true parameter values, starting values for the MCMC and MLE estimations, posterior mean estimates as well as MLE estimates. $MLEONCE$ uses the same data for estimation as IWLS. $MLEAVG$ are averaged ML estimates over 50 replications of the data (with the same true values as the individual data used for MLE and IWLS). MCMC is performed with a total of $M = 50000$ draws. Acceptance rates are about 50% for all parameters.

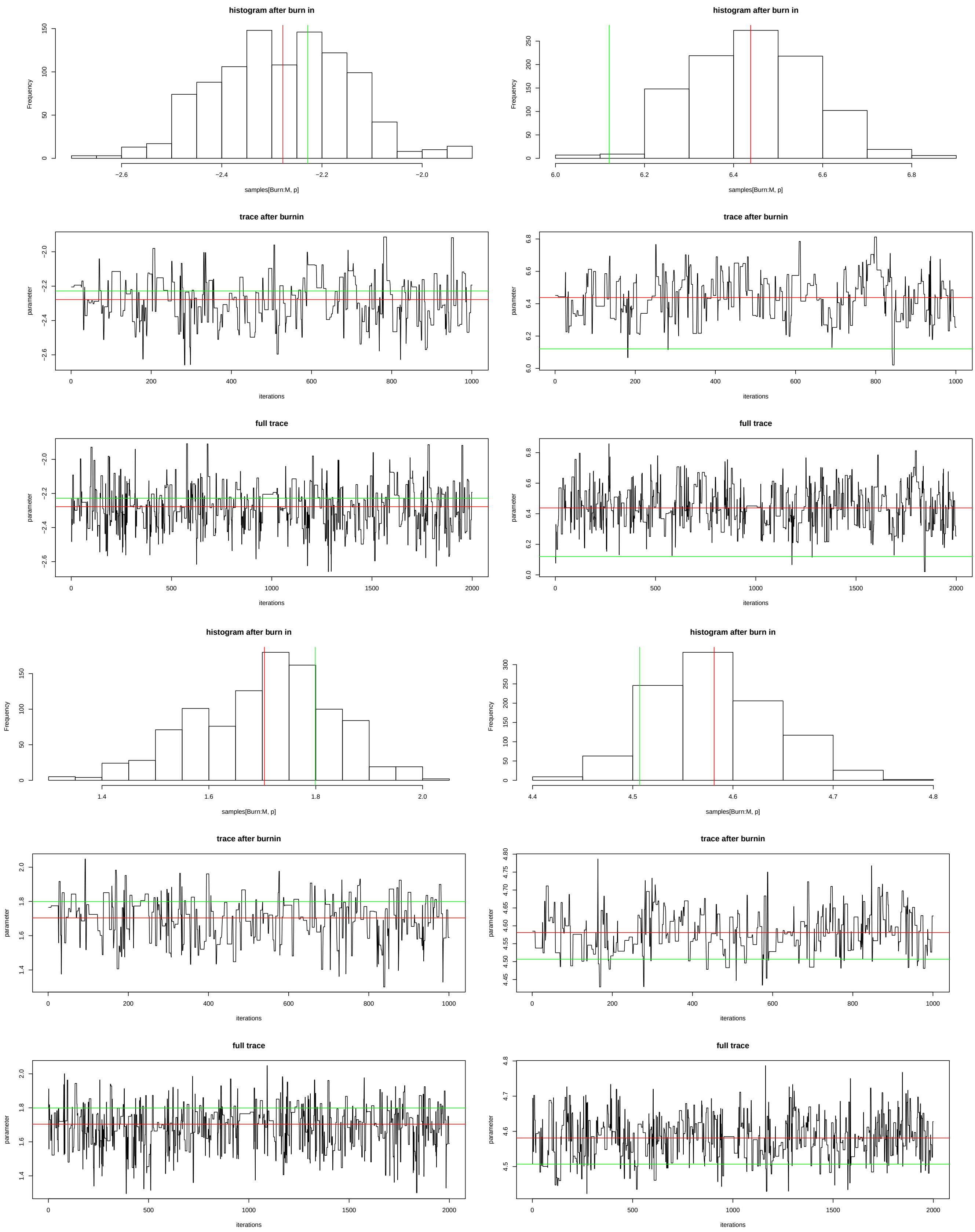
	$TRUE\ VAL$	$START\ VAL$	$IWLS$	$MLE\ ONCE$	$MLE\ AVG$
$\beta_{1,1}$	-2.8024	0	-2.8871	-2.8830	-2.8354
$\beta_{1,2}$	-1.1509	0	-1.2622	-1.2641	-1.2502
$\beta_{1,3}$	7.7935	0	7.6265	7.6259	7.7537
$\beta_{1,lag}$	9.9346	0	10.0122	10.0125	9.9387
$\beta_{2,1}$	0.3525	0	0.3296	0.3233	0.3245
$\beta_{2,2}$	0.6464	0	0.7467	0.7461	0.5841
$\beta_{2,3}$	8.5753	0	8.7177	8.7248	8.5768
$\beta_{2,lag}$	3.4893	0	3.5908	3.5918	3.5030
$\beta_{3,1}$	2.3046	0	2.2844	2.2933	2.3989
$\beta_{3,2}$	-6.3253	0	-6.4151	-6.4049	-6.3204
$\beta_{3,3}$	-3.4343	0	-3.4542	-3.4494	-3.2951
$\beta_{3,lag}$	-8.8331	0	-9.3210	-9.2284	-8.7508
$\beta_{4,1}$	-2.2283	0	-2.1367	-2.1434	-2.3060
$\beta_{4,2}$	6.1204	0	6.5359	6.5395	6.1904
$\beta_{4,3}$	1.7991	0	1.8214	1.8190	1.9573
$\beta_{4,lag}$	4.5068	0	4.6077	4.6109	4.5131
$\beta_{5,1}$	2.0039	0	2.1292	2.1441	2.1838
$\beta_{5,2}$	0.5534	0	0.7361	0.7416	0.5697
$\beta_{5,3}$	-2.7792	0	-2.7652	-2.7539	-2.6734
$\beta_{5,lag}$	-1.3640	0	-0.5929	-0.5340	-1.4939

The next page shows histograms, trace plots and traceplots after burn in (=25000) for all 20 parameters in the exact same order as they appear in the previous table.
GREEN: TRUE VALUES, RED: POSTERIOR MEANS

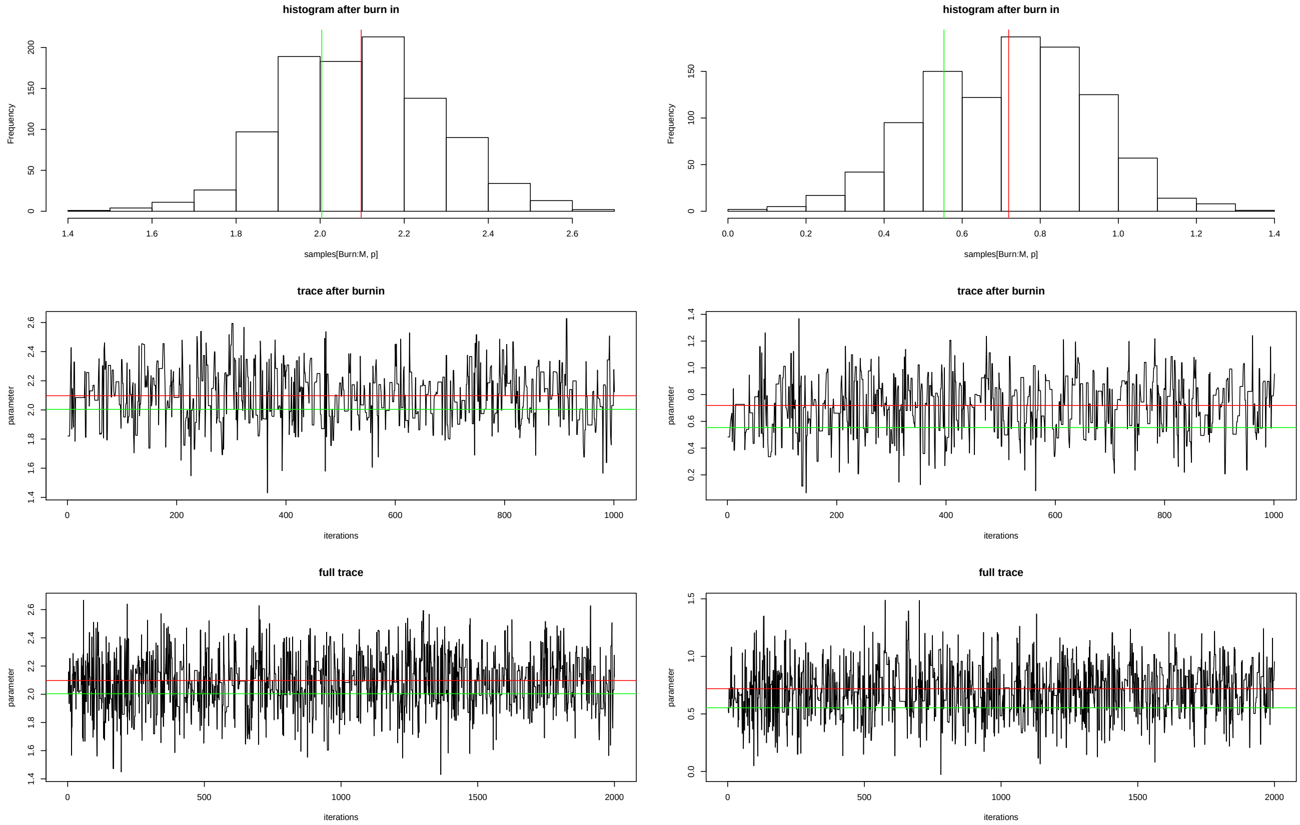

histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations

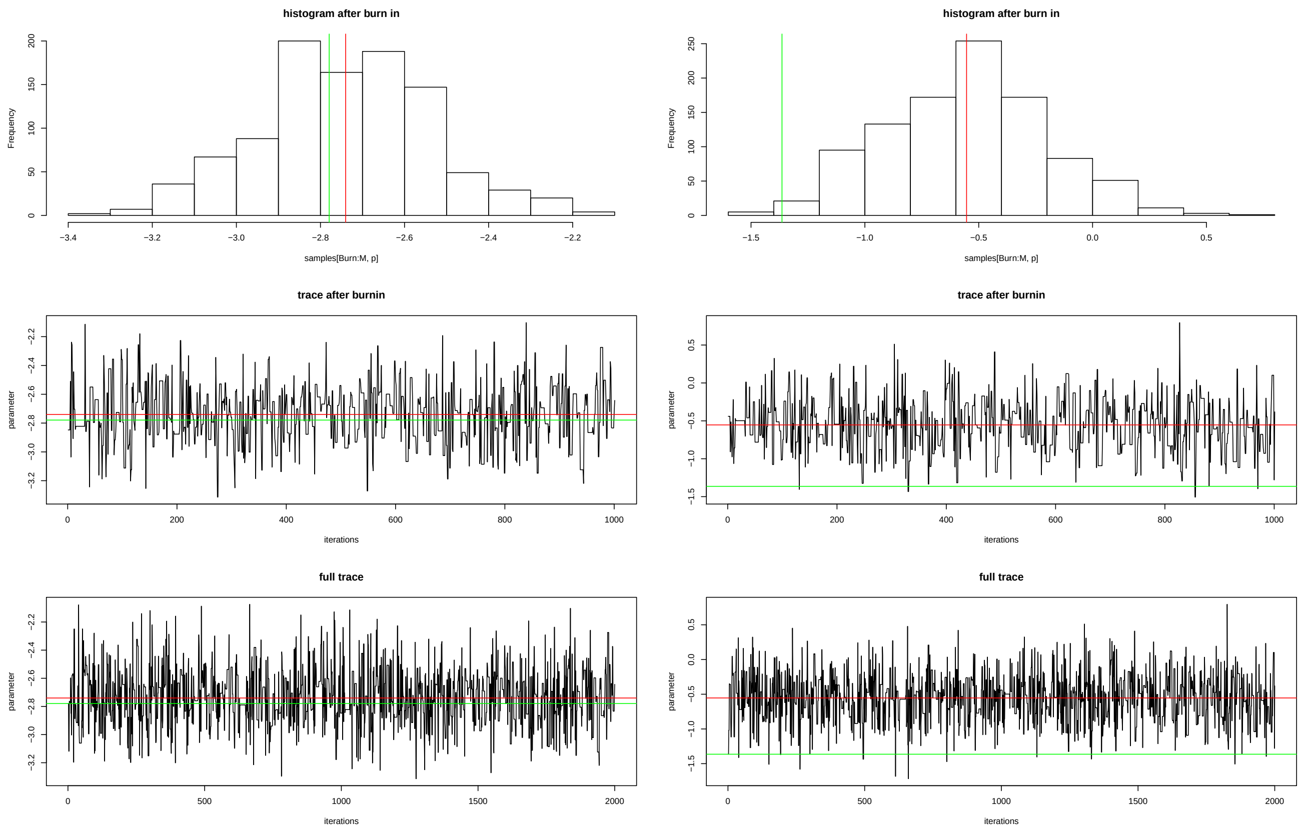

histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations


histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations



histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations


histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations


histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations
histogram after burn in
Frequency
samples[Burn:M, p]
trace after burnin
parameter
iterations
full trace
parameter
iterations

5 PGAS Implementation

5.1 Over time T

5.1.1 The Model:

parameter vector: for $d = 1, \dots, D$

$$\boldsymbol{\theta_d} = (\phi_d, \boldsymbol{\beta_{Z,d}}, \boldsymbol{\beta_{W,d}}, u_d)$$

state equations: for $d = 1, \dots, D$

$$\begin{aligned}
x_{t,d} &= \phi_d x_{t-1,d} + \boldsymbol{z_t}^\top \boldsymbol{\beta_{Z,d}} + \boldsymbol{w}^\top \boldsymbol{V_t} \boldsymbol{\beta_{W,d}} + u_d + \epsilon_t \ , \epsilon_t \sim N(0, \sigma_\epsilon^2) \\
p(x_{t,d} | x_{t-1,d}, \boldsymbol{\theta_d}) &= N(x_{t,d} | \mu_{x_d}, \sigma_{x_d}^2) \\
\mu_{x_d} &= \phi_d x_{t-1,d} + \boldsymbol{z_t}^\top \boldsymbol{\beta_{Z,d}} + \boldsymbol{w}^\top \boldsymbol{V_t} \boldsymbol{\beta_{W,d}} + u_d \\
\sigma_{x_d}^2 &= \sigma_\epsilon^2 \\
\Rightarrow \boldsymbol{x_t} &= (x_{t,1}, \dots, x_{t,D})
\end{aligned}$$

measurement equations

$$\begin{aligned}
\boldsymbol{y_t} &= B\left(\exp(\boldsymbol{x_t})\right)^{-1} \times \prod_{d=1}^{D} (y_{t,d})^{\exp((x_{t,d}-1)} \\
p(\boldsymbol{y_t} | \boldsymbol{x_t}) &= \mathbf{Dir}\left(\boldsymbol{y_t} | \exp(\boldsymbol{x_t})\right)
\end{aligned}$$

In Detail:

5.1.2 PGAS:

$s = 0$ Initialization: Set $\boldsymbol{\theta}(0)$ and $X^{\mathcal{R}}_{1:T}(0)$ arbitrarily or by sampling from a SMC-AS pre-run targeting $p(x_{1:T}|y_{1:T}, \boldsymbol{\theta}(0))$ i.e. $X^{\mathcal{R}}_{1:T}(0) \sim \widehat{p}(x_{1:T}|y_{1:T}, \boldsymbol{\theta}(0))$.

$s \geq 1$ Sample between $p(\boldsymbol{\theta}|x_{1:T}, y_{1:T})$ and a SMC-AS approx. to $p(x_{1:T}|y_{1:T}, \boldsymbol{\theta})$:

I. Sample $\boldsymbol{\theta}(s) \sim p(\cdot|X^{\mathcal{R}}_{1:T}(s-1), y_{1:T})$.

II. Run a SMC-AS targeting $p(x_{1:T}|y_{1:T}, \boldsymbol{\theta}(s))$ conditional on $X^{\mathcal{R}}_{1:T}(s-1)$ following the ALGORITHM below .

III. Sample $X^{\mathcal{R}}_{1:T}(s) \sim \widehat{p}(\cdot|y_{1:T}, \boldsymbol{\theta}(s))$.

In Detail:

5.1.3 Conditional SMC/BPF with AS:

`Algorithm:`

`START` <u>*For* $t = 1$*:*</u>

1. For $i = 1, \ldots, M-1$:
   Sample $X_1^i \sim q_1(\cdot)$
2. Conditioning: For $i = M$:
   Set $X_1^M = X_1^{\mathcal{R}}$
3. For $i = 1, \ldots, M$:
   Compute unnormalized $w_1(X_1^i)$ and normalized weights W_1^i.

   <u>*For* $t = 2$ *to* T*:*</u>

   For $i = 1, \ldots, M-1$:
4. Sample $A_t^i \sim \mathcal{M}(\cdot|\boldsymbol{W}_{t-1})$
5. Sample $X_t^i \sim q_t(\cdot|X_{t-1}^{A_t^i})$ for $i = 1, \ldots, M-1$

   For $i = M$:
6. Conditioning: Set $X_t^M = X_t^{\mathcal{R}}$

   For $i = 1, \ldots, M$:
7. AS: Compute $\widehat{W}_{n-1|N}^i$ and sample $A_t^M \sim \mathcal{M}(1, \boldsymbol{p} = \widehat{\boldsymbol{W}}_{n-1|N})$
8. Set $X_{1:t}^i \leftarrow \left(X_{1:t-1}^{A_t^i}, X_t^i\right)$ for $i = 1, \ldots, M$
9. Compute unnormalized $w_t(X_t^i)$ and normalized weights W_t^i

`END`

Above `Algorithm` allows to obtain a sample $X_{1:T}^*$ from the SMC-AS approximation to $p_{\boldsymbol{\theta}}(x_{1:N}|y_{1:T})$. To this end, draw an index k with probability $P(k = i) = W_T^i$ i.e. $k \sim \mathcal{M}(1, \boldsymbol{p} = \boldsymbol{W_T})$ and set $X_{1:T}^* = X_{1:T}^k$ where $\{W_T^i, X_{1:T}^i\}_{i=1}^M$ is obtained from step 8.-9. at iteration $t = T$. To shorten notation, denote this step by $X_{1:T}^* \sim \widehat{p}_{\boldsymbol{\theta}}(x_{1:T}|y_{1:T})$.

In Detail:

5.2 Over cross section N and time T

5.2.1 The Model:

In Detail:

5.2.2 PGAS:

$s = 0$ Initialization: Set $\boldsymbol{\theta}(0)$ and $X^{\mathcal{R}}_{1:N,1:T}(0)$ arbitrarily (or by sampling from a SMC-AS pre-run targeting $p_{\boldsymbol{\theta}(0)}(x_{1:N,1:T}|y_{1:N,1:T})$ i.e. $X^{\mathcal{R}}_{1:N,1:T}(0) \sim \widehat{p}_{\boldsymbol{\theta}(0)}(x_{1:N,1:T}|y_{1:N,1:T})$).

$s \geq 1$ Sample between $p(\boldsymbol{\theta}|x_{1:N,1:T}, y_{1:N,1:T})$ and a SMC-AS approx. to $p_{\boldsymbol{\theta}}(x_{1:N,1:T}|y_{1:N,1:T})$:

I. Sample $\boldsymbol{\theta}(s) \sim p(\cdot|X^{\mathcal{R}}_{1:N,1:T}(s-1), y_{1:N,1:T})$.

II. Run a SMC-AS targeting $p_{\boldsymbol{\theta}(s)}(x_{1:N,1:T}|y_{1:N,1:T})$ conditional on $X^{\mathcal{R}}_{1:N,1:T}(s-1)$ following the Algorithm below .

III. Sample $X^{\mathcal{R}}_{1:N,1:T}(s) \sim \widehat{p}_{\boldsymbol{\theta}(s)}(\cdot|y_{1:N,1:T})$.

In Detail:

5.2.3 Conditional SMC/BPF with AS:

`Algorithm:`

`START`

<u>*For $n = 1, \ldots, N$:*</u>

<u>*For $t = 1$:*</u>

1. For $i = 1, \ldots, M-1$:
   Sample $X^i_{n,1} \sim q_1(\cdot)$
2. Conditioning: For $i = M$:
   Set $X^M_{n,1} = X^{\mathcal{R}}_{n,1}$
3. For $i = 1, \ldots, M$:
   Compute unnormalized $w_{n,1}(X^i_{n,1})$ and normalized weights $W^i_{n,1}$.

<u>*For $t = 2$ to T:*</u>

For $i = 1, \ldots, M-1$:

4. Sample $A^i_{n,t} \sim \mathcal{M}(\cdot|\boldsymbol{W}_{n,t-1})$
5. Sample $X^i_{n,t} \sim q_t(\cdot|X^{A^i_{n,t}}_{n,t-1})$ for $i = 1, \ldots, M-1$

   For $i = M$:

6. Conditioning: Set $X^M_{n,t} = X^{\mathcal{R}}_{n,t}$

   For $i = 1, \ldots, M$:

7. AS: Compute $\widehat{W}^i_{n-1|N}$ and sample $A^M_{n,t} \sim \mathcal{M}(1, \boldsymbol{p} = \widehat{\boldsymbol{W}}_{n-1|N})$
8. Set $X^i_{n,1:t} \leftarrow \left(X^{A^i_{n,t}}_{n,1:t-1}, X^i_{n,t}\right)$ for $i = 1, \ldots, M$
9. Compute unnormalized $w_{n,t}(X^i_{n,t})$ and normalized weights $W^i_{n,t}$

`END`

Above `Algorithm` allows to obtain a sample $X^*_{1:N,1:T}$ from the SMC-AS approximation to $p_{\boldsymbol{\theta}}(x_{1:N}|y_{1:N,1:T})$. To this end, draw an index k with probability $P(k = i) = W^i_{N,T}$ i.e. $k \sim \mathcal{M}(1, \boldsymbol{p} = \boldsymbol{W_{N,T}})$ and set $X^*_{1:N,1:T} = X^k_{1:N,1:T}$ where $\left\{W^i_{N,T}, X^i_{1:N,1:T}\right\}_{i=1}^M$ is obtained from step 8.-9. at iteration $n = N, t = T$. To shorten notation, denote this step by $X^*_{1:N,1:T} \sim \widehat{p}_{\boldsymbol{\theta}}(x_{1:N,1:T}|y_{1:N,1:T})$.

In Detail: