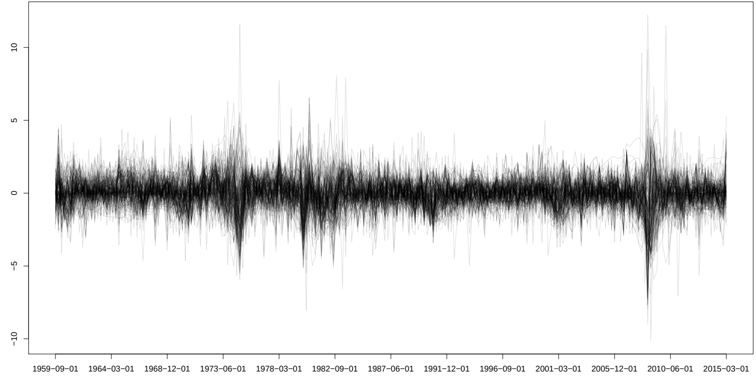
Bayesian Inference in Many Dimensions: Examples from Macroeconomics and Finance

Gregor Kastner | WU Vienna University of Economics and Business



based on joint work with

Martin Feldkircher
OeNB

Sylvia Frühwirth-Schnatter
WU Vienna

Florian Huber
University of Salzburg

Hedibert Freitas Lopes
Insper São Paulo

Bayesians Statistics in the Big Data Era | CIRM | Marseille Luminy | November 28, 2018

Three simple steps in 25 minutes (or so)

1. Step 1: Time-Varying Covariance
2. Step 2: Vector Autoregression with Time-Varying Covariance
3. Step 3: Thresholded Time-Varying Parameter Vector Autoregression with Time-Varying Covariance Matrix

Factor stochastic volatility

Suppose that $\boldsymbol{y}_t \in \mathbb{R}^m$, $t = 1, \ldots, T$,

$$\boldsymbol{y}_t \sim \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{\Omega}_t),$$

where

$$\boldsymbol{\Omega}_t = \boldsymbol{\Lambda} \boldsymbol{V}_t \boldsymbol{\Lambda} + \boldsymbol{\Sigma}_t$$

and

- $\boldsymbol{\Sigma}_t = \text{diag}(\sigma^2_{1t}, \ldots, \sigma^2_{mt})$ and $\boldsymbol{V}_t = \text{diag}(\sigma^2_{m+1,t}, \ldots, \sigma^2_{m+q,t})$
- $\boldsymbol{\Lambda}$: $m \times q$ factor loadings matrix
- AR(1) processes for the log variances (non-linear state space model)

Factor stochastic volatility

Suppose that $\boldsymbol{y}_t \in \mathbb{R}^m$, $t = 1, \ldots, T$,

$$\boldsymbol{y}_t \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Omega}_t),$$

where

$$\boldsymbol{\Omega}_t = \boldsymbol{\Lambda} \boldsymbol{V}_t \boldsymbol{\Lambda} + \boldsymbol{\Sigma}_t$$

and

> $\boldsymbol{\Sigma}_t = \text{diag}(\sigma^2_{1t}, \ldots, \sigma^2_{mt})$ and $\boldsymbol{V}_t = \text{diag}(\sigma^2_{m+1,t}, \ldots, \sigma^2_{m+q,t})$

> $\boldsymbol{\Lambda}$: $m \times q$ factor loadings matrix

> AR(1) processes for the log variances (non-linear state space model)

Known as the factor stochastic volatility model (see e.g. Pitt and Shephard, 1999; Aguilar and West, 2000; Kastner et al., 2017) with latent variable representation

$$\boldsymbol{\varepsilon}_t \sim \mathcal{N}_m(\boldsymbol{\Lambda} \boldsymbol{f}_t, \boldsymbol{\Sigma}_t), \quad \boldsymbol{f}_t \sim \mathcal{N}_q(\mathbf{0}, \boldsymbol{V}_t)$$

> Off-diagonal entries of $\boldsymbol{\Omega}_t$ exclusively stem from the volatilities of the q factors

> Diagonal entries of $\boldsymbol{\Omega}_t$ are allowed to feature idiosyncratic deviations driven by the elements of $\boldsymbol{\Sigma}_t$

> Reduces the number of free elements in $\boldsymbol{\Omega}_t$ from $m(m+1)/2$ to $m(q+1)$

Sparse factor models

Factor models are a sparse representation of $\mathbf{\Omega}$ and $\mathbf{\Omega}^{-1}$. To achieve additional sparsity, use shrinkage priors a.k.a. penalized likelihood.

- Point mass priors
  - Basic factor model (West, 2003; Carvalho et al., 2008; Frühwirth-Schnatter and Lopes, 2018)
  - Bayesian dedicated factor analysis (Conti et al., 2014)
  - Sparse dynamic factor models (Kaufmann and Schumacher, 2018)
- Continuous shrinkage priors, e.g. in sparse Bayesian infinite factor models (Bhattacharya and Dunson, 2011)
- Latent thresholding approaches (Nakajima and West, 2013a; Zhou et al., 2014)
- ...

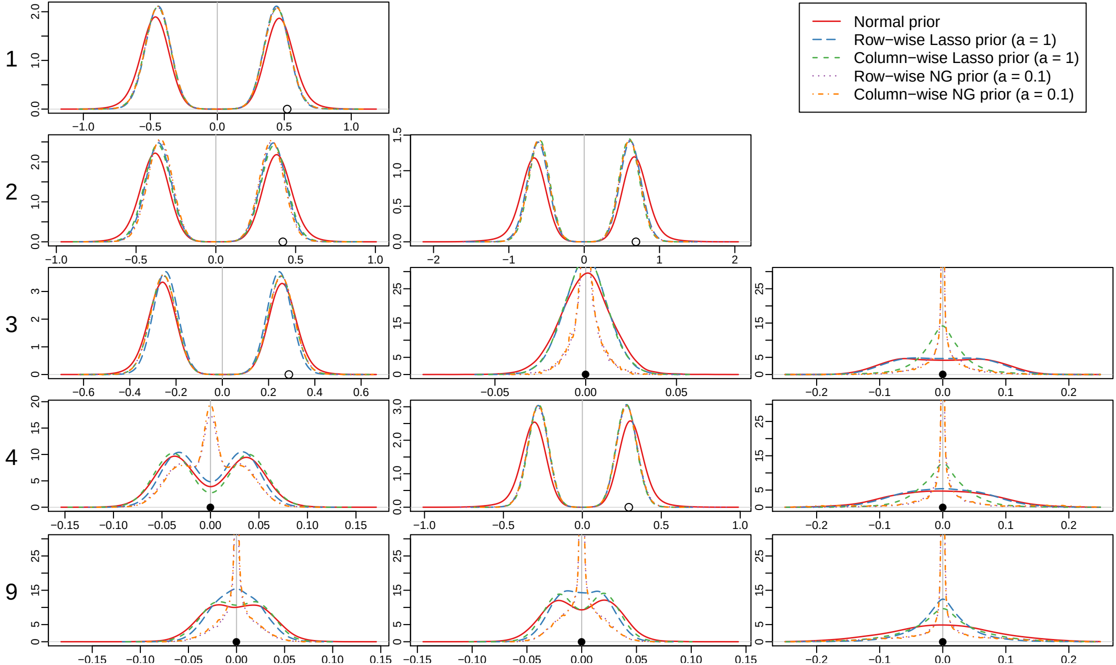
Sparse factor SV models

Use continuous shrinkage priors such as the Normal-Gamma prior (Griffin and Brown, 2010):

$$\Lambda_{ij}|\lambda_j, \tau_{ij} \sim \mathcal{N}\left(0, \tau_{ij}^2/\lambda_j^2\right), \quad \lambda_j^2 \sim \mathcal{G}\left(c, d\right), \quad \tau_{ij}^2 \sim \mathcal{G}\left(a, a\right).$$

- $\mathbb{V}(\Lambda_{ij}|\lambda_j^2) = 1/\lambda_j^2$
- Excess kurtosis of Λ_{ij} is $3/a$ if it exists
- Shrink globally (column-wise) through λ_j^2 (or row-wise through λ_i^2, industry-wise, ...)
- Adjust locally (element-wise) through τ_{ij}: $\tau_{ij} < 1$ more, $\tau_{ij} > 1$ less shrinkage
- Bayesian Lasso (Park and Casella, 2008) arises for $a = 1$

Marginal factor loadings posteriors, $r_{\text{true}} = 2$, $r = 3$, $m = 10$, $T = 1000$



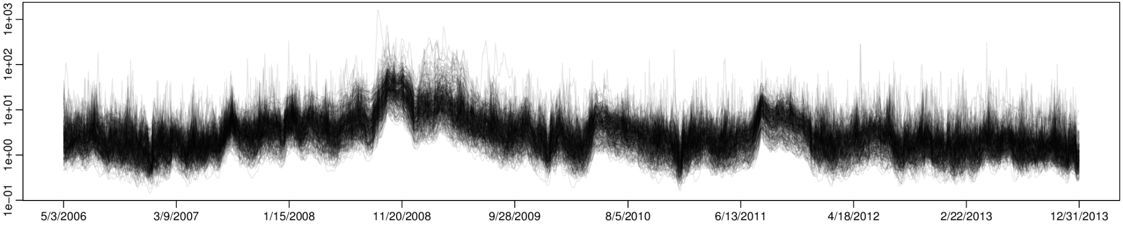
Application to S&P 500 members

- Only firms which have been listed from November 1994 onwards, resulting in $m = 300$ stock prices on 5001 days, ranging from 11/1/1994 to 12/31/2013.
- Data was obtained from Bloomberg Terminal in January 2014.
- Investigate $T = 5000$ demeaned percentage log-returns.
- Time-varying covariance matrix with 45150 nontrivial elements on 5000 days can be well explained by 4 factors with many factor loadings shrunken to 0.

Application to S&P 500 members

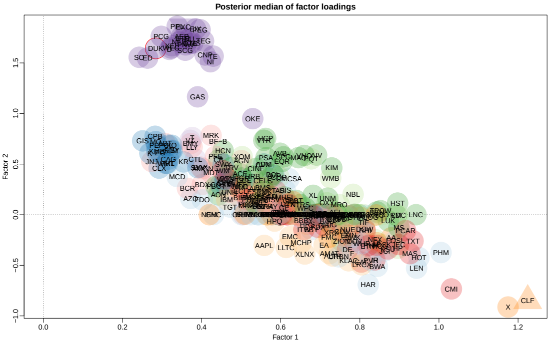
- Only firms which have been listed from November 1994 onwards, resulting in $m = 300$ stock prices on 5001 days, ranging from 11/1/1994 to 12/31/2013.
- Data was obtained from Bloomberg Terminal in January 2014.
- Investigate $T = 5000$ demeaned percentage log-returns.
- Time-varying covariance matrix with 45150 nontrivial elements on 5000 days can be well explained by 4 factors with many factor loadings shrunken to 0.

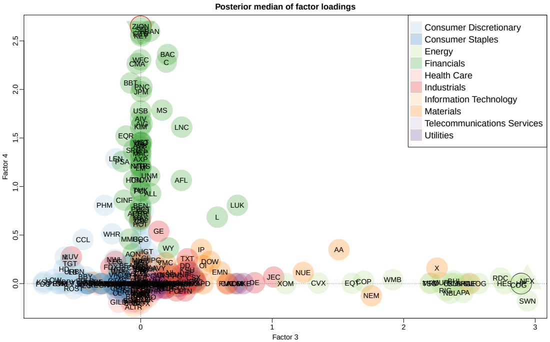
Marginal conditional variances (posterior mean, log scale)

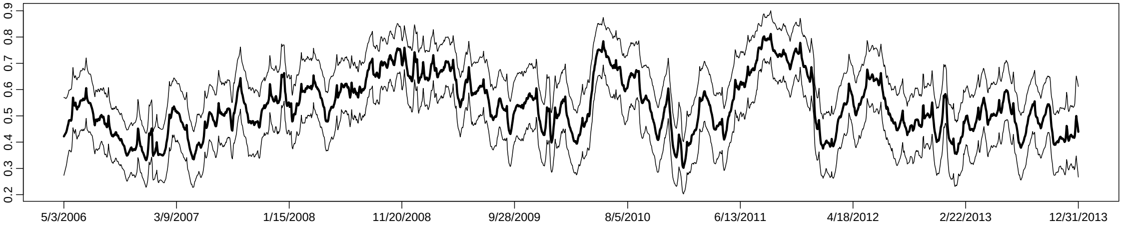


- Substantial co-movement (generally)
- Idiosyncratic deviations (at certain stretches in time)

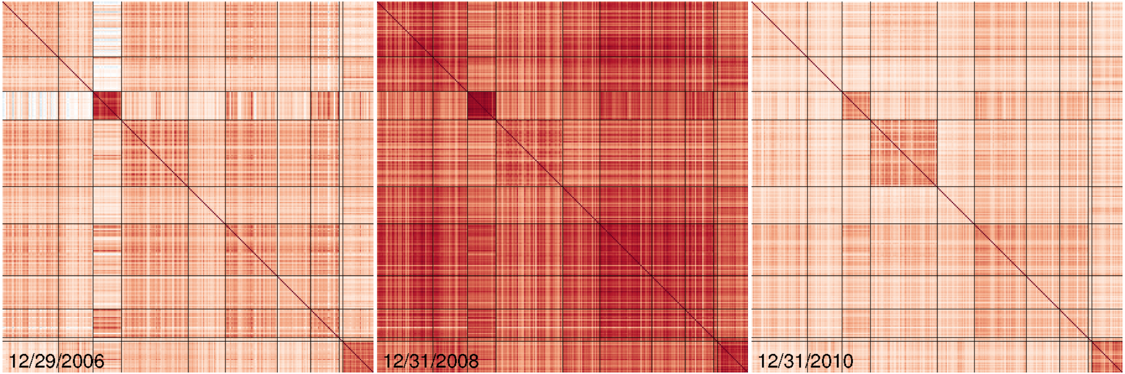
Median factor loadings and joint communalities (mean $\pm$ 2 sd)







Mean posterior correlations



Video

Works well for

- > density predictions
- > minimum variance portfolio constructions

(Kastner, 2018)

Three simple steps in 25 minutes (or so)

1. Step 1: Time-Varying Covariance
2. Step 2: Vector Autoregression with Time-Varying Covariance
3. Step 3: Thresholded Time-Varying Parameter Vector Autoregression with Time-Varying Covariance Matrix

The VAR model

Suppose that $\boldsymbol{y}_t \in \mathbb{R}^m, t = 1, \ldots, T$, follows a zero-mean heteroskedastic VAR(p) process,

$$\boldsymbol{y}_t = \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Omega}_t)$$

> $\boldsymbol{A}_j$ ($j = 1, \ldots, p$): $m \times m$ matrices of autoregressive coefficients

The VAR model

Suppose that $\boldsymbol{y}_t \in \mathbb{R}^m$, $t = 1, \dots, T$, follows a zero-mean heteroskedastic VAR(p) process,

$$\boldsymbol{y}_t = \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Omega}_t)$$

- $\boldsymbol{A}_j$ ($j = 1, \dots, p$): $m \times m$ matrices of autoregressive coefficients
- $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \dots, \boldsymbol{y}'_{t-p})'$ and a $m \times mp$ coefficient matrix $\boldsymbol{B} = (\boldsymbol{A}_1, \dots, \boldsymbol{A}_p)$ to rewrite the model more compactly as

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_m(\mathbf{0}, \boldsymbol{\Omega}_t).$$

(Including an intercept is straightforward but omitted here for simplicity of exposition)

Shrinking the coefficients

Dirichlet-Laplace prior (Bhattacharya et al., 2015) for $\boldsymbol{b} = \text{vec}(\boldsymbol{B})$. In what follows, we impose the DL prior on each of the $K = m^2 p$ elements of $\boldsymbol{b}$, for $j = 1, \ldots, K$,

$$b_j \sim \mathcal{N}(0, \psi_j \vartheta_j^2 \zeta)$$

> ψ_j are local scaling parameters $\psi_j \sim \text{Exp}(1/2)$
> $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_K)' \in \mathcal{S}^{K-1} = \{\boldsymbol{\vartheta} : \vartheta_j \geq 0, \sum_{j=1}^{K} \vartheta_j = 1\}$, $\boldsymbol{\vartheta} \sim \text{Dir}(a, \ldots, a)$
> ζ is a global shrinkage parameter $\zeta \sim \mathcal{G}(Ka, 1/2)$

Shrinking the coefficients

Dirichlet-Laplace prior (Bhattacharya et al., 2015) for $\boldsymbol{b} = \text{vec}(\boldsymbol{B})$. In what follows, we impose the DL prior on each of the $K = m^2 p$ elements of $\boldsymbol{b}$, for $j = 1, \ldots, K$,

$$b_j \sim \mathcal{N}(0, \psi_j \vartheta_j^2 \zeta)$$

- ψ_j are local scaling parameters $\psi_j \sim \text{Exp}(1/2)$
- $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_K)' \in \mathcal{S}^{K-1} = \{\boldsymbol{\vartheta} : \vartheta_j \geq 0, \sum_{j=1}^{K} \vartheta_j = 1\}$, $\boldsymbol{\vartheta} \sim \text{Dir}(a, \ldots, a)$
- ζ is a global shrinkage parameter $\zeta \sim \mathcal{G}(Ka, 1/2)$

Nice features:

- strong shrinkage with plenty of flexibility
- mimicking "real" spike and slab priors, but much lower computational burden
- implementation is simple, only requires one single structural hyperparameter a (smaller $a \Rightarrow$ stronger spike)
- good contraction guarantees in stylized models when $a = K^{-(1+\epsilon)}$ with ϵ small

MCMC at a glance

- Local shrinkage parameters $\psi_j|\bullet \sim iG(\vartheta_j\zeta/|b_j|, 1)$, $j = 1, \dots, K$ via efficient and stable rejection sampler from Hörmann and Leydold (2013), R-package `GIGrvg`
- Global shrinkage parameter $\zeta|\bullet \sim \mathcal{GIG}\left(K(a-1), 1, 2\sum_{j=1}^{K} |b_j|/\vartheta_j\right)$, `GIGrvg`
- Scaling parameters ϑ_j by first sampling L_j from $L_j|\bullet \sim \mathcal{GIG}(a-1, 1, 2|b_j|)$, and then setting $\vartheta_j = L_j/\sum_{i=1}^{K} L_i$ (Bhattacharya et al., 2015), `GIGrvg`
- Conditionally "univariate" SV states and parameters of volatility processes via auxiliary mixture sampling (Omori et al., 2007) with ASIS (Yu and Meng, 2011) via R-package `stochvol`
- Factors and factor loadings: "deep interweaving" (Kastner et al., 2017) via R-package `factorstochvol`
- VAR parameters, see next slide

Sampling the VAR parameters

Exploiting the data augmentation representation for the factor SV model, the model can be cast as a system of m conditionally unrelated regression models

$$z_{it} := y_{it} - \boldsymbol{\Lambda}_{i\bullet}\boldsymbol{f}_t = \boldsymbol{B}_{i\bullet}\boldsymbol{x}_t + \eta_{it}, \quad i = 1, \dots, m, \quad t = 1, \dots, T$$

Full conditional posteriors:

$$\boldsymbol{B}'_{i\bullet}|\bullet \sim \mathcal{N}(\boldsymbol{b}_i, \boldsymbol{Q}_i), \quad \boldsymbol{Q}_i = (\tilde{\boldsymbol{X}}'_i\tilde{\boldsymbol{X}}_i + \boldsymbol{\Phi}_i^{-1})^{-1}, \quad \boldsymbol{b}_i = \boldsymbol{Q}_i(\tilde{\boldsymbol{X}}'_i\tilde{\boldsymbol{z}}_i)$$

- $\boldsymbol{\Phi}_i$ is the respective $k \times k$ diagonal submatrix of $\boldsymbol{\Phi} = \zeta \times \text{diag}(\psi_1\vartheta_1^2, \dots, \psi_K\vartheta_K^2)$
- $\tilde{\boldsymbol{X}}_i$ is a $T \times k$ matrix with typical row t given by $\boldsymbol{X}_t/\sigma_{it}$
- $\tilde{\boldsymbol{z}}_i$ is a T-dimensional vector with the tth element given by z_{it}/σ_{it}.

Sampling the VAR parameters

Exploiting the data augmentation representation for the factor SV model, the model can be cast as a system of m conditionally unrelated regression models

$$z_{it} := y_{it} - \boldsymbol{\Lambda}_{i\bullet}\boldsymbol{f}_t = \boldsymbol{B}_{i\bullet}\boldsymbol{x}_t + \eta_{it}, \quad i = 1, \dots, m, \quad t = 1, \dots, T$$

Full conditional posteriors:

$$\boldsymbol{B}'_{i\bullet}|\bullet \sim \mathcal{N}(\boldsymbol{b}_i, \boldsymbol{Q}_i), \quad \boldsymbol{Q}_i = (\tilde{\boldsymbol{X}}'_i\tilde{\boldsymbol{X}}_i + \boldsymbol{\Phi}_i^{-1})^{-1}, \quad \boldsymbol{b}_i = \boldsymbol{Q}_i(\tilde{\boldsymbol{X}}'_i\tilde{\boldsymbol{z}}_i)$$

- $\boldsymbol{\Phi}_i$ is the respective $k \times k$ diagonal submatrix of $\boldsymbol{\Phi} = \zeta \times \text{diag}(\psi_1\vartheta_1^2, \dots, \psi_K\vartheta_K^2)$
- $\tilde{\boldsymbol{X}}_i$ is a $T \times k$ matrix with typical row t given by $\boldsymbol{X}_t/\sigma_{it}$
- $\tilde{\boldsymbol{z}}_i$ is a T-dimensional vector with the tth element given by z_{it}/σ_{it}.

$\Rightarrow$ allows for equation by equation estimation: each draw costs $O(m^4p^3 + Tm^3p^2)$ – as opposed to $O(m^6p^3)$ in the naive approach (cf. Carriero et al., 2015)

Yet another approach

In typical macro data $T \lesssim 200$ … even faster sampling is possible via an algorithm proposed by Bhattacharya et al. (2016) for univariate regressions, applied to each equation:

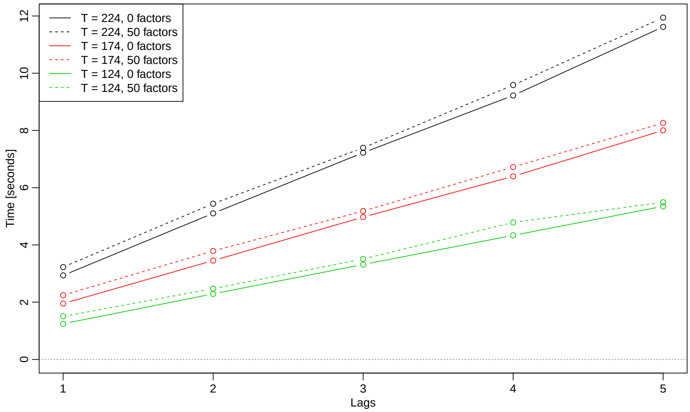
1. Sample independently $\boldsymbol{u}_i \sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{\Phi}_i)$ and $\boldsymbol{\delta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_T)$
2. Use $\boldsymbol{u}_i$ and $\boldsymbol{\delta}_i$ to construct $\boldsymbol{v}_i = \tilde{\boldsymbol{X}}_i \boldsymbol{u}_i + \boldsymbol{\delta}_i$
3. Solve $(\tilde{\boldsymbol{X}}_i \boldsymbol{\Phi}_i \tilde{\boldsymbol{X}}_i' + \boldsymbol{I}_T)\boldsymbol{w}_i = (\tilde{\boldsymbol{z}}_i - \boldsymbol{v}_i)$ for $\boldsymbol{w}_i$
4. Set $\boldsymbol{B}_{i\bullet}' = \boldsymbol{u}_i + \boldsymbol{\Phi}_i \tilde{\boldsymbol{X}}_i' \boldsymbol{w}_i$

Yet another approach

In typical macro data $T \lesssim 200$ …even faster sampling is possible via an algorithm proposed by Bhattacharya et al. (2016) for univariate regressions, applied to each equation:

1. Sample independently $\boldsymbol{u}_i \sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{\Phi}_i)$ and $\boldsymbol{\delta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_T)$
2. Use $\boldsymbol{u}_i$ and $\boldsymbol{\delta}_i$ to construct $\boldsymbol{v}_i = \tilde{\boldsymbol{X}}_i \boldsymbol{u}_i + \boldsymbol{\delta}_i$
3. Solve $(\tilde{\boldsymbol{X}}_i \boldsymbol{\Phi}_i \tilde{\boldsymbol{X}}_i' + \boldsymbol{I}_T)\boldsymbol{w}_i = (\tilde{\boldsymbol{z}}_i - \boldsymbol{v}_i)$ for $\boldsymbol{w}_i$
4. Set $\boldsymbol{B}_{i\bullet}' = \boldsymbol{u}_i + \boldsymbol{\Phi}_i \tilde{\boldsymbol{X}}_i' \boldsymbol{w}_i$

Cost is now $O(m^2 T^2 p)$, e.g. for $m = 215$ we have:

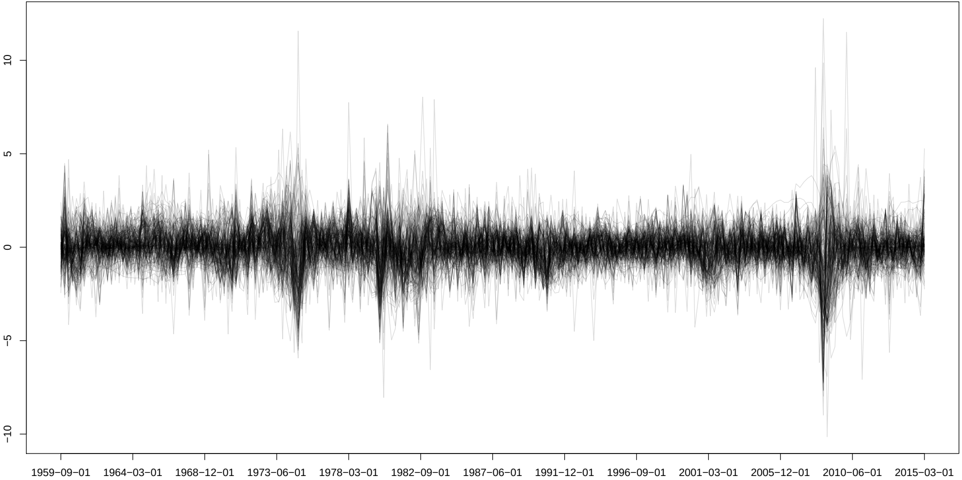


…and thanks to Aki's I now know that there is even more room for improvement (Nishimura and Suchard, 2018; Zhang et al., 2018)!

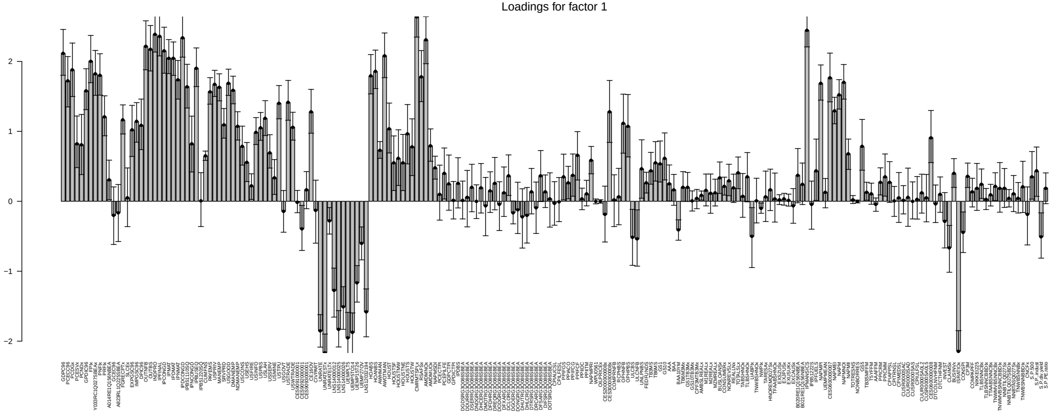
Modeling the US economy

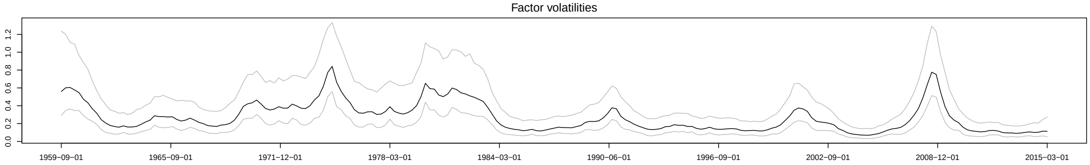
Quarterly dataset from McCracken and Ng (2016) including suggested transformations

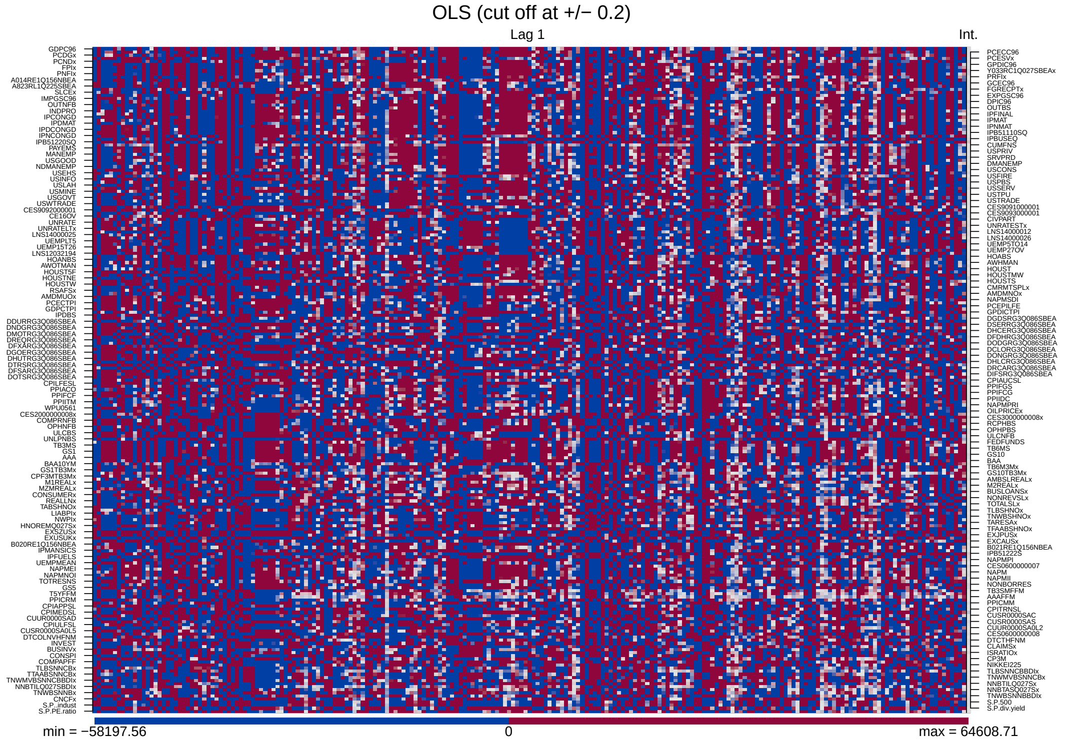
- Ranging from 1959:Q1 to 2015:Q4, $m = 215$
- Component-wise standardization (zero mean, variance one)
- For presentation purposes: One lag, one factor



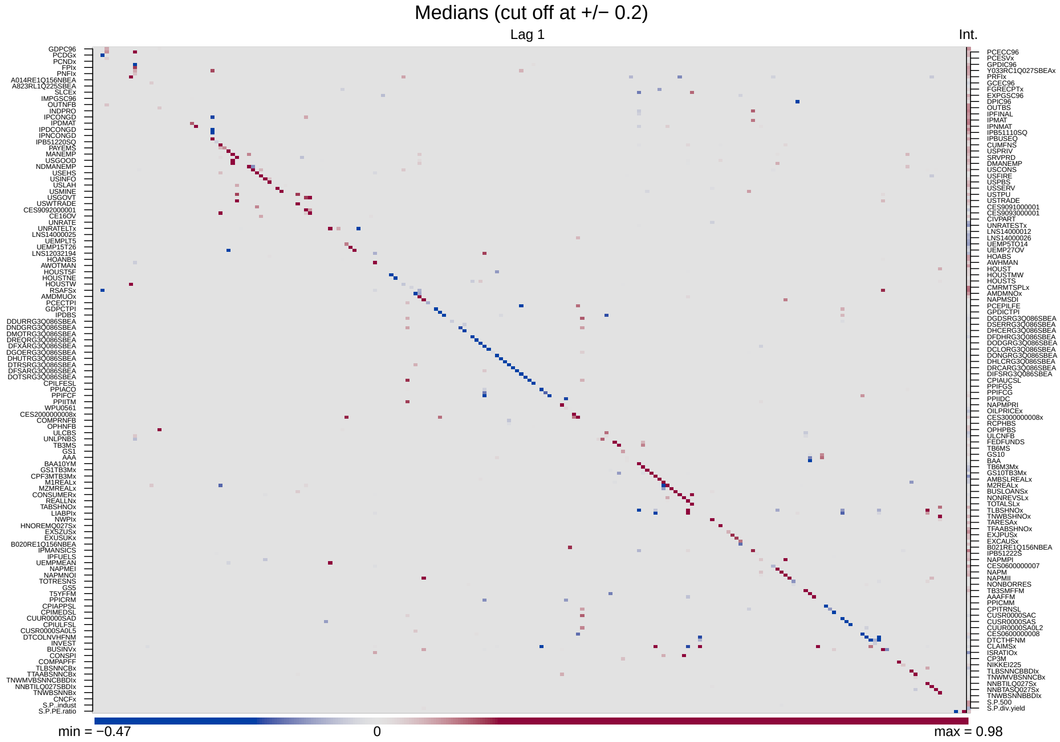
Factor loadings and factor volatilities







OLS (cut off at +/− 0.2)
Lag 1
Int.
GDPC96
PCDGx
PCNDx
FPIx
PNFIx
A014RE1Q156NBEA
A823RL1Q225SBEA
SLCEx
IMPGSC96
OUTNFB
INDPRO
IPCONGD
IPDMAT
IPDCONGD
IPNCONGD
IPB51220SQ
PAYEMS
MANEMP
USGOOD
NDMANEMP
USEHS
USINFO
USLAH
USMINE
USGOVT
USWTRADE
CES9092000001
CE16OV
UNRATE
UNRATELTx
LNS14000025
UEMPLT5
UEMP15T26
LNS12032194
HOANBS
AWOTMAN
HOUST5F
HOUSTNE
HOUSTW
RSAFSx
AMDMUOx
PCECTPI
GDPCTPI
IPDBS
DDURRG3Q086SBEA
DNDGRG3Q086SBEA
DMOTRG3Q086SBEA
DREQRG3Q086SBEA
DFXARG3Q086SBEA
DGOERG3Q086SBEA
DHUTRG3Q086SBEA
DTRSRG3Q086SBEA
DFSARG3Q086SBEA
DOTSRG3Q086SBEA
CPILFESL
PPIACO
PPIFCF
PPIITM
WPU0561
CES2000000008x
COMPRNFB
OPHNFB
ULCBS
UNLPNBS
TB3MS
GS1
AAA
BAA10YM
GS1TB3Mx
CPF3MTB3Mx
M1REALx
MZMREALx
CONSUMERx
REALLNx
TABSHNOx
LIABPIx
NWPIx
HNOREMQ027Sx
EXSZUSx
EXUSUKx
B020RE1Q156NBEA
IPMANSICS
IPFUELS
UEMPMEAN
NAPMEI
NAPMNOI
TOTRESNS
GS5
T5YFFM
PPICRM
CPIAPPSL
CPIMEDSL
CUUR0000SAD
CPIULFSL
CUSR0000SA0L5
DTCOLNVHFNM
INVEST
BUSINVx
CONSPI
COMPAPFF
TLBSNNCBx
TTAABSNNCBx
TNWMVBSNNCBBDIx
NNBTILQ027SBDIx
TNWBSNNBx
CNCFx
S.P..indust
S.P.PE.ratio
PCECC96
PCESVx
GPDIC96
Y033RC1Q027SBEAx
PRFIx
GCEC96
FGRECPTx
EXPGSC96
DPIC96
OUTBS
IPFINAL
IPMAT
IPNMAT
IPB51110SQ
IPBUSEQ
CUMFNS
USPRIV
SRVPRD
DMANEMP
USCONS
USFIRE
USPBS
USSERV
USTPU
USTRADE
CES9091000001
CES9093000001
CIVPART
UNRATESTx
LNS14000012
LNS14000026
UEMP5TO14
UEMP27OV
HOABS
AWHMAN
HOUST
HOUSTMW
HOUSTS
CMRMTSPLx
AMDMNOx
NAPMSDI
PCEPILFE
GPDICTPI
DGDSRG3Q086SBEA
DSERRG3Q086SBEA
DHCERG3Q086SBEA
DFDHRG3Q086SBEA
DODGRG3Q086SBEA
DCLORG3Q086SBEA
DONGRG3Q086SBEA
DHLCRG3Q086SBEA
DRCARG3Q086SBEA
DIFSRG3Q086SBEA
CPIAUCSL
PPIFGS
PPIFCG
PPIIDC
NAPMPRI
OILPRICEx
CES3000000008x
RCPHBS
OPHPBS
ULCNFB
FEDFUNDS
TB6MS
GS10
BAA
TB6M3Mx
GS10TB3Mx
AMBSLREALx
M2REALx
BUSLOANSx
NONREVSLx
TOTALSLx
TLBSHNOx
TNWBSHNOx
TARESAx
TFAABSHNOx
EXJPUSx
EXCAUSx
B021RE1Q156NBEA
IPB51222S
NAPMPI
CES0600000007
NAPM
NAPMII
NONBORRES
TB3SMFFM
AAAFFM
PPICMM
CPITRNSL
CUSR0000SAC
CUSR0000SAS
CUUR0000SA0L2
CES0600000008
DTCTHFNM
CLAIMSx
ISRATIOx
CP3M
NIKKEI225
TLBSNNCBBDIx
TNWMVBSNNCBx
NNBTILQ027Sx
NNBTASQ027Sx
TNWBSNNBBDIx
S.P.500
S.P.div.yield
min = −58197.56
0
max = 64608.71


Medians (cut off at +/− 0.2)
Lag 1
Int.
GDPC96
PCDGx
PCNDx
FPIx
PNFIx
A014RE1Q156NBEA
A823RL1Q225SBEA
SLCEx
IMPGSC96
OUTNFB
INDPRO
IPCONGD
IPDMAT
IPDCONGD
IPNCONGD
IPB51220SQ
PAYEMS
MANEMP
USGOOD
NDMANEMP
USEHS
USINFO
USLAH
USMINE
USGOVT
USWTRADE
CES9092000001
CE16OV
UNRATE
UNRATELTx
LNS14000025
UEMPLT5
UEMP15T26
LNS12032194
HOANBS
AWOTMAN
HOUST5F
HOUSTNE
HOUSTW
RSAFSx
AMDMUOx
PCECTPI
GDPCTPI
IPDBS
DDURRG3Q086SBEA
DNDGRG3Q086SBEA
DMOTRG3Q086SBEA
DREQRG3Q086SBEA
DFXARG3Q086SBEA
DGOERG3Q086SBEA
DHUTRG3Q086SBEA
DTRSRG3Q086SBEA
DFSARG3Q086SBEA
DOTSRG3Q086SBEA
CPILFESL
PPIACO
PPIFCF
PPIITM
WPU0561
CES2000000008x
COMPRNFB
OPHNFB
ULCBS
UNLPNBS
TB3MS
GS1
AAA
BAA10YM
GS1TB3Mx
CPF3MTB3Mx
M1REALx
MZMREALx
CONSUMERx
REALLNx
TABSHNOx
LIABPIx
NWPIx
HNOREMQ027Sx
EXSZUSx
EXUSUKx
B020RE1Q156NBEA
IPMANSICS
IPFUELS
UEMPMEAN
NAPMEI
NAPMNOI
TOTRESNS
GS5
T5YFFM
PPICRM
CPIAPPSL
CPIMEDSL
CUUR0000SAD
CPIULFSL
CUSR0000SA0L5
DTCOLNVHFNM
INVEST
BUSINVx
CONSPI
COMPAPFF
TLBSNNCBx
TTAABSNNCBx
TNWMVBSNNCBBDIx
NNBTILQ027SBDIx
TNWBSNNBx
CNCFx
S.P..indust
S.P.PE.ratio
PCECC96
PCESVx
GPDIC96
Y033RC1Q027SBEAx
PRFIx
GCEC96
FGRECPTx
EXPGSC96
DPIC96
OUTBS
IPFINAL
IPMAT
IPNMAT
IPB51110SQ
IPBUSEQ
CUMFNS
USPRIV
SRVPRD
DMANEMP
USCONS
USFIRE
USPBS
USSERV
USTPU
USTRADE
CES9091000001
CES9093000001
CIVPART
UNRATESTx
LNS14000012
LNS14000026
UEMP5TO14
UEMP27OV
HOABS
AWHMAN
HOUST
HOUSTMW
HOUSTS
CMRMTSPLx
AMDMNOx
NAPMSDI
PCEPILFE
GPDICTPI
DGDSRG3Q086SBEA
DSERRG3Q086SBEA
DHCERG3Q086SBEA
DFDHRG3Q086SBEA
DODGRG3Q086SBEA
DCLORG3Q086SBEA
DONGRG3Q086SBEA
DHLCRG3Q086SBEA
DRCARG3Q086SBEA
DIFSRG3Q086SBEA
CPIAUCSL
PPIFGS
PPIFCG
PPIIDC
NAPMPRI
OILPRICEx
CES3000000008x
RCPHBS
OPHPBS
ULCNFB
FEDFUNDS
TB6MS
GS10
BAA
TB6M3Mx
GS10TB3Mx
AMBSLREALx
M2REALx
BUSLOANSx
NONREVSLx
TOTALSLx
TLBSHNOx
TNWBSHNOx
TARESAx
TFAABSHNOx
EXJPUSx
EXCAUSx
B021RE1Q156NBEA
IPB51222S
NAPMPI
CES0600000007
NAPM
NAPMII
NONBORRES
TB3SMFFM
AAAFFM
PPICMM
CPITRNSL
CUSR0000SAC
CUSR0000SAS
CUUR0000SA0L2
CES0600000008
DTCTHFNM
CLAIMSx
ISRATIOx
CP3M
NIKKEI225
TLBSNNCBBDIx
TNWMVBSNNCBx
NNBTILQ027Sx
NNBTASQ027Sx
TNWBSNNBBDIx
S.P.500
S.P.div.yield
min = −0.47
0
max = 0.98

Three simple steps in 25 minutes (or so)

1. Step 1: Time-Varying Covariance
2. Step 2: Vector Autoregression with Time-Varying Covariance
3. Step 3: Thresholded Time-Varying Parameter Vector Autoregression with Time-Varying Covariance Matrix

Making the VAR coefficients time-varying I

- Especially in larger dimensional contexts (think of time-varying parameter VARs), allowing for too much flexibility quickly leads to overfitting issues
- Questions such as "should all parameters or only a subset of parameters drift over time?" or "do parameters evolve according to a random walk process or are their dynamics better captured by model that implies relatively few, large breaks?" typically arise
- Literature offers several solutions (e.g. McCulloch and Tsay, 1993; Gerlach et al., 2000; Giordani and Kohn, 2008; Koop et al., 2009; Frühwirth-Schnatter and Wagner, 2010; Koop and Korobilis, 2012; Koop and Korobilis, 2013; Belmonte et al., 2014; Kalli and Griffin, 2014; Eisenstat et al., 2016; Bitto and Frühwirth-Schnatter, 2018)

Making the VAR coefficients time-varying II

- Mixture innovation models provide a flexible means of shrinking time variation of the latent states within a state space model
  - Generally straightforward to implement and quite flexible (allows for large swings in the parameters over certain time intervals)
  - However, not feasible in large applications (think of densely parameterized time varying parameter VAR models)
- Combine the literature on threshold and latent threshold models (Nakajima and West, 2013b; Neelon and Dunson, 2004) with the literature on mixture innovation models (McCulloch and Tsay, 1993; Carter and Kohn, 1994; Gerlach et al., 2000; Koop and Potter, 2007; Giordani and Kohn, 2008)
- We assume that the indicator that controls the mixture component used is determined by the absolute period-on-period change of the states $\rightarrow$ small changes are effectively set equal to zero
- Provides a great deal of flexibility while keeping the additional computational burden tractable

Econometric framework

We consider the following dynamic regression model (TVP model),

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + u_t, \;\; u_t \sim \mathcal{N}(0, \sigma^2),$$
$$\beta_{jt} = \beta_{j,t-1} + e_{jt}, \;\; e_{jt} \sim \mathcal{N}(0, \vartheta_j),$$

where

> $\boldsymbol{x}_t$ is a $K \times 1$ vector of explanatory variables

> $\boldsymbol{\beta}_t$ is a $K \times 1$ vector of time varying regression coefficients

> σ^2 and ϑ_j $(j = 1, \ldots, K)$ are innovation variances

This specification assumes that parameters evolve according to a random walk with small movements governed by ϑ_j.

The threshold mixture innovation model (Huber et al., forthcoming)

By contrast, the threshold mixture innovation model (TTVP model) assumes that

$$e_{jt} \sim \mathcal{N}(0, \theta_{jt}),$$
$$\theta_{jt} = s_{jt}\vartheta_{j1} + (1 - s_{jt})\vartheta_{j0},$$

with $\vartheta_{j1} \gg \vartheta_{j0}$.

As in standard mixture innovation models, the indicators s_{jt} are (unconditionally) Bernoulli distributed. In contrast to those models, we however assume that

$$s_{jt}|\Delta\beta_{jt}, d_j = \begin{cases} 1 & \text{if} \quad |\Delta\beta_{jt}| > d_j, \\ 0 & \text{if} \quad |\Delta\beta_{jt}| \leq d_j, \end{cases}$$

where d_j is a coefficient-specific threshold to be estimated. Note that we "only" have one extra parameter per equation (and the indicators are conditionally deterministic).

Can use standard MCMC with one additional griddy Gibbs (Ritter and Tanner, 1992) step per equation.

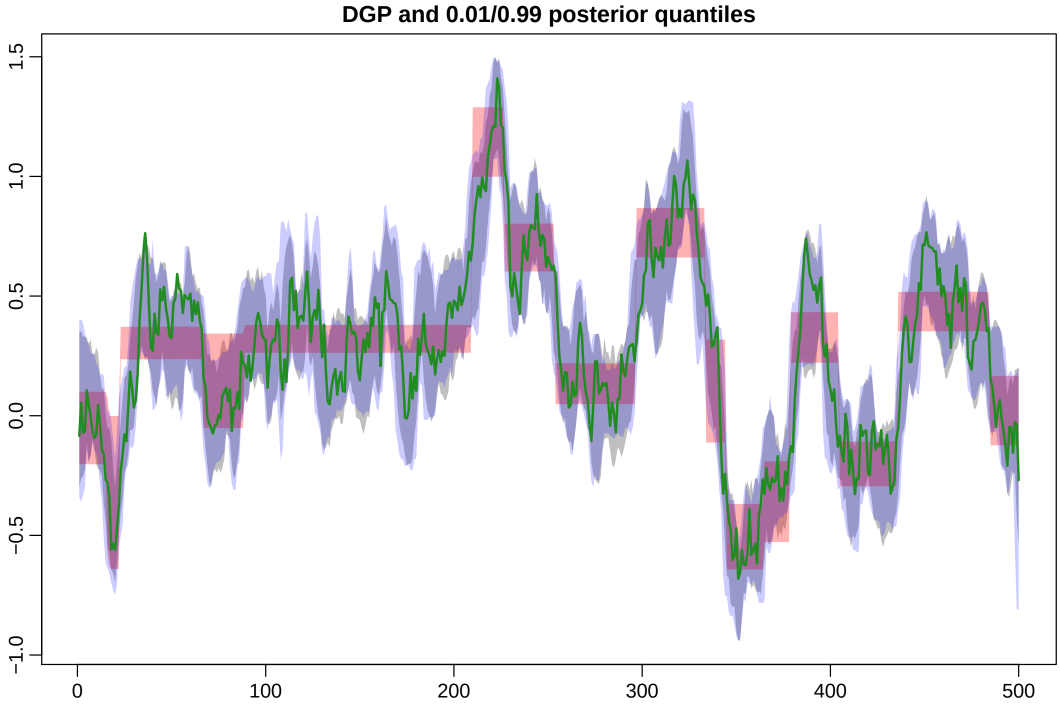
Four illustrative examples

Consider four simple DGPs with $K = 1$: TVP, many, few, and no breaks. Independently for all t, we generate $x_{1t} \sim \mathcal{U}(-1, 1)$ and set $\beta_{1,0} = 0$.

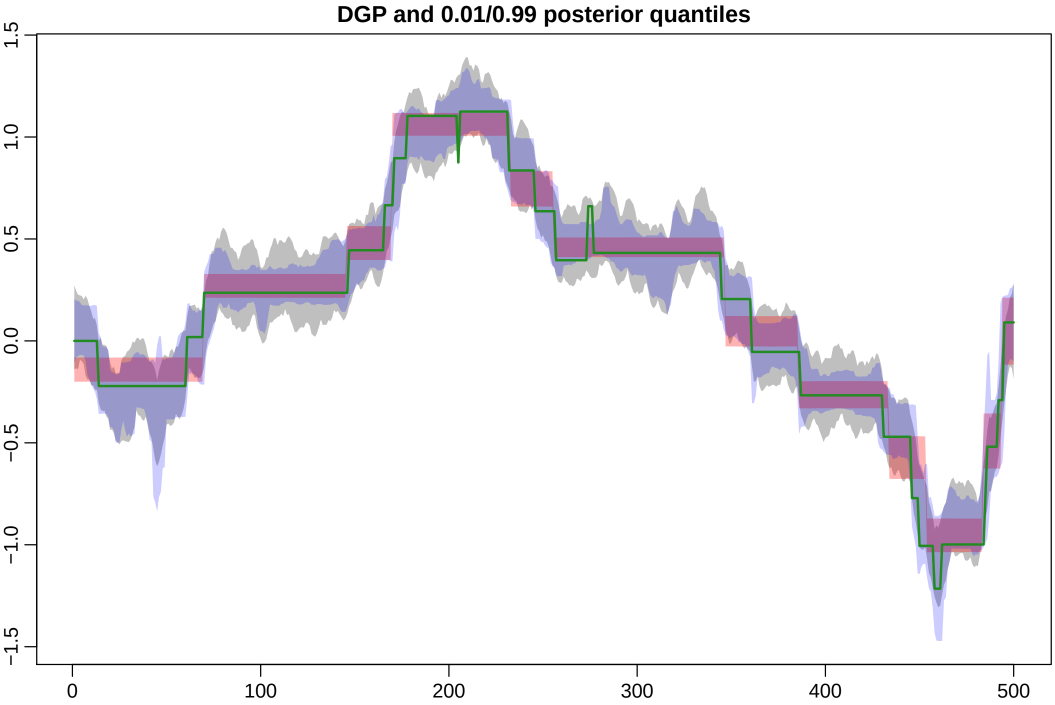
In order to assess how different models perform in recovering the latent processes, we run:

- standard TVP model
- mixture innovation model estimated using the algorithm outlined in Gerlach et al. (GCK, 2000)
- TTVP

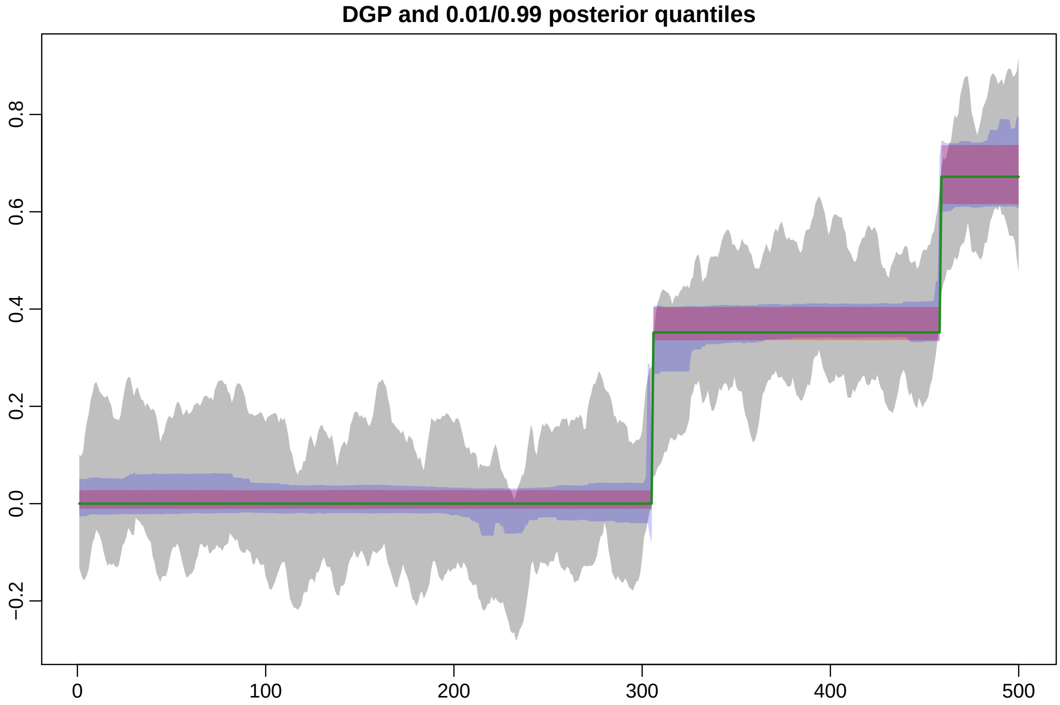
To ease comparison between the models, we impose a similar prior setup for all models.

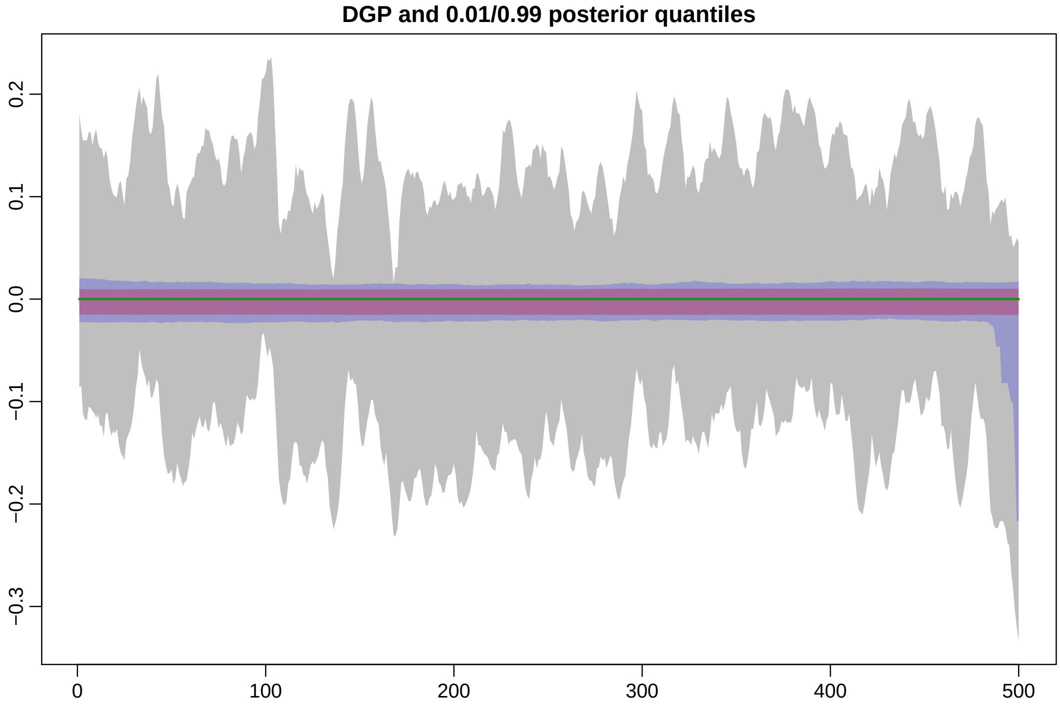


DGP and 0.01/0.99 posterior quantiles
−1.0
−0.5
0.0
0.5
1.0
1.5
0
100
200
300
400
500



DGP and 0.01/0.99 posterior quantiles
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
100
200
300
400
500


DGP and 0.01/0.99 posterior quantiles
0.8
0.6
0.4
0.2
0.0
−0.2
0
100
200
300
400
500


DGP and 0.01/0.99 posterior quantiles
0.2
0.1
0.0
−0.1
−0.2
−0.3
0
100
200
300
400
500

Generalization to the multivariate case

It's straightforward to generalize the framework to a TTVP-VAR-SV model:

$$\boldsymbol{y}_t = \boldsymbol{B}_{1t}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{B}_{Pt}\boldsymbol{y}_{t-P} + \boldsymbol{u}_t,$$

where

- $\boldsymbol{B}_{pt}$ $(p = 1, \ldots, P)$ are $m \times m$ matrices of thresholded dynamic coefficients
- $\boldsymbol{u}_t \sim \mathcal{N}(\boldsymbol{0}_m, \boldsymbol{\Sigma}_t)$ with $\boldsymbol{\Sigma}_t = \boldsymbol{V}_t\boldsymbol{H}_t\boldsymbol{V}_t'$ being a time-varying variance-covariance matrix
- $\boldsymbol{V}_t$ is lower triangular with unit diagonal; free elements have thresholded dynamic coefficients
- $\boldsymbol{H}_t = \text{diag}(e^{h_{1t}}, \ldots, e^{h_{mt}})$ with $h_{it} = \mu_i + \rho_i(h_{i,t-1} + \mu_i) + \nu_{it},\ \nu_{it} \sim \mathcal{N}(0, \zeta_i)$

This structure imposes order-dependence but allows to estimate the model equation-by-equation with the same techniques as before.

Forecasting the US term structure of interest rates

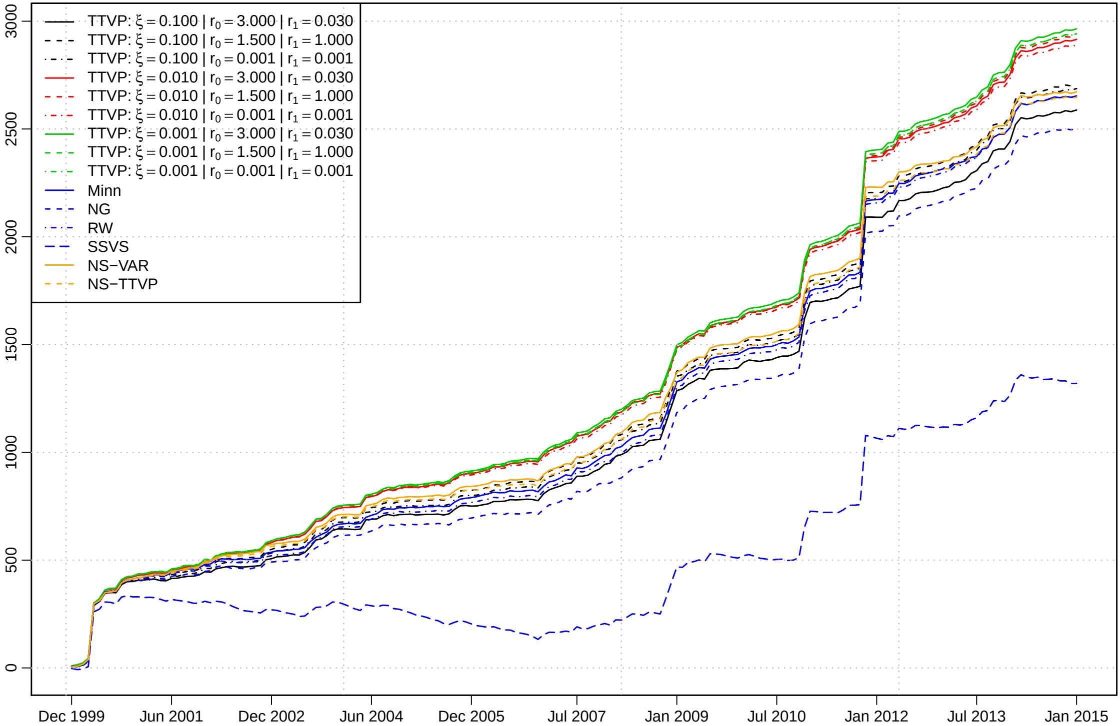
We use monthly Fama-Bliss zero coupon yields obtained from the CRSP database as well as the dataset described in Gürkaynak et al. (2007). The data spans the period from 1960:M01 to 2014:M12 and the maturities included are 1, 2, 3, 4, 5, 7, and 10 years. Moreover, we include 3 lags of the endogenous variables.

Forecasting the US term structure of interest rates

We use monthly Fama-Bliss zero coupon yields obtained from the CRSP database as well as the dataset described in Gürkaynak et al. (2007). The data spans the period from 1960:M01 to 2014:M12 and the maturities included are 1, 2, 3, 4, 5, 7, and 10 years. Moreover, we include 3 lags of the endogenous variables.

Competitors (all models include SV):

- Benchmark: TVP-VAR as in Primiceri (2005).
- Three constant parameter VAR models:
  - Minnesota-type VAR (Doan et al., 1984)
  - Normal-Gamma (NG) VAR (Huber and Feldkircher, 2018)
  - VAR coupled with an SSVS prior (George et al., 2008)
- Nelson-Siegel (NS, Nelson and Siegel, 1987) VAR as in Diebold and Li (2006)
- NS-TTVP-VAR
- Random walk



TTVP: ξ = 0.100 | r0 = 3.000 | r1 = 0.030
TTVP: ξ = 0.100 | r0 = 1.500 | r1 = 1.000
TTVP: ξ = 0.100 | r0 = 0.001 | r1 = 0.001
TTVP: ξ = 0.010 | r0 = 3.000 | r1 = 0.030
TTVP: ξ = 0.010 | r0 = 1.500 | r1 = 1.000
TTVP: ξ = 0.010 | r0 = 0.001 | r1 = 0.001
TTVP: ξ = 0.001 | r0 = 3.000 | r1 = 0.030
TTVP: ξ = 0.001 | r0 = 1.500 | r1 = 1.000
TTVP: ξ = 0.001 | r0 = 0.001 | r1 = 0.001
Minn
NG
RW
SSVS
NS−VAR
NS−TTVP
0
500
1000
1500
2000
2500
3000
Dec 1999
Jun 2001
Dec 2002
Jun 2004
Dec 2005
Jul 2007
Jan 2009
Jul 2010
Jan 2012
Jul 2013
Jan 2015

Wrap-up

Step 1: Sparse factor SV models

- "Hybrid approach" to dynamic covariance modeling, i.e. parsimony through factor structure, plus additional shrinkage
- "Plug-and-play" friendly due to the modular nature of MCMC

Step 2: High-dimensional VARs

- Equation-by-equation estimation of VARs possible due to latent variable representation of residual covariance matrix
- Faster algorithm(s) for sampling VAR coefficients if $T \lesssim mp$
- Continuous global-local shrinkage priors (DL prior and friends) help to cure the curse of dimensionality and provide a viable alternative to Minnesota-type priors
- Multivariate SV is crucial for macro application and improves joint density predictions markedly

Step 3: TTVP

- Computationally feasible variant of a mixture innovation model
- Allows for jumps in the parameters if the conditional absolute change exceeds a threshold value to be estimated
- Works well for term structure modeling, especially during crisis times

References I

Aguilar, O. and M. West (2000). "Bayesian dynamic factor models and portfolio allocation". Journal of Business & Economic Statistics 18, pp. 338–357 (cit. on pp. 3, 4).

Belmonte, M. A., G. Koop, and D. Korobilis (2014). "Hierarchical Shrinkage in Time-Varying Parameter Models". Journal of Forecasting 33, pp. 80–94 (cit. on p. 27).

Bhattacharya, A. and D. B. Dunson (2011). "Sparse Bayesian Infinite Factor Models". Biometrika 98, pp. 291–309. DOI: 10.1093/biomet/asr013 (cit. on p. 5).

Bhattacharya, A., D. Pati, N. S. Pillai, and D. B. Dunson (2015). "Dirichlet–Laplace priors for optimal shrinkage". Journal of the American Statistical Association 110, pp. 1479–1490 (cit. on pp. 15–17).

Bhattacharya, A., A. Chakraborty, and B. K. Mallick (2016). "Fast sampling with Gaussian scale-mixture priors in high-dimensional regression". Biometrika 4, pp. 985–991. DOI: 10.1093/biomet/asw042 (cit. on pp. 20, 21).

Bitto, A. and S. Frühwirth-Schnatter (2018). "Achieving Shrinkage in a Time-Varying Parameter Model Framework". Journal of Econometrics. DOI: 10.1016/j.jeconom.2018.11.006 (cit. on p. 27).

Carriero, A., T. Clark, and M. Marcellino (2015). Large Vector Autoregressions with Asymmetric Priors and Time Varying Volatilities. Working Papers. Queen Mary University of London, School of Economics and Finance. URL: http://EconPapers.repec.org/RePEc:qmw:qmwecw:wp759 (cit. on pp. 18, 19).

Carter, C. K. and R. Kohn (1994). "On Gibbs sampling for state space models". Biometrika 81, pp. 541–553 (cit. on p. 28).

Carvalho, C. M. et al. (2008). "High-dimensional sparse factor modeling: Applications in gene expression genomics". Journal of the American Statistical Association 103, pp. 1438–1456. DOI: 10.1198/016214508000000869 (cit. on p. 5).

Conti, G., S. Frühwirth-Schnatter, J. J. Heckman, and R. Piatek (2014). "Bayesian exploratory factor analysis". Journal of Econometrics 183, pp. 31–57. DOI: doi:10.1016/j.jeconom.2014.06.008 (cit. on p. 5).

References II

Diebold, F. X. and C. Li (2006). “Forecasting the term structure of government bond yields”. Journal of Econometrics 130, pp. 337–364 (cit. on pp. 37, 38).

Doan, T. R., B. R. Litterman, and C. A. Sims (1984). “Forecasting and Conditional Projection Using Realistic Prior Distributions”. Econometric Reviews 3, pp. 1–100 (cit. on pp. 37, 38).

Eisenstat, E., J. C. Chan, and R. W. Strachan (2016). “Stochastic model specification search for time-varying parameter VARs”. Econometric Reviews, pp. 1–28 (cit. on p. 27).

Frühwirth-Schnatter, S. and H. F. Lopes (2018). Sparse Bayesian Factor Analysis when the Number of Factors is Unknown. arXiv pre-print 1804.04231. URL: https://arxiv.org/abs/1804.04231 (cit. on p. 5).

Frühwirth-Schnatter, S. and H. Wagner (2010). “Stochastic model specification search for Gaussian and partial non-Gaussian state space models”. Journal of Econometrics 154, pp. 85–100 (cit. on p. 27).

George, E. I., D. Sun, and S. Ni (2008). “Bayesian stochastic search for VAR model restrictions”. Journal of Econometrics 142, pp. 553–580 (cit. on pp. 37, 38).

Gerlach, R., C. Carter, and R. Kohn (2000). “Efficient Bayesian inference for dynamic mixture models”. Journal of the American Statistical Association 95, pp. 819–828 (cit. on pp. 27, 28, 31).

Giordani, P. and R. Kohn (2008). “Efficient Bayesian inference for multiple change-point and mixture innovation models”. Journal of Business & Economic Statistics 26, pp. 66–77 (cit. on pp. 27, 28).

Griffin, J. E. and P. J. Brown (2010). “Inference with Normal-Gamma Prior Distributions in Regression Problems”. Bayesian Analysis 5, pp. 171–188. DOI: 10.1214/10-BA507 (cit. on p. 6).

Gürkaynak, R. S., B. Sack, and J. H. Wright (2007). “The US Treasury yield curve: 1961 to the present”. Journal of Monetary Economics 54, pp. 2291–2304 (cit. on pp. 37, 38).

Hörmann, W. and J. Leydold (2013). “Generating Generalized Inverse Gaussian Random Variates”. Statistics and Computing 24, pp. 1–11. DOI: 10.1007/s11222-013-9387-3 (cit. on p. 17).

References III

Huber, F. and M. Feldkircher (2018). "Adaptive shrinkage in Bayesian vector autoregressive models". Journal of Business and Economic Statistics forthcoming (cit. on pp. 37, 38).

Huber, F., G. Kastner, and M. Feldkircher (forthcoming). "Should I Stay or Should I Go? A Latent Threshold Approach to Large-Scale Mixture Innovation Models". Journal of Applied Econometrics. URL: https://arxiv.org/abs/1607.04532 (cit. on p. 30).

Kalli, M. and J. E. Griffin (2014). "Time-varying sparsity in dynamic regression models". Journal of Econometrics 178, pp. 779–793 (cit. on p. 27).

Kastner, G. (2017). factorstochvol: Bayesian Estimation of (Sparse) Latent Factor Stochastic Volatility Models. R package version 0.8.4. URL: https://CRAN.R-project.org/package=factorstochvol.

Kastner, G. (2018). "Sparse Bayesian Time-Varying Covariance Estimation in Many Dimensions". Journal of Econometrics. DOI: 10.1016/j.jeconom.2018.11.007 (cit. on p. 11).

Kastner, G., S. Frühwirth-Schnatter, and H. F. Lopes (2017). "Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models". Journal of Computational and Graphical Statistics 26, pp. 905–917. DOI: 10.1080/10618600.2017.1322091 (cit. on pp. 3, 4, 17).

Kaufmann, S. and C. Schumacher (2018). "Bayesian Estimation of Sparse Dynamic Factor Models With Order-Independent and Ex-Post Mode Identification". Journal of Econometrics. DOI: 10.1016/j.jeconom.2018.11.008 (cit. on p. 5).

Koop, G. and D. Korobilis (2012). "Forecasting inflation using dynamic model averaging". International Economic Review 53, pp. 867–886 (cit. on p. 27).

Koop, G. and D. Korobilis (2013). "Large time-varying parameter VARs". Journal of Econometrics 177, pp. 185–198 (cit. on p. 27).

References IV

Koop, G. and S. M. Potter (2007). “Estimation and forecasting in models with multiple breaks”. The Review of Economic Studies 74, pp. 763–789 (cit. on p. 28).

Koop, G., R. Leon-Gonzalez, and R. W. Strachan (2009). “On the evolution of the monetary policy transmission mechanism”. Journal of Economic Dynamics and Control 33, pp. 997–1017 (cit. on p. 27).

McCracken, M. W. and S. Ng (2016). “FRED-MD: A monthly database for macroeconomic research”. Journal of Business & Economic Statistics 34, pp. 574–589 (cit. on p. 22).

McCulloch, R. E. and R. S. Tsay (1993). “Bayesian inference and prediction for mean and variance shifts in autoregressive time series”. Journal of the American Statistical Association 88, pp. 968–978 (cit. on pp. 27, 28).

Nakajima, J. and M. West (2013a). “Dynamic Factor Volatility Modeling: A Bayesian Latent Threshold Approach”. Journal of Financial Econometrics 11, pp. 116–153. DOI: 10.1093/jjfinec/nbs013 (cit. on p. 5).

Nakajima, J. and M. West (2013b). “Bayesian analysis of latent threshold dynamic models”. Journal of Business & Economic Statistics 31, pp. 151–164 (cit. on p. 28).

Neelon, B. and D. B. Dunson (2004). “Bayesian Isotonic Regression and Trend Analysis”. Biometrics 60, pp. 398–406. DOI: 10.1111/j.0006-341X.2004.00184.x (cit. on p. 28).

Nelson, C. R. and A. F. Siegel (1987). “Parsimonious modeling of yield curves”. Journal of Business 60, pp. 473–489 (cit. on pp. 37, 38).

Nishimura, A. and M. A. Suchard (2018). “Prior-preconditioned conjugate gradient for accelerated Gibbs sampling in “large n & large p” sparse Bayesian logistic regression models”. arXiv preprint arXiv:1810.12437 (cit. on pp. 20, 21).

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). “Stochastic volatility with leverage: Fast and efficient likelihood inference”. Journal of Econometrics 140, pp. 425–449. DOI: 10.1016/j.jeconom.2006.07.008 (cit. on p. 17).

References V

Park, T. and G. Casella (2008). "The Bayesian Lasso". Journal of the American Statistical Association 103, pp. 681–686. DOI: 10.1198/016214508000000337 (cit. on p. 6).

Pitt, M. and N. Shephard (1999). "Time varying covariances: a factor stochastic volatility approach". Bayesian statistics 6, pp. 547–570 (cit. on pp. 3, 4).

Primiceri, G. E. (2005). "Time varying structural vector autoregressions and monetary policy". The Review of Economic Studies 72, pp. 821–852 (cit. on pp. 37, 38).

Ritter, C. and M. A. Tanner (1992). "Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler". Journal of the American Statistical Association 87, pp. 861–868 (cit. on p. 30).

West, M. (2003). "Bayesian factor regression models in the "large p, small n" paradigm". In: Bayesian Statistics 7 – Proceedings of the Seventh Valencia International Meeting. Ed. by J. M. Bernardo et al. Oxford: Oxford University Press, pp. 307–326 (cit. on p. 5).

Yu, Y. and X.-L. Meng (2011). "To Center or Not to Center: That Is Not the Question—An Ancillarity-Suffiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency". Journal of Computational and Graphical Statistics 20, pp. 531–570. DOI: 10.1198/jcgs.2011.203main (cit. on p. 17).

Zhang, L., A. Datta, and S. Banerjee (2018). "Practical Bayesian Modeling and Inference for Massive Spatial Datasets On Modest Computing Environments". arXiv preprint arXiv:1802.00495 (cit. on pp. 20, 21).

Zhou, X., J. Nakajima, and M. West (2014). "Bayesian Forecasting and Portfolio Decisions Using Dynamic Dependent Sparse Factor Models". International Journal of Forecasting 30, pp. 963–980. DOI: 10.1016/j.ijforecast.2014.03.017 (cit. on p. 5).