Streamlining Time-varying VAR with a Factor Structure in the Parameters

Simon Beyeler

Working Paper 19.03

Streamlining Time-varying VAR with a Factor Structure in the Parameters

Simon Beyeler*†

October 2018

Abstract

I introduce a factor structure on the parameters of a Bayesian TVP-VAR to reduce the dimension of the model's state space. To further limit the scope of over-fitting the estimation of the factor loadings uses a new generation of shrinkage priors. A Monte Carlo study illustrates the ability of the proposed sampler to well distinguish between time-varying and constant parameters. In an application with Swiss data the model proves useful to capture changes in the economy's dynamics due to the lower bound on nominal interest rates.

JEL classification: C11, C22, E52
Key words: Bayesian VAR; time-varying parameter; dimension reduction

*Study Center Gerzensee and University of Bern, Dorfstrasse 2, CH-3115 Gerzensee, simon.beyeler@szgerzensee.ch

†I thank Luca Benati, Sylvia Kaufmann, Rodney Strachan and Mark Watson for valuable comments and discussions. Omissions and errors are mine.

1 Introduction

Capturing instabilities in the economic environment or in the behavior of economic agents is a challenging problem in modern macroeconometric analysis. Periods such as the great moderation during which the volatility of major economic variables strongly declined and other occurrences such as the recent shift form conventional to unconventional monetary policy measures by central banks who were constrained by the lower bound on nominal interest rates, call into question the assumption of fixed parameters present in most traditional econometric models. Instead, the relationships that describe an economy might well be evolving over time as the behavior of economic agents is likely to adapt to changes of the "rules of the game", such as policy changes or modifications of the institutional settings. This point was famously emphasized in the Lucas critique (Lucas, 1976). Such variation in the structural relationships naturally lead to shifts in the reduced form parameters of macroeconometric models, such as vector auto regressions (VAR). Indeed, using a battery of tests Stock and Watson (1996) found evidence of instability for a substantial number of US macroeconomic time series, supporting these doubts.
A widely used method to approach these issues are time-varying parameter (TVP) models in combination with stochastic volatility such as the TVP-VAR proposed by Cogley and Sargent (2005) or Primiceri (2005).[1] These models offer the flexibility to handle transitory as well as permanent changes in the coefficients and/or in the covariance structure of the error terms. However, this great flexibility goes hand in hand with the risk of over-fitting as (i) the number of parameters to be estimated grows exponentially and (ii) not necessarily all the relationships among the variables are truly time-varying. It is of course important to capture possible breaks in the model's parameters but at the same time it is inefficient to assume time variation where there is none. Different approaches to overcome these problems have been proposed in the recent literature. Belmonte et al. (2014) and Bitto and Frühwirth-Schnatter (2016) among others apply Bayesian shrinkage priors to the innovation variances in the state equation of the parameters. This allows them to discriminate between time-varying and constant parameters by eventually shrinking the state variance of a constant parameter towards zero. Another important issue with TVP-VAR is the size of the model's underlying state space representation. A common assumption in TVP models is that each coefficient enters the model as a state variable typically evolving as a random walk. This allows for a high degree

[1] See e.g. Nakajima (2011) for an overview of the Methodology.

of flexibility in the parameter's time-varying behavior but leads to a large number of states that have to be filtered from the data. The question arises whether this great extent of flexibility is indeed necessary, as already Cogley and Sargent (2005) discovered that the estimated time-varying coefficients showed strong similarities in their behavior. A new strand of the literature takes advantage of this feature, either by directly introducing a factor structure in the TVP-VAR model to reduce the number of state variables such as Stevanovic (2016), or by working with a rank reduced state covariance matrix as de Wind and Gambetti (2014) or Eisenstat et al. (2018). The latter in turn also imposes a factor-like structure upon the model's coefficients. Both approaches rest on the assumption that the parameters share a high degree of commonality in their time-varying behavior and that the number of sources for parameter instability is in fact (much) smaller than the actual number of parameters of the model.

The aim of this paper is to combine these two recent developments and adopt the factor-TVP structure for the VAR parameters together with shrinkage priors in the estimation of the factor loadings. The introduction of the factor structure leads to a tremendous reduction of the state space as the time variation in the model parameters is fully captured by a small set of factors. The application of shrinkage priors is based on the notion that only a subset of the unobserved parameters is truly time-varying while the remainder is in fact constant, implying that a substantial fraction of entries in the factor loading matrix is essentially equal to zero. The estimation of the factor loadings is based on a novel generation of Normal-Gamma priors as applied in Bitto and Frühwirth-Schnatter (2016) in the context of TVP models or in Huber and Feldkircher (2017) for VARs with constant parameters. Their advantage is that they provide a high degree of flexibility which ensures that non-zero entries are not too strongly shrunk towards zero, allowing to properly discriminate between time-varying and constant but at the same time also between significant and insignificant parameters. In sum, this delivers a model that is able to capture structural breaks (or trends) by allowing for time variation wherever necessary but in a well structured way. This reduces model complexity and together with the application of shrinkage offers a remedy for the over-fitting problem typically occurring in these types of models. It further allows working with larger sets of endogenous variables while TVP-VARs are usually applied to small systems. In a Monte Carlo study with simulated data the proposed sampler shows its ability to correctly estimate the degree of time variation and to correctly distinguish between truly constant and truly time-varying parameters.

An additional problem that arises in the context of TVP models is how to test for stability of the underlying dynamic system. As the eigenvalue analysis of the "time-frozen" coefficient matrices borrowed from VAR with constant parameters is non-informative with regard to the dynamic properties of the system (an issue that is typically ignored in the existing literature) a correct alternative needs to be found. Neusser (2018) addresses this question and advocates the analysis of the Lyapunov exponents as a valid solution.

In an application for Switzerland the model is estimated with monthly data to study the effect of an appreciation shock on the economy and its evolution during the lower bound period for the short-term nominal interest rate. Although the estimated time variation in the coefficients is muted, the model captures a stark drop in the effect of the exchange rate shock on the nominal interest rate once it has been lowered to basically zero. The response of the remaining variables also changes but to a much lesser extent. A forecasting exercise based on the same dataset for Switzerland further reveals a superior forecasting performance of the factor-TVP model (henceforth FacTVP) compared to the traditional TVP-VAR. A second application is based on historical inflation data for the United Kingdom, Norway, Sweden and the United States covering almost 200 years of history. It clearly emphasizes the role of time variation among model parameters when working with time series covering a longer time span. The remainder of the paper proceeds as follows. Section 2 contains the specification of the model and the prior distributions. Section 3 presents the Bayesian MCMC sampling scheme. Section 4 summarizes the results of a Monte Carlo study with simulated data. Section 5 presents an application with empirical data for the Swiss economy while Section 6 presents results for an application with long run inflation series, and Section 7 concludes.

2 Model

2.1 Model specification

Consider the following TVP-VAR with stochastic volatility (SV) in the spirit of Cogley and Sargent (2005) or Primiceri (2005),

$$y_t = b_{0,t} + \sum_{j=1}^{p} B_{j,t} y_{t-j} + e_t, \tag{1}$$

where y_t is a N-dimensional vector of non-trending variables observed at time t, with time-varying intercepts $b_{0,t}$, and coefficient matrices $B_{j,t}$. Further the $N \times 1$ vector of one step ahead prediction errors e_t is assumed Gaussian with zero mean and time-varying covariance matrix Σ_t, i.e.

$$e_t \sim N(0, \Sigma_t). \tag{2}$$

Defining the regressor matrix as $X_t \equiv \left[I_N \otimes (1, y'_{t-1}, ..., y'_{t-p})\right]$ and collecting the time-varying coefficients in the vector b_t yields a more parsimonious formulation of equation (1):

$$y_t = X_t b_t + e_t, \tag{3}$$

this helps simplifying the notation throughout the rest of the paper. The length of the vector b_t is equal to the number of coefficients of the VAR and depends on the number of included variables N and the lag order p , i.e. $C \equiv pN^2 + N$ (if a constant is included in the model). In contrast to the traditional formulation of TVP-VAR models, where each non-constant parameter enters the model as a state variable and evolves as a random walk, the propagation of the autoregressive coefficients (the elements of b_t) in the present model is different. Each coefficient builds the sum of two components: a constant component λ^c that is independent of time, and a time-varying component $\lambda^f f_t$, where f_t describes a k-dimensional vector of latent factors that govern the time variation of the coefficients. The number of factors is assumed to be substantially lower than the total number of autoregressive coefficients such that $k << C$, further we assume that the time-varying behavior of all coefficient is fully described by these k factors. The factor themselves are assumed to evolve as independent random walks, this is similar to the traditional formulation of TVP-VARs where the evolution of the state variables is governed by a random walk structure to allow for transitory as well as permanent shifts in the VAR parameters. The propagation of the model's autoregressive coefficients in b_t is therefore governed by

$$b_t = \lambda^c + \lambda^f f_t \tag{4}$$

$$f_t = f_{t-1} + \eta_t \tag{5}$$

Such a formulation implies that the possible sources of parameter instability are reduced from C to k. This structure is based on the idea that the time variation in the different VAR coefficients shares a high degree of commonality and therefore,

can be well described by a small set of common factors. Empirical evidence of such a common behavior of the model's parameters is as old as the TVP-VAR models. Cogley and Sargent (2005) were among the first to discover a high degree of commonality in the behavior of the coefficients. Later on Canova and Ciccarelli (2009) introduced reduced-rank time variation in the context of multi-country VARs and more recently Stevanovic (2016) developed a factor-TVP-VAR where all the model's coefficients are time-varying possess a factor representation. He provides empirical evidence for the model's ability to capture time variation in macroeconomic data. The present work extends the FacTVP setup by introducing recently developed Bayesian shrinkage priors for the factor loadings that allow to better discriminate between constant and time-varying coefficients. This helps overcoming the problem of over-fitting in the presence of constant true parameters. The approach is similar to Belmonte et al. (2014) and Bitto and Frühwirth-Schnatter (2016) who work with shrinkage priors for the innovation variances to limit the degree of time variation in the estimated coefficients. Kastner (2016) uses a similar setup in a factor stochastic volatility model.

In their work Eisenstat et al. (2018) show that working with a reduced-rank covariance matrix for the innovations of the state equation leads to a law of motion of the exact same form, as the reduction of the rank imposes a factor-like structure upon the time-varying parameters in b_t. Note that there is no time-varying idiosyncratic component in b_t, as adding this would introduce an additional random effect at each point in time which would no longer allow us to discriminate between time-varying and constant parameters and secondly, it would complicate the estimation of the factors by introducing additional identification issues. Modeling the time-varying coefficients based on a factor structure reduces the dimension of the state space system tremendously, this would allow to estimate the model with a larger number of endogenous variables.

Combining equations (3) and (4) leads to an extended version of the observation equation that relates the unobserved factors directly to the observable variables in y_t,

$$y_t = X_t\lambda^c + X_t\lambda^f f_t + e_t. \tag{6}$$

Together with the law of motion of the factors and conditional on the factor loadings as well as the time-varying covariance matrix of the VAR innovations this forms a linear Gaussian state space system. This can be used to recover the factors from the

data using the algorithm of Carter and Kohn (1994) or the more efficient precision sampler proposed in Chan and Jeliazkov (2009).
Following Primiceri (2005), the time-varying error covariance matrix is decomposed into two parts

$$\Sigma_t = A_t^{-1} H_t A_t^{-1'} \tag{7}$$

where H_t is a diagonal matrix and A_t is a lower diagonal matrix with ones on the diagonal. H_t captures the volatility states, while A_t contains the covariance states (or rather the inverse). This implies that the reduced form innovations e_t are linear combinations of N independent standard normal increments ε_t,

$$e_t = A_t^{-1} H_t^{\frac{1}{2}} \varepsilon_t. \tag{8}$$

Note that the innovations ε_t often are assigned a structural interpretation (based on the lower triangular structure of A_t this yields the well known Cholesky type timing restrictions), however, the decomposition holds in general even if they cannot be assigned any structural interpretation. Let h_t be a vector containing the natural logarithm of the volatilities, i.e. the diagonal elements of H_t, and collect the non-fixed elements of A_t in a vector a_t. To reduce the dimension of the state space even further, the covariance states possess an identical factor structure as the autoregressive coefficients. Instead of $N(N-1)/2$ covariance states there are only k_a covariance factors to be filtered from the data. This leads to a similar state space representation as above:

$$a_t = \lambda_a^c + \lambda_a^f f_t^a \tag{9}$$

$$f_t^a = f_{t-1}^a + \eta_t^a \tag{10}$$

Finally, the log volatilities h_t are assumed to evolve as random walk, which is analogue to the specification used in Primiceri (2005).

$$h_t = h_{t-1} + \nu_t \tag{11}$$

An interesting further step would be to reduce the dimension of the volatilities as suggested in Carriero et al. (2016). The gains of the reduction is smaller here because the number of volatilities in the model grows linearly with the number of included variables. But once the model is applied to larger data sets this might become crucial

and is an interesting path for future work. The innovations of the state equations are assumed to be Gaussian with zero mean and block diagonal covariance matrix

$$V = Var \begin{bmatrix} \varepsilon_t \\ \eta_t \\ \eta_t^{\alpha} \\ \nu_t \end{bmatrix} = \begin{bmatrix} I_N & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix} \tag{12}$$

Factor normalization is achieved by setting $Q = \kappa_Q \cdot I_k$ and $S = \kappa_S \cdot I_{k_a}$, due to the random walk structure κ_Q and κ_S are set to small numbers. This does however not limit the extent of time variation but it ensures that the scale of the factor loadings is not getting too small, and facilitates estimation.

2.2 Separating the intercepts

In order to decouple the time-varying intercepts from the autoregressive VAR coefficients the model is slightly adapted. In this alternative specification an additional set of factors f_t^0 governing only the time variation of the intercepts is introduced.

$$b_{0,t} = \lambda_0^c + \lambda_0^f f_t^0 \tag{13}$$

$$b_{-0,t} = \lambda^c + \lambda^f f_t \tag{14}$$

This representation allows to estimate a model where only the intercepts or only the autoregressive coefficients can move over time. This feature is helpful in a situation where the data appear to have breaks in the mean but where their dynamic behavior such as the persistence seems to be the same over the observed sample period, or in cases where one wants to allow the time-varying intercepts to enter the model as separate state variables (by setting $k_0 = N$ and λ_0^f to the identity matrix).
The factors related to the intercepts can then simply be added to the states for the autoregressive coefficients as they both evolve as independent random walks

$$\begin{bmatrix} f_t^0 \\ f_t \end{bmatrix} = \begin{bmatrix} f_{t-1}^0 \\ f_{t-1} \end{bmatrix} + \begin{bmatrix} \eta_t^0 \\ \eta_t \end{bmatrix} \tag{15}$$

where analogously to the original specification above the innovations are assumed to be orthogonal across factors.

2.3 Stability in TVP models

The notion of stability in time-varying parameter models is a delicate point that requires careful thought. Early advances in the literature addressing this issue are Brandt (1986) and Bougerol and Picard (1992). The typical way to deal with questions concerning the stability of a model found in the literature is to work with "time frozen" system matrices and to mimic what one would do in the case with constant parameters i.e. checking that the eigenvalues of the relevant system matrix lie within the unit circle (eigenvalue-condition). However, this is not the right way to proceed, as in an environment with time-varying parameters the eigenvalues are non-informative about the stability of the model. The problem is that even a system fulfilling the eigenvalue-condition at each point in time may still show an explosive behavior in the long run (see e.g. Neusser (2018), Elaydi (2013) or Francq and Zakoïan (2001) for examples). At the same time, a system that violates the eigenvalue-condition over certain periods of time does not necessarily possess explosive properties in the longer run. In the context of TVP models, stability is not a property that should be studied at each point in time separately by inspecting the properties of the "time frozen" coefficients. Rather, it should be evaluated as a global property of the whole system including the coefficient matrices $B_{1,1}$ up to $B_{p,T}$. A correct alternative to the eigenvalue-condition in the TVP framework are the system's Lyapunov exponents as discussed in Francq and Zakoïan (2001) or Neusser (2018). Rewriting the VAR stated in equation (1) in its companion form we obtain

$$Y_t = \mathcal{B}_t Y_{t-1} + E_t$$

where $Y_t = \left[y_t', y_{t-1}', \ldots, y_{t-p+1}'\right]'$ and $E_t = [e_t', 0, \ldots, 0]'$. The top Lyapunov exponent of the system is then defined as

$$\varphi = \lim_{T\to\infty} \inf \frac{1}{T} \log || \prod_{t=1}^{T} \mathcal{B}_t || \tag{16}$$

where $||\cdot||$ is a suitable norm. Stability is a function of the product of the system matrices $\mathcal{B}_t$ over all periods of the sample. The system is said to be stable whenever φ is negative. In principle this enlarges the scope of interesting models, because it allows for "time frozen" coefficient matrices with explosive behavior in certain periods, as long as the system as a whole remains stable.

2.4 Prior specification

Given the different structure of the model compared to traditional TVP-VAR models also the prior distributions applied here are different. For the estimation of the factor loadings I rely on two alternative versions of shrinkage priors. In the following, the prior distributions used are explained in more detail.

2.4.1 Priors for the factor loadings

Conditional on the remaining parameters of the model λ^c and λ^f can be recovered from a linear regression model. Defining $\lambda \equiv \begin{bmatrix} \lambda^c & \lambda^f \end{bmatrix}$ and slightly rearranging equation (6) leads to

$$y_t = \underbrace{\left(\begin{bmatrix} 1 \\ f_t \end{bmatrix}' \otimes X_t \right)}_{\equiv Z_t} vec\left(\lambda'\right) + e_t$$

Since the dimension of λ can become quite large as the number of variables N, the number of lags p or the number of factors k grows, one can face over-fitting issues when estimating the factor loadings. To overcome these problems and to support the view that potentially some elements in b_t do not vary over time Bayesian shrinkage priors that were recently proposed in the literature are introduced for the estimation of λ. One way to introduce shrinkage in this model is to follow the line of Huber and Feldkircher (2017) who extend the Normal-Gamma prior proposed by Griffin and Brown (2010) and apply it to a constant parameter VAR with stochastic volatility. The Normal-Gamma prior is closely related to the Bayesian LASSO by Park and Casella (2008) but allows for additional flexibility which leads to richer shrinkage properties. For the factor loadings the Normal Gamma prior (henceforth NG) writes

$$\pi\left(\lambda_{rj}^{\{s\}} | \tau_{rj}^{\{s\}}\right) \sim N\left(0, \tau_{rj}^{\{s\}}\right), \qquad \pi\left(\tau_{rj}^{\{s\}} | \rho_j^2\right) \sim \mathcal{G}(\nu_j, \nu_j \rho_j^2/2)$$
$$\pi(\rho_j^2) \sim \mathcal{G}(a_{j1}, a_{j2})$$
$$\pi(\nu_j) \sim Exp(1)$$

where the index s denotes the form of shrinkage. Similar to the model specifications in Huber and Feldkircher (2017) I introduce either global, (VAR-)equation-specific, factor-specific, or lag-specific shrinkage. To facilitate the notation $\lambda^{\{s\}}$ denotes a reshuffled version of the original λ that collects all the elements of λ belonging to

the same level of common shrinkage in the same column. The number of rows (r) and columns (j) of $\lambda^{\{s\}}$ depends on the form of shrinkage. $\tau^{\{s\}}$ contains the prior variances of the factor loadings and has exactly the same shape as $\lambda^{\{s\}}$.

1. **Global shrinkage:** Under global shrinkage the same shrinkage parameters are applied to all factor loadings. In the notation introduced above this means that $\lambda^{\{s\}}$ is equal to $vec(\lambda')$, i.e. a vector of length equal to the total number of factor loadings.

2. **Equation-specific shrinkage:** In contrast to global shrinkage, different shrinkage parameters ρ_j^2 and ν_j and possibly also different hyperparameters a_1^j, a_2^j are specified for each equation ($j = 1, \ldots, N$) of the VAR. $\lambda^{\{s\}}$ is a matrix with N columns and as many rows as there are factor loadings per equation ($(k+1)pN$).

3. **Factor-specific shrinkage:** The same shrinkage parameters are applied to all the factor loadings related to the same factor. In this case $\lambda^{\{s\}} = \lambda$ and no reshuffling takes place.

4. **Lag-specific shrinkage:** All factor loadings related to the same lag of the VAR are treated in the same way and $\lambda^{\{s\}}$ is a matrix the columns of which contain all the loadings related to the same lag, whereby the ordering of the coefficients in column j corresponds to vec(B_j'). The number of rows of $\lambda^{\{s\}}$ is equal to $p(k+1)N^2$.

An alternative way to prohibit over-parametrization of the model is to perform stochastic search variable selection (SSVS) as in George et al. (2008). This can be achieved by placing a hierarchical mixture prior on the elements of λ that assigns some non-zero probability to all the possible submodels of the original model. Updating the prior with the likelihood of the data will then increase the weights on those models that are favored by the data. Given the dimensions of the problem at hand it is in general infeasible to compute the posterior probability of every single submodel, but it is possible to simulate a Markov chain sample in which the submodels that are highly supported by the data are visited more often. The hierarchical prior takes on the form

$$\pi(\lambda_{rj}|\gamma) = (1 - \gamma_{rj})N(0, \tau_0) + \gamma_{rj}N(0, \tau_1) \tag{17}$$

$$\pi(\gamma_{rj}) = Bernoulli(p_{rj}) \tag{18}$$

The prior variances τ_0 and τ_1 are set such that $\tau_0 << \tau_1$, this implies that when $\gamma_{rj} = 0$ λ_{rj} is restricted to be close to zero and left unrestricted for $\gamma_{rj} = 1$. Here p_{rj} denotes the a priori probability that λ_{rj} is unrestricted.

2.4.2 Factor loadings for the covariance states

The factors loadings for the covariance states, λ_a^c and λ_a^f can essentially be treated in the same way as those associated with b_t. The only thing that needs to be adjusted, is the regression model used in the estimation. Let $\lambda_a \equiv \begin{bmatrix}\lambda_a^c & \lambda_a^f\end{bmatrix}$ and $\hat{y}_t \equiv y_t - X_t b_t$, then conditional on b_t and h_t the covariance states can be recovered from

$$\begin{aligned} A_t \hat{y}_t &= e_t \\ \hat{y}_t &= -\left(A_t - I_N\right)\hat{y}_t + e_t, \end{aligned}$$

Given the factor structure of A_t and applying the same transformation as above we obtain

$$\hat{y}_t = \underbrace{\left(\begin{bmatrix} 1 \\ f_t^a \end{bmatrix}' \otimes \tilde{X}_t^a\right)}_{\equiv Z_t^a} vec\left(\lambda_a'\right) + e_t$$

where $\tilde{X}_t^a$ is formed of the columns of $X_t^a = -\left(I_{N-1} \otimes \hat{y}_{t;2:N}'\right)$ related to the non-constant elements of A_t (i.e. the ones on the diagonal and the zeros above the diagonal). The priors that are used in the estimation are of the same form as described above. For the NG prior the shrinkage level can be chosen to be either global, equation- or factor-specific.

2.4.3 Innovation variances of volatility states

The specification of the prior variances for the innovations of the volatility states, $W = diag(w_1^2, \ldots w_N^2)$, is given by

$$w_i^2 \sim IG\left(df_w, V_w\right) \tag{19}$$

where IG stands for the inverse Gamma distribution, with degree of freedom parameter df_w and scale parameter V_w. The scale parameter V_w is chosen such that the prior mean of w_i^2 is equal to 0.01. This is a value that is often found in the

related literature and it insures that the changes in the log-volatilities stay in a reasonable range. An alternative approach where W is assumed to be a full matrix and the scaling parameter for the innovation variances is estimated as suggested in Amir-Ahmadi et al. (2016) is described in Section B of the Appendix.

3 Bayesian Inference

To describe the estimation of the model, I introduce some additional notation. While y_t denotes observations in period t, y^t indicates observations up to period t, this applies to all other variables. All the parameters as well as hyperparameters are included in $\Theta = \{\lambda, \lambda_a, Q, S, W, \vartheta\}$, where all the hyperparameters depending on the chosen model specification enter in ϑ.

3.1 Posterior sampler

To obtain a sample from the joint posterior distribution

$$\pi(f^T, f_a^T, h^T, \Theta|y^T) \propto L(Y^T|f^T, f_a^T, h^T, \Theta)\pi(f^T, f_a^T, h^T|\Theta)\pi(\Theta)$$

the following algorithm is applied:

(i) Draw the volatilities:

   (a) Draw the volatility states h^T from $\pi(h^T|y^T, f^T, f_a^T, \Theta)$. This step uses the approach of Kim et al. (1998) in combination with the precision sampler of Chan and Jeliazkov (2009) to efficiently generate draws from a nonlinear state space system.

   (b) Draw the covariance matrix of the volatility state innovations W.

(ii) Draw the VAR coefficients:

   (a) Draw f^T from $\pi(f^T|y^T, f_a^T, h^T, \Theta)$ using the precision sampler.

   (b) Draw λ^c, λ^f from $\pi\left(\lambda|y^T, h^T, f^T, f_a^T, \Theta_{(-\lambda)}\right)$ and update the hyper parameters.

   (c) Compute the Lyapunov exponent and verify the condition $\varphi < 0$. Otherwise discard the draw.

(iii) Draw a^T:

(a) Draw f_a^T from $\pi(f_a^T|y^T, f^T, h^T, \Theta)$ using the precision sampler.

(b) Draw λ_a^c, λ_a^f from $\pi\left(\lambda_a|y^T, f^T, f_a^T, h^T, \Theta_{(-\lambda_a)}\right)$ and update the hyper parameters.

Additional details for the single steps are presented in more detail below.

3.2 Sampling the time-varying factors

To draw from the conditional posterior of the factors

$$\pi(f^T|y^T, f_a^T, h^T, \Theta)$$

we first subtract the constant component from the observed variables which yields a re-centered version of the data denoted by

$$y_t^f \equiv y_t - X_t\lambda^c.$$

One can then apply the precision sampler of Chan and Jeliazkov (2009) for linear state space systems on

$$y_t^f = X_t\lambda^f f_t + e_t \tag{20}$$

$$f_t = f_{t-1} + \eta_t \tag{21}$$

where the increments e_t and η_t are normally distributed

$$\begin{bmatrix} e_t \\ \eta_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \Sigma_t & 0 \\ 0 & Q \end{bmatrix}\right)$$

to efficiently obtain a posterior draw of f^T. The precision sampler of Chan and Jeliazkov (2009) is presented in more detail in Appendix A. The set of factors associated with the covariance states, f_t^a, can be drawn in the exact same manner from $\pi(f_a^T|y^T, H^T, B^T, \Theta)$ using the state space representation,

$$\hat{y}_t^f = X_t^a\lambda_a^f f_t^a + e_t \tag{22}$$

$$f_t^a = f_{t-1}^a + \eta_t^a, \tag{23}$$

where $\hat{y}_t^f \equiv \hat{y}_t - X_t^a\lambda_a^c$ denotes the re-centered version of $\hat{y}_t$.

3.2.1 Scaling the factors: Data augmentation

Since, conditional on the data, only the mean of the autoregressive coefficients b_t but not the mean of the factors f_t and the factor loadings λ are identified, the parameter space is expanded by the working parameter $\bar{f}$ capturing the factor mean of f_t. This data augmentation step stabilizes the sampler by ensuring that the posterior mean of the factors stays centered around zero. Due to the random walk structure of the factors we can write the factor in period t as $f_t^* = \bar{f} + \sum_{j=1}^{t} \eta_t$, where $\bar{f}$ is distributed as $N(\bar{f}, t \cdot Q)$. Assuming a Normal prior for $\bar{f}$, $\pi(\bar{f}) \sim N(0, B)$ its conditional posterior is $\pi(\bar{f}|\cdot) \sim N(\mu^{\bar{f}}, \Sigma^{\bar{f}})$ with

$$\begin{aligned}\Sigma^{\bar{f}} &= \left[[1_T \otimes I_k]'[\tau \otimes Q]^{-1}[1_T \otimes I_k] + B^{-1}\right]^{-1} \\ \mu^{\bar{f}} &= \Sigma^{\bar{f}} \left[[1_T \otimes I_k]'[\tau \otimes Q]^{-1} vec(f^{*T})\right],\end{aligned}$$

where τ is a vector of length T with the t^{th} element equal to $\tau_t = t^{-1}$. The sampling of f_t is then based on the following steps: first f_t^* is drawn as outlined above, then $\bar{f}$ is drawn from its conditional posterior and finally the factor is centered by subtracting $\bar{f}$ from f_t^*: $f_t = f_t^* - \bar{f}$. The same identification issue applies to the covariance factors, and therefore, f_t^a is demeaned in the same way.

3.3 Factor loadings

For the different prior specifications described in Section 2.4.1 the conditional posterior of the factor loadings is always Gaussian and is described by

$$\pi\left(vec\left(\lambda'\right)|\cdot\right) \sim N\left(\overline{m}, \overline{V}\right)$$

where the posterior variance $\overline{V}$ and mean $\overline{m}$ are equal to

$$\begin{aligned}\overline{V} &= \left[\sum_{t=p+1}^{T} Z_t' \Sigma_t^{-1} Z_t + \underline{\mathrm{V}}^{-1}\right]^{-1} \\ \overline{m} &= \overline{V}\left[\sum_{t=p+1}^{T} Z_t' \Sigma_t^{-1} y_t\right].\end{aligned}$$

$\underline{\mathrm{V}}$ is a diagonal matrix containing the prior variances for each element in λ.

3.3.1 Update the hyperparameters

Under the Normal-Gamma prior:

The posterior distributions for the hyperparameters of the NG prior for the different shrinkage specifications described in Section 2.4.1 can be summarized as follows. The conditional posterior distributions of $\tau_{rj}^{\{s\}}$ and ρ_j^2 follow a generalized inverse Gaussian and a Gamma distribution respectively, with

$$\pi\left(\tau_{rj}^{\{s\}}|\lambda_{rj},\rho^2,\nu_s\right) \sim GIG\left(\nu_s - \frac{1}{2}, \nu_s\rho_j^2, \lambda_{rj}^2\right)$$
$$\pi(\rho_s^2|\tau_{rj}^{\{s\}},\nu_s) \sim \mathcal{G}\left(a_1^s + \nu_s C, a_2^s + \frac{\nu_s}{2}\mathcal{S}_j\right)$$

where the same notation as introduced in Section 2.4.1 applies and $\mathcal{S}_j = \sum_{r=1}^{R^{\{s\}}} \tau_{rj}^{\{s\}}$ equals the sum over all elements in column j of τ^s, with $R^{\{s\}}$ equal to the number of rows in $\lambda^{\{s\}}$ and $\tau^{\{s\}}$.

The conditional posterior of ν_j does not inherit a well-known form, therefore sampling requires a Metropolis Hastings step. Following Huber and Feldkircher (2017) the proposal density q^ν is chosen to be log-normal with the parameters $\mu^\nu = \ln(\nu_j)$ and $\sigma^\nu = \kappa_\nu$ being, respectively, the mean and the standard deviation of the underlying normal distribution. The standard deviation is used as tuning parameter to adjust the acceptance rate during the first half of the burn-in phase. The proposal ν_j^* is accepted with probability

$$\min\left[1, \frac{q^\nu(\nu_j|\nu_j^*)\pi(\nu_j^*)(\nu_j^*\rho_j^2/2)^{\nu_j^*}\Gamma(\nu_j)}{q^\nu(\nu_j^*|\nu_j)\pi(\nu_j)(\nu_j\rho_j^2/2)^{\nu_j}\Gamma(\nu_j^*)}\mathcal{P}_j^{\nu_j^*-\nu_j}\right]$$

where $\Gamma(\cdot)$ denotes the Gamma function and $\mathcal{P}_j = \prod_{r=1}^{R^{\{s\}}} \tau_{rj}^{\{s\}}$ denotes the product over all $\tau_{rj}^{\{s\}}$ related to the same level of common shrinkage.

SSVS:

Setting $p_{rj} = p_0$ for all r, j and assuming that the elements of λ are independent a priori leads to a conditional posterior for the factor loadings that is Gaussian as described above and hence the posterior update of the loadings is exactly as described at the beginning of this section. The update of the hyperparameters of the SSVS prior is as follows. The conditional posterior of a generic γ_{rj} conditional on all the

other elements $\gamma_{(-rj)}$ is

$$\begin{aligned}\pi(\gamma_{rj}|\gamma_{(-rj)},\lambda,\cdot) &\sim Bernoulli(u^1_{rj}/(u^0_{rj}+u^1_{rj}))\\ u^0_{rj} &= p(\lambda_{rj}|\gamma_{(-rj)},\gamma_{rj}=0)(1-p_0)\\ u^1_{rj} &= p(\lambda_{rj}|\gamma_{(-rj)},\gamma_{rj}=1)p_0\end{aligned}$$

If the components of λ are assumed to be independent a priori, then

$$\begin{aligned}\hat{u}^0_{rj} &= \frac{1}{\tau_0} exp\left(-\frac{\bar{\lambda}^2_l}{2\tau_0}\right)(1-p_0)\\ \hat{u}^1_{rj} &= \frac{1}{\tau_1} exp\left(-\frac{\bar{\lambda}^2_l}{2\tau_1}\right)p_0.\end{aligned}$$

The hyperparameter controlling the prior variance is then simply updated by setting it equal to $\tau_{rj} = (1-\gamma_{rj})\tau_0 + \gamma_{rj}\tau_1$.

3.3.2 For the alternative specification with separated intercepts

Under the alternative model specification with separate dynamics for the intercepts and the autoregression coefficients, the estimation of the factor loadings proceeds in two steps. First, conditional on $b_{-0,1:T}$, $\lambda_0 \equiv \begin{bmatrix}\lambda^c_0 & \lambda^f_0\end{bmatrix}$ is estimated from

$$y^0_t = \underbrace{\begin{bmatrix}1\\ f^0_t\end{bmatrix}'}_{\equiv Z^0_t} vec\,(\lambda_0') + e_t$$

where $y^0_t \equiv y_t - \tilde{X}_t b_{-0,t}$ and $\tilde{X}_t$ does not contain the columns of X_t related to the intercepts whose elements are all equal to one. Due to the smaller number of elements in λ_0 only global shrinkage is imposed under the Normal-Gamma prior.[2] The second step consists of estimating λ conditional on $b_{0,1:T}$ from

$$\tilde{y}_t = \underbrace{\left(\begin{bmatrix}1\\ f_t\end{bmatrix}' \otimes \tilde{X}_t\right)}_{\equiv \tilde{Z}_t} vec\,(\lambda') + e_t$$

[2]There is no distinction between elements associated with a certain equation or factor, instead the same set of shrinkage parameters are applied to all the elements of λ_0.

where $\tilde{y}_t \equiv y_t - b_{0,t}$. The conditional posterior of both λ^0 and λ is Gaussian and analogue to the one described above. The hyperparameters are then updated according to the same steps as they are described in the previous section.

3.4 Controlling time variation: Data augmentation

A critical issue with TVP models is how to control the overall amount of time variation for each parameter, especially because the data might not always be very informative in that respect. In the traditional TVP-VAR literature this is achieved by selecting the scaling factors for the prior variances of the innovations accordingly (see also Section B of the Appendix). In the present setup this issue is a little bit more involved because of the factor structure for the autoregressive coefficients b_t and the covariance states a_t. A measure for the amount of time variation is given by

$$\Omega = \lambda^f Q \lambda^{f'}.$$

To stabilize the sampler at reasonable values of Ω, another data augmentation is used, and the working parameter Ω^* is introduced. At each iteration we draw a new value Ω^* from

$$\pi(\Omega|\cdot) = IW(\overline{V}_\Omega, \overline{d}_\Omega),$$

where $\overline{V}_\Omega = \underline{V}_\Omega + 0.5 \sum_{t=1}^{T} (b_t - b_{t-1})(b_t - b_{t-1})'$ and $\overline{d}_\Omega = \underline{d}_\Omega + \frac{T}{2}$ and rescale the factor loadings by

$$D = \begin{bmatrix} \left(\frac{\Omega_1^*}{\Omega_1}\right)^{\frac{1}{2}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \left(\frac{\Omega_C^*}{\Omega_C}\right)^{\frac{1}{2}} \end{bmatrix}$$

An appropriate choice of $\underline{V}_\Omega$ and $\underline{d}_\Omega$ helps to control the overall amount of time variation for the autoregressive coefficients. The same procedure is applied to the covariance states a_t.

4 Monte Carlo study

In order to illustrate the ability of the proposed sampler to correctly distinguish between constant and time-varying parameters, this section contains a Monte Carlo exercise based on a set of 100 simulated histories. A model with two factors governing the time variation of the autoregressive coefficients and one additional factor for the

covariance states serves as data generating process. The number of variables is set to $N = 5$. To introduce a fraction of completely insignificant coefficients 20 percent of the elements in b_t are set to zero over the whole sample period, this implies that the corresponding row of the factor loading matrix contains only zeros. The probability for a non-zero loading of the remaining coefficients is set to 0.6 for the first factor and 0.4 for the second factor. The probability for a non-zero loading on the covariance factor is set to 0.6 for each element in a_t. For every simulation the position of the zero and non-zero factor loadings is randomly assigned. To ensure stationarity and rule out explosive behavior, only paths for b_t for which the corresponding Lyapunov exponent is negative are considered. All the simulations with a positive Lyapunov exponent are discarded and replaced with a stable simulation. The FacTVP model is then estimated twice for each simulated data set, once with the NG prior and once with the SSVS prior. The hyperparameters for the Normal-Gamma prior are set to $a_1 = 0.01$ and $a_2 = 0.01$, while the respective ones for the SSVS prior are set to $\tau_0 = 0.01$ and $\tau_1 = 0.5$. The Gibbs sampler is run for 6,000 iterations after a burn in phase of 4,000 iterations, and every third draw is stored.[3] The degree of shrinkage is stronger under the Normal-Gamma prior. Table 1 contains the relative root mean squared errors (RMSE) of the NG relative to the SSVS prior averaged over all parameters, and all simulations (a value below unity implies that NG outperforms SVSS). It turns out that the NG prior is slightly superior in case of estimating the autoregressive coefficients while the estimates for the covariance states based on the SVSS prior show a higher accuracy on average. In the case of the volatilities the two models perform almost equally well, which is not surprising because no shrinkage is involved here.

Table 1: Relative RMSE (average over all coefficients and time periods) of the estimation with the NG prior in comparison to the estimation with the SSVS prior along with the mean RMSE for the SSVS prior.

	b_t	a_t	h_t
Mean	0.9244	1.0576	1.0047
Min	0.8119	0.8641	0.9131
Max	1.0632	1.4152	1.1020
RMSE	0.0670	0.1154	0.3287

Figure 1 computes the difference between the estimated and the true time variation

[3] I have been working with longer chains, but due to fast convergence of the model I successively lowered the number of draws to economize on computational time.

(measured as the standard deviation of each coefficient over the entire sample period) for b_t and a_t for all the simulations based on the NG prior. The coefficients are ordered according to their true degree of time variation, starting with those that exhibit a strong variation on the left hand side of the figure. The model tends to underestimate the degree of time variation for the parameters that are strongly time-varying, while it slightly overestimates the time variation for the weakly varying and constant parameters. This observation holds equally for the two different priors.

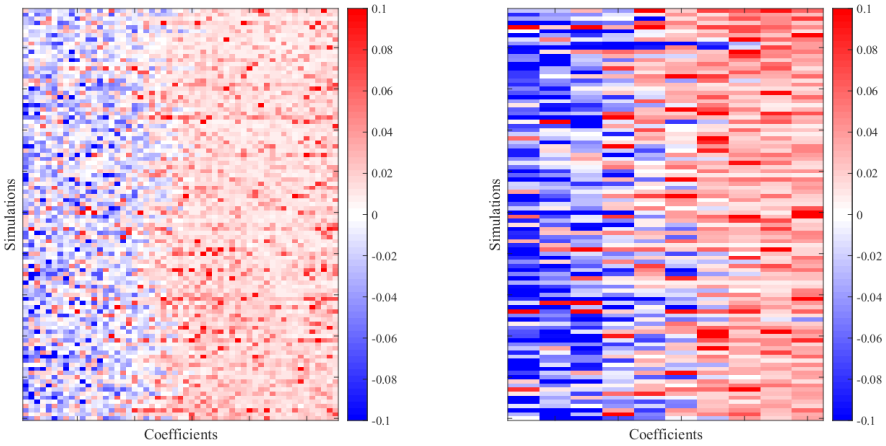

Figure 1: Difference between estimated and effective time variation for b_t (left panel) and a_t (right panel) for the FacTVP model estimated with the NG prior. Coefficients appear in descending order of simulated time variation.

4.1 Comparison to traditional TVP-VAR

For each simulation I estimate a traditional TVP-VAR following the approach of Primiceri (2005), and compute the root mean squared errors for all the parameters of interest and compare them to the estimates above. The errors are computed as the difference between the estimated and the simulated coefficients for each posterior draw. A training sample of length $T_0 = 40$ is used to initialize the priors. The relative RMSE defined as the RMSE of the FacTVP model estimated with the NG prior divided by the RMSE of the TVP-VAR are shown in Table 2.

Table 2: Relative RMSE (average over all coefficients and time periods) of the FacTVP in comparison to the TVP-VAR along with the mean RMSE for the TVP-VAR.

	b_t	a_t	h_t
Mean	0.6596	0.6774	0.9929
Min	0.4639	0.4093	0.8379
Max	0.9844	1.0153	1.3971
RMSE	0.0938	0.1801	0.3326

There is an obvious gain in precision for FacTVP for both, the autoregressive coefficients b_t as well as the covariance states a_t. This gain is mainly related to the fact that parameters that are effectively constant or exhibit only a minor degree of time variation are fitted much better in the case of the FacTVP model. This is well illustrated in Figures 2 and 3 that show the difference between the estimated and the true time variation for b_t and a_t for all the simulations. The coefficients are ordered according to their true degree of time variation, starting with those that exhibit a strong variation on the left side of the figure (the left panels correspond to the two panels of Figure 1). The FacTVP model with the NG prior has a tendency to slightly underestimate the degree of time variation for the parameters that vary a lot over time (blue squares on the left hand side of the left panel of Figure 2 and 3). On the other hand, the degree of time variation for parameters which only slightly vary over time or are constant is relatively precisely estimated while the TVP-VAR has a strong tendency of overestimating the time variation in that case (red squares in the right panel of Figures 2 and 3). For the stochastic volatilities, the FacTVP model does slightly worse, that is however not very astonishing, because the block of the sampler that is related to the volatilities is exactly the same in both models, so there is no shrinkage or dimension reduction involved.

5 Application to the Swiss economy

The Swiss economy turns out to be an interesting laboratory for the use of TVP models. Switzerland is a small open economy with a strong international currency and a large financial sector. During the financial crisis, all major central banks dramatically lowered short term interest rates, so did the Swiss National Bank (SNB). In August 2009 it narrowed its target band for the 3 month London Interbank Rate in CHF (3M-Libor) to 0-0.25 percent. It found itself in an unpleasant situation when

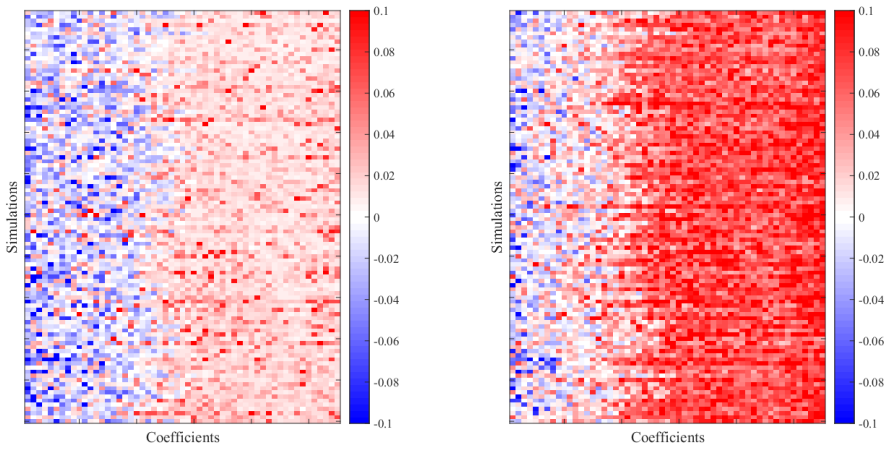

Figure 2: Difference between estimated and effective time variation for b_t for the FacTVP model with NG prior (left panel) and the TVP-VAR (right panel). Coefficients appear in descending order of simulated time variation.

the outbreak of the European debt crisis in 2010 led to strong appreciation pressure on the Swiss franc. The export sector accounts for a large fraction of Swiss GDP, hence changes in the exchange rate especially vis-à-vis the Euro have a direct effect on the demand of Swiss export goods. Lowering interest rates further to achieve a sufficient interest rate differential to the euro area to ease appreciation pressures was not an option by that time. Instead, SNB engaged in "quantitative easing", mainly in the form of foreign exchange acquisition which led to a substantial increase in the length of SNB's balance sheet.[4] In summer 2011 appreciation pressures accelerated strongly and the Swiss franc reached almost parity vis-à-vis the Euro. This finally led to the introduction of a minimum exchange rate of 1.20 Swiss francs against the Euro on September 6, 2011, which SNB announced to enforce by all available means. This step was a profound change in the conduct of Swiss monetary policy. In December 2014 for the first time in its history the SNB introduced negative short-term interest rates by changing its target band for the 3M-Libor from $0 - 0.25$ percent to $-0.75 - 0.25$ percent. A more dramatic move followed on January 15, 2015 when SNB announced that it would discontinue the minimum exchange rate and lower the target band for its policy rate deeper into negative territory to -1.25 – -0.25 percent. The sudden discontinuation of the exchange rate floor took market participants by surprise. In the following, the Swiss franc appreciated strongly versus

[4]In the second half of 2016 foreign exchange reserves amounted to 100 percent of Switzerland's nominal annual GDP.

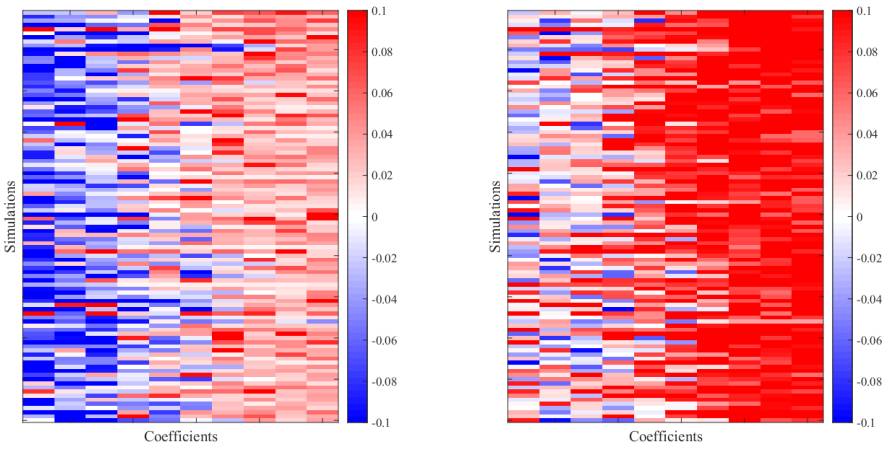

Figure 3: Difference between estimated and effective time variation for a_t for the FacTVP model (left panel) and the TVP-VAR (right panel). Coefficients appear in descending order of simulated time variation.

the Euro.[5] It is obvious that a VAR with constant parameters is not suitable to analyze a period where short term interest rates are stuck at their respective lower bound. In addition, the presence of an exchange rate floor is another non-linearity that needs to be accounted for when looking at the Swiss case.

5.1 The impact of an appreciation shock

In the following, I study the effect of a nominal exchange rate shock on the Swiss economy to understand the changes in the systematic behavior of the endogenous variables once the central bank is constrained by the lower bound on short term interest rates. For this purpose, I estimate the FacTVP model for monthly data including the CPI inflation rate, the monthly business cycle index (BCI) for Switzerland, the spread between the 10 year government bond yield and the 3M-Libor, the index for the nominal effective exchange rate (NEER), and the 3M-Libor.[6] The data set covers the time period from January 1990 up to February 2018, the sample period is chosen according to the availability of the data. The data series are standardized prior to the estimation to avoid scaling issues. To allow for sufficient flexibility in the time-varying behavior of the parameters I include two factors for the intercepts, the autoregressive coefficients and covariance states each, i.e. $k_0 = k = k_a = 2$. The computations are based on the NG-prior with factor-specific shrinkage for b_t and a_t and global shrinkage for the intercepts $b_{0,t}$. The prior hyperparameters are set to $a_1 = a_2 = 0.001$ for the factor loadings associated with the two factors that govern the time variation and $a_1^c = a_2^c = 0.01$ for the factor loadings capturing the constant part. The same settings are used for the factor loadings of the covariance states, while for the time-varying intercepts $a_1^0 = a_2^0 = 0.01$. After a burn in phase of 100,000 iterations the sampler is run for an additional 100,000 iterations out of which every 10^{th} draw is stored. To check the stability of the system, the Lyapunov exponent is computed as outlined in Section 2.3 and only draws with a negative Lyapunov exponent are kept.

[5] For a detailed summary of Swiss monetary history see Baltensperger and Kugler (2017)

[6] The time series are available from the SNB's online database (data.snb.ch). A detailed description of the BCI and how it is computed is provided in Galli (2017)

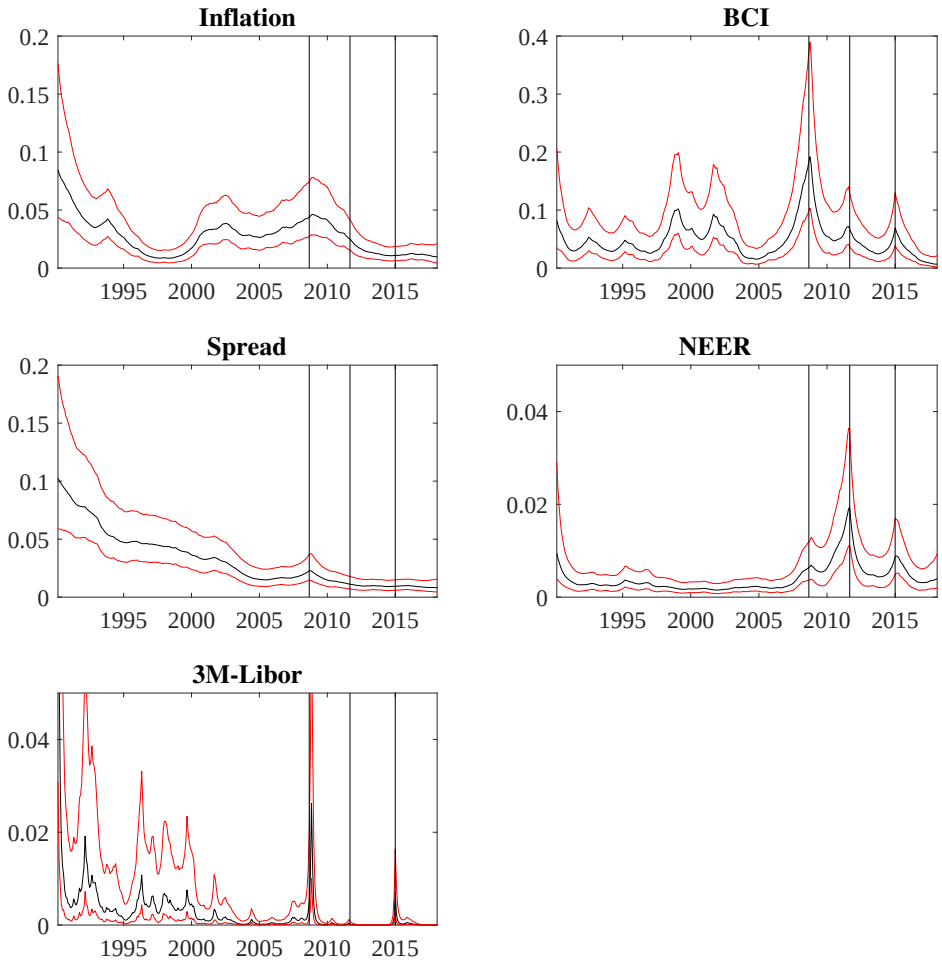

Figure 4: Volatility of reduced form errors, median (in black) along with 68% HPDI (marked in red). The three vertical lines indicate the collapse of Lehman Brothers in September 2008, the introduction of the exchange rate floor vis-à-vis the Euro by the SNB in September 2011 and its discontinuation in January 2015.

Figure 4 shows the time-varying volatilities of the reduced form errors. It reveals that each of them experienced substantial movements throughout the sample period. The three vertical lines indicate the collapse of Lehman Brothers in September 2008, the introduction of the exchange rate floor vis-à-vis the Euro by SNB in September 2011 and its discontinuation in January 2015. During the financial crisis the error volatility in inflation, the BCI and the nominal exchange rate went up, and the reduction of the interest rates left behind a spike in the volatility of the errors associated with the 3M-Libor. In addition the introduction, as well as the abolishment of the minimum exchange rate are clearly visible in the volatilities, especially for the NEER and the

BCI.

The degree of time variation present in the remaining parameters of the VAR is shown in Figure 5. It reports the estimated standard deviation of the different parameters over time (median over all posterior draws). Interestingly there is almost no time variation in the intercepts (left column of the upper panel in Figure 5). The autoregressive coefficients (marked with B_1 and B_2) show very little time variation. Time variation additionally plays a crucial role for the covariance states (A in the lower panel of Figure 5).

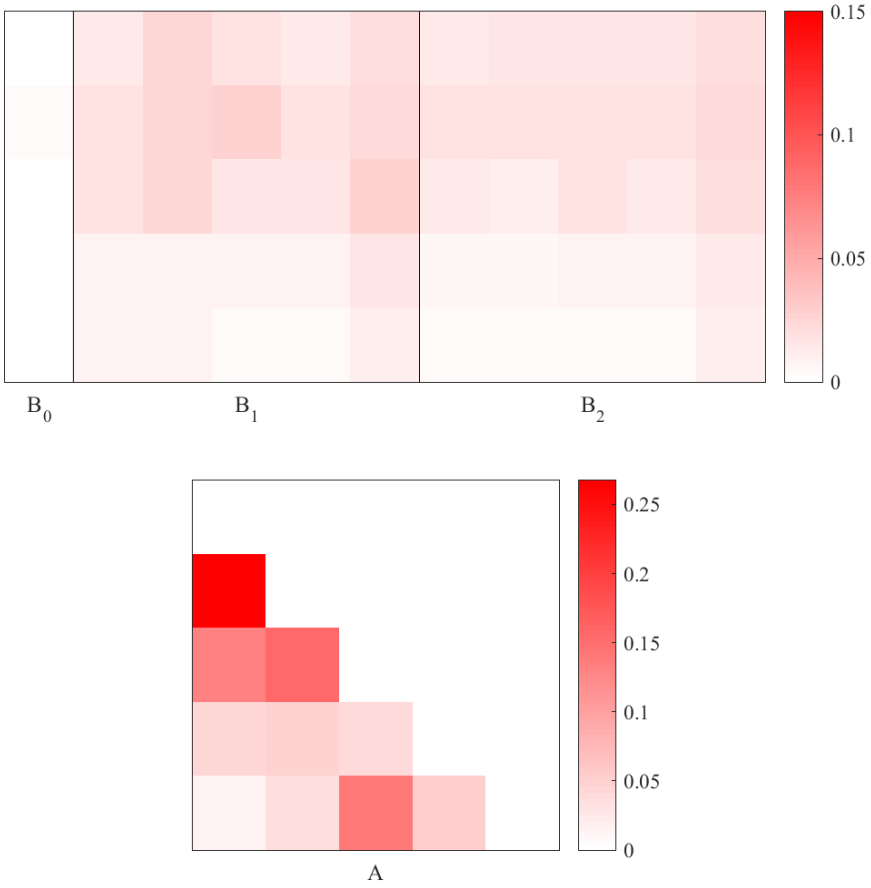

Figure 5: Time variation (median standard deviation over time) of the intercepts and autoregressive coefficients (upper panel) and of the covariance states (lower panel).

Figure 6 shows the impulse response of the 3M-Libor to an appreciation shock, normalized to a 10 point increase of the NEER, at the 4-month horizon. The exchange rate shock is identified using a simple recursive scheme based on the Cholesky decomposition of the time-varying covariance matrix of the reduced form VAR. The NEER is ordered second to last such that it is the only variable along with

the 3M-Libor (which is ordered last) that can respond to the shock on impact. While at the beginning of the sample period an appreciation shock triggers a reduction of the short term interest rate, the effect on the interest rate shrinks along with the decreasing level of the short-term interest rate and is basically zero during the lower bound period that started when the SNB lowered its target band for the 3M-Libor to 0-0.25 percent in August 2009 (indicated by the black vertical line). Figure 7 shows the median impulse responses of inflation, the BCI, the spread and the 3M-Libor to an appreciation shock for the period before interest rates approached the lower bound (black line) and lower bound period (red line). An appreciation shock has a negative effect on the real economy, it lowers the spread and leads to deflationary pressure (although the uncertainty for the response of inflation is large). Once the economy is at the lower bound and short-term interest rates cannot be lowered, the negative effect on the BCI is slightly more persistent and deflationary pressures are slightly stronger while the depression of the spread is also slightly stronger. Although the estimated degree of time variation among the model's parameters turns out to be muted the model is still able to mimic an economically sensible transformation in the dynamics of the underlying economy.

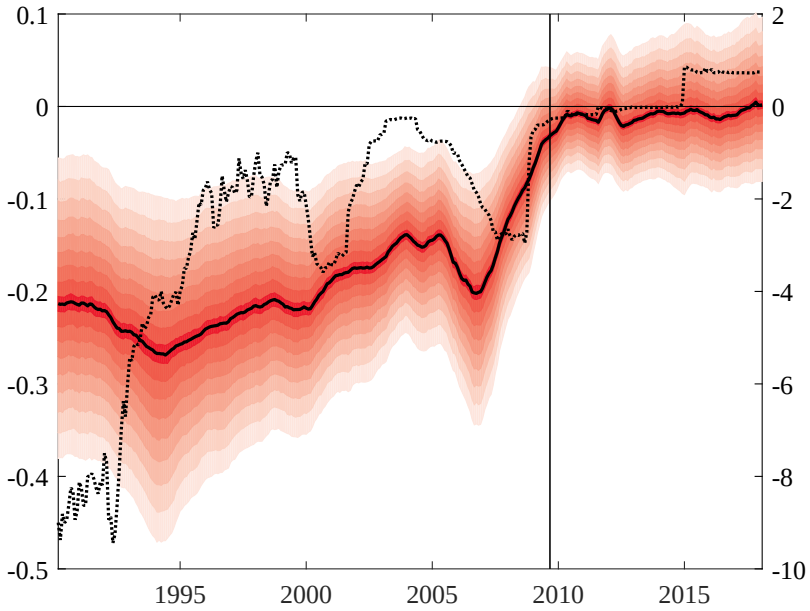

Figure 6: Impulse response of the 3M-Libor to an appreciation shock at the 12-month horizon (left axis) along with the level of the 3M-Libor multiplied by minus one (dotted line, right axis).

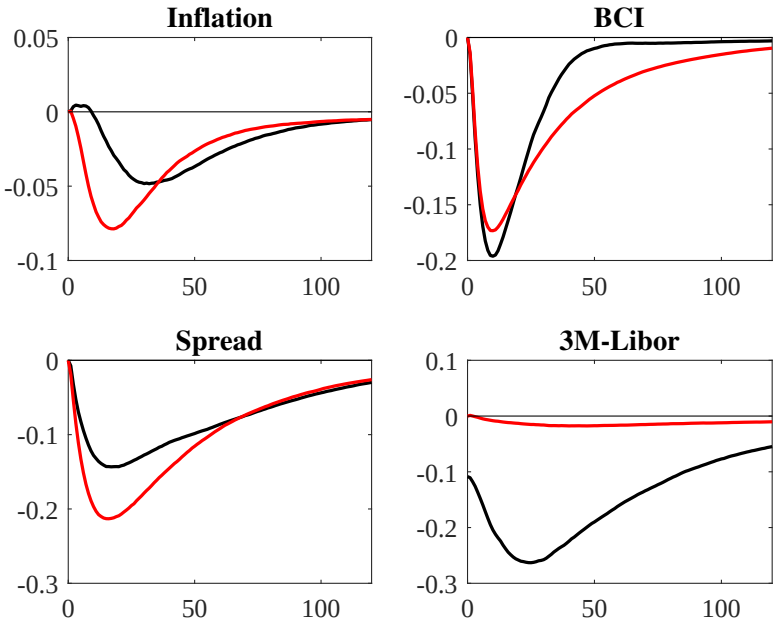

Figure 7: Impulse responses to an exchange rate shock (appreciation) during "normal times" (black line) and during lower bound period (red line).

The small amount of time variation present in the autoregressive coefficients raises the question whether it would be enough to let only the covariance states a_t and the volatilities h_t be time-varying, while holding b_t constant. However, it turns out that this is not the case. Figure 8 presents the evolution of the 3M-Libor's median response to the exchange rate appreciation shock over the estimation period for three different models. The blue line corresponds to the full model where two factors for each block are included, i.e. $k_a = k_b = 2$. The red line corresponds to the model where b_t are not allowed to change over time ($k_a = 2$, $k_b = 0$) while the yellow line represents the case where a_t is fixed ($k_a = 0$, $k_b = 2$).

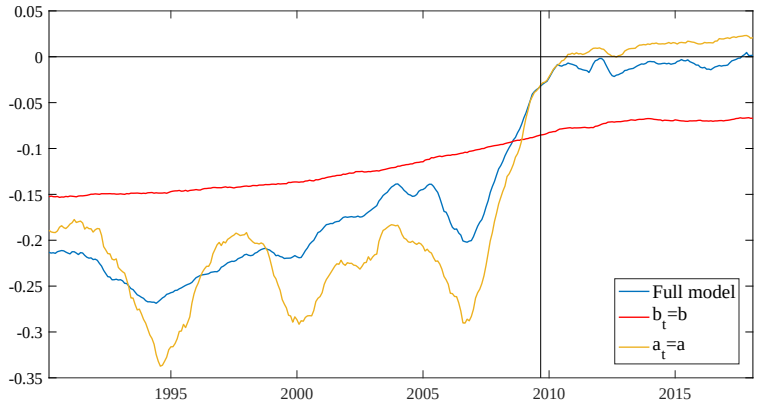

Figure 8: Median impulse response of the 3M-Libor to an appreciation shock at the 12-month horizon for the full model (blue), a model where only A_t varies over time (red) and a model where only b_t is time-varying (yellow).

It demonstrates impressively that, although barely visible, there is still important action among the elements of b_t. Fixing b_t to constant values for the whole sample period implies that only the impact matrix changes for the impulse responses while the propagation dynamics remain unchanged. This is clearly not sufficient to capture the break that occurred in the late 2000s when the short-term interest rate was lowered to almost zero. The impulse response of the model with time-varying VAR coefficients (in yellow) is much closer to the corresponding impulse response of the full model, while the one stemming from the model b_t held constant (in red) shows only a slight change over time. Although the full model and the model where a_t is held constant deliver similar results for the median impulse responses, the uncertainty around these estimates increases substantially when a_t is fixed (not shown in the figure for the sake of clarity). This is most likely due to the fact that when a_t is fixed the time variation present in the autoregressive coefficients increases quite a bit compared to the estimates of the full model. The time variation of a_t is basically shifted into b_t in this case.

5.2 A small forecasting exercise

To test the model's forecasting ability I perform a small forecasting exercise using the same variables as outlined above and compare the outcomes to the forecasts obtained with a traditional TVP-VAR. The forecast horizon is set to one year, i.e $h = 12$ because of the monthly frequency of the data. The forecast period starts in February 2014 which leads to a total of 50 estimation samples where for each sample one observation is added to the information set at a time. The two models

are then subsequently re-estimated for each estimation sample and forecasts for all the endogenous variables are computed. In order to evaluate the accuracy of the forecasts I compute the average root mean squared forecast errors (ARMSFE) for each forecasting horizon over the 50 samples. Table 3 reports the relative ARMSFE of the FacTVP model to the TVP-VAR (a value below 1 means better accuracy of FacTVP). It turns out that the FacTVP model performs relatively well, the relative ARMSFE for all variables and all horizons lie below one indicating a better performance of the FacTVP.

Table 3: ARMSFE of the FacTVP model relative to the traditional TVP-VAR at different forecast horizons for the five endogenous variables.

Horizon	Inflation	BCI	Spread	NEER	3M-Libor
1	0.8199	0.8495	0.8608	0.9510	0.9596
2	0.8097	0.9264	0.8194	0.9764	0.9491
3	0.7960	0.9547	0.8390	0.9222	0.9433
4	0.8322	0.9313	0.8188	0.9110	0.9211
5	0.8351	0.9702	0.7889	0.9385	0.8986
6	0.8303	0.9428	0.7663	0.9468	0.8789
7	0.8571	0.8657	0.7142	0.9525	0.8718
8	0.8868	0.8361	0.7096	0.9448	0.8619
9	0.9001	0.8521	0.7047	0.9512	0.8422
10	0.9276	0.8804	0.6683	0.9770	0.8294
11	0.9198	0.8632	0.6356	0.9766	0.8318
12	0.9234	0.8238	0.6162	0.9886	0.8231

6 Application to historical inflation

Allowing for time-varying parameters becomes especially interesting when working with data sets that cover a long time span. This section presents some results based on the estimation of the FacTVP model with data on historical inflation rates for the United Kingdom (UK), Norway (NOR), Sweden (SWE) and the United States (US) covering the period from 1830 to 2017 shown in Figure 9. The four series measure annual year-on-year inflation rates either based on GDP deflators (UK and US) or consumer prices (NOR and SWE).[7] Similar to the previous example I include

[7]The data for the UK and the US is available from the Bank of England: A millennium of macroeconomic data. The data for Norway stems from Norges Bank, and the series for Sweden is available from the Warren E. Weber Historical Data Archives at the Federal Reserve Bank

two factors for the intercepts, the autoregressive coefficients and the covariance states each (i.e. $k_0 = k = k_a = 2$) to achieve a sufficient level of flexibility.[8] The computations are based on the NG-prior with factor-specific shrinkage for b_t and a_t and global shrinkage for the intercepts $b_{0,t}$. The prior hyperparameters are set to $a_1 = a_2 = 0.001$ for the factor loadings associated with the two factors that govern the time variation in the different parameters and $a_1^c = a_2^c = 0.01$ for the factor loadings capturing the constant part. The same settings are used for the factor loadings of the covariance states, while for the time-varying intercepts $a_1^0 = a_2^0 = 0.01$ is imposed. After a burn in phase of 100,000 iterations the sampler is run for an additional 100,000 iterations out of which every 10^{th} draw is stored. To check the stability of the system, the Lyapunov exponent is computed as outlined in Section 2.3 and only draws with a negative Lyapunov exponent are kept when sampling.

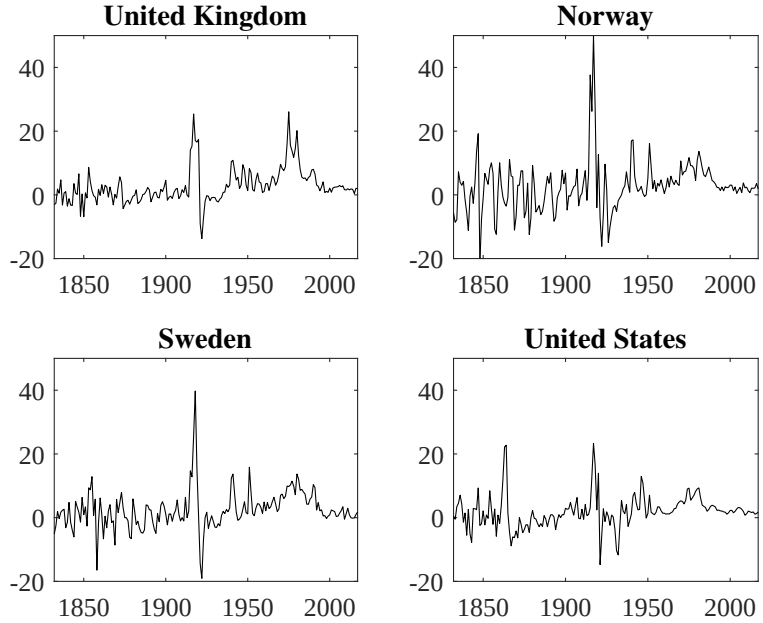

Figure 9: Annual inflation series in percent from 1830 to 2017.

Figure 10 presents the estimated error volatilities for the four countries in the sample. There is notable movement over time among all of them. The four volatility series further share certain commonalities such as the spikes around 1850 and around 1920

of Minneapolis. When necessary the series have been linked with actual data from the FRED database to extend them up to 2017.

[8]Since the sample only covers inflation series whose properties are expected to be somewhat similar one could also think of estimating the model with only one factor for each block. In fact that changes the quality of the results only slightly, it reduces the overall amount of time variation especially among the off-diagonal elements of $B_{\cdot,t}$

in the aftermath of World War I, as well as a strong decline in the volatility of errors towards the end of the sample period.

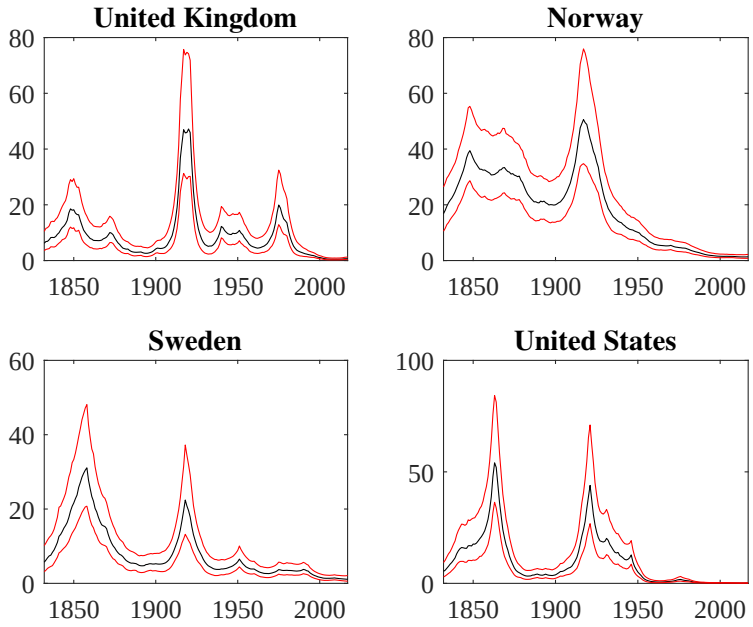

Figure 10: The estimated volatilities of the errors (median in black along with the 68% HPDI in red).

The degree of time variation for the intercepts and the autoregressive coefficients is substantially larger than in the previous example for the Swiss economy, and is especially large for the first own lag of each variable (i.e. the diagonal elements of $B_{1,t}$). The evolution of these four coefficients is presented in Figure 12. The size of the first own lag of each series steadily increases over the first part of the sample, then this increase comes to a halt and in the last part of the sample the coefficients start to decrease again. This pattern points towards an overall increase in persistence present in the inflation rates of the four countries. Indeed, having a look at the sum of the two included own lags for each series (shown in Figure 13) reveals that the estimated persistence increased for the UK and Norway and to a much lesser extent also for Sweden where it stayed at a muted level, while it remained more or less constant for US inflation. Without time variation in the parameters we would clearly not be able to capture such interesting features of the data.

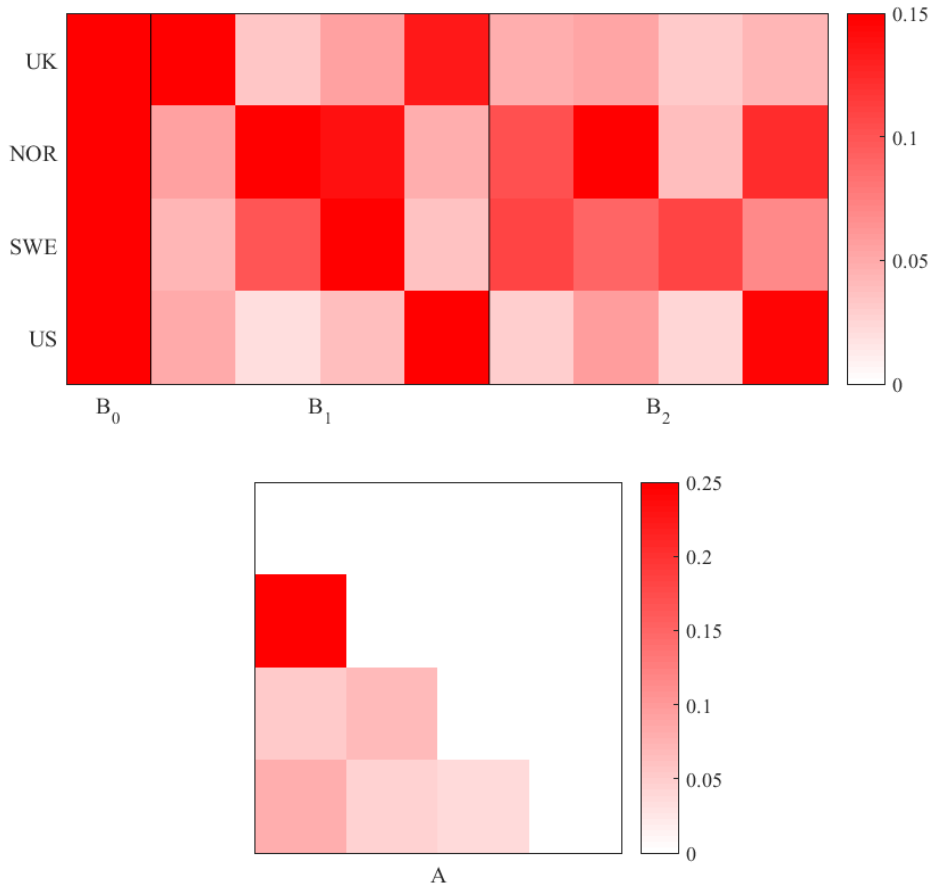

Figure 11: Time variation of the intercepts and autoregressive coefficients (upper panel) and of the covariance states (lower panel).

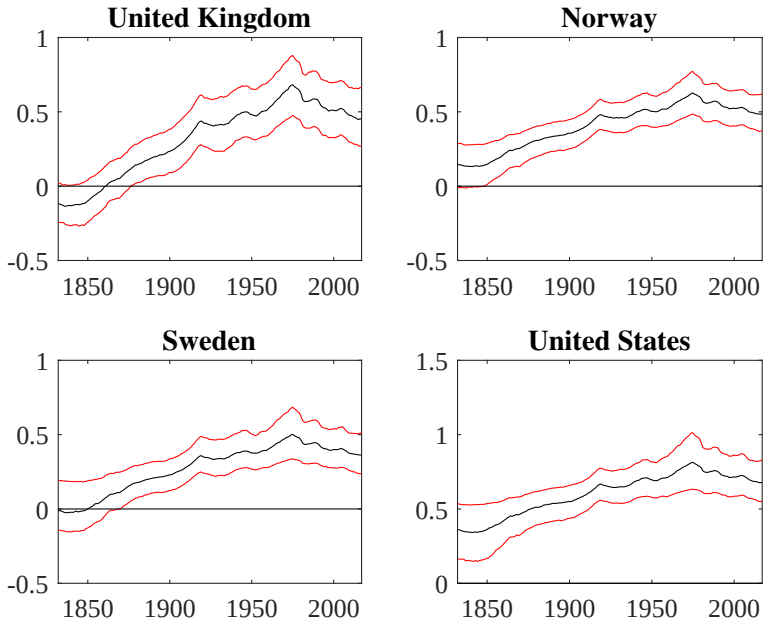

Figure 12: The evolution of the first own lag of each inflation series over time (median in black along with the 68% HPDI in red).

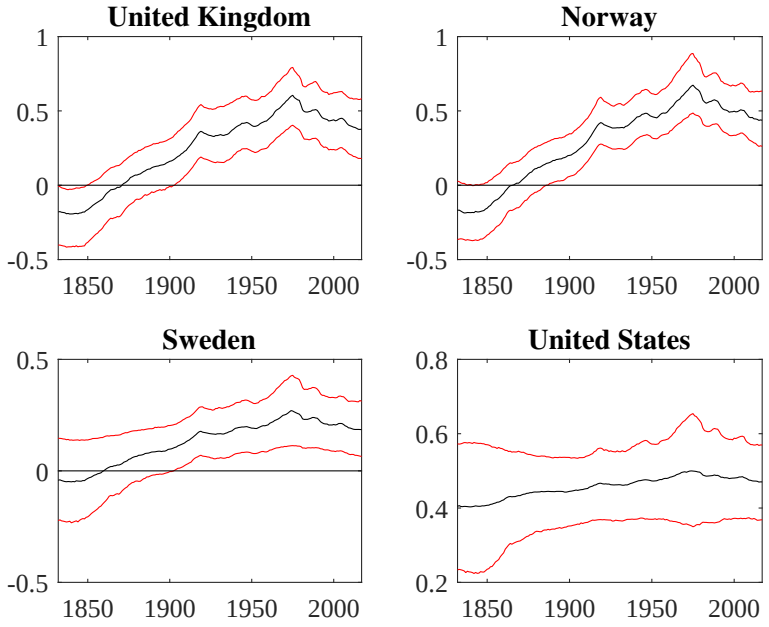

Figure 13: The evolution of the sum of own lags of each inflation series over time (median in black along with the 68% HPDI in red).

7 Conclusion

In this paper I present a novel approach to estimate a Bayesian TVP-VAR assuming a factor structure for the time-varying coefficients and covariance states. The factor structure leads to a strong reduction in the dimension of the model's state space, rendering the TVP models suitable for large VARs. In combination with Bayesian shrinkage priors, the approach also allows for a better distinction between constant and truly time-varying parameters. Marginal data augmentation is applied to stabilize the factor's mean when sampling, and to control the overall amount of time variation in the parameters. A Monte Carlo study with simulated data illustrates the ability of the proposed sampler to correctly estimate the degree of time variation and its superiority to distinguish between constant and truly time-varying parameters when compared to the traditional TVP-VAR approach.

In an empirical application with monthly data for the Swiss economy the model is well able to capture structural changes in the economic structure related to the fact that monetary policy operated at the lower bound of nominal interest rates. It delivers sensible impulse responses for the case of an appreciation shock. The effect of the shock on the nominal interest rate strongly diminishes when the economy enters the lower bound period. A small forecasting exercise based on the same dataset reveals that the forecasting performance of the FacTVP model turns out to be better compared to a TVP-VAR without dimension reduction.

A second application with data on historical inflation shows that the role for time variation in the model parameters gains in importance when the time span of the sample period increases.

Interesting future steps include extending the dimension reduction to the time varying volatilities in a similar fashion and applying the model to larger data sets where the possible gains of the dimension reduction and the role for shrinkage are even greater.

References

Amir-Ahmadi, P., C. Matthes, and M.-c. Wang (2016): "Choosing Prior Hyperparameters," .

Baltensperger, E. and P. Kugler (2017): *Swiss Monetary History since the Early 19th Century*, Studies in Macroeconomic History, Cambridge University Press.

Belmonte, M. A. G., G. Koop, and D. Korobilis (2014): "Hierarchical shrinkage in time-varying parameter models," *Journal of Forecasting*, 33, 80–94.

Bitto, A. and S. Frühwirth-Schnatter (2016): "Achieving Shrinkage in a Time-Varying Parameter Model Framework," .

Bougerol, P. and N. Picard (1992): "Strict stationarity of generalized autoregressive processes," *The Annals of Probability*, 20, 1714–1730.

Brandt, A. (1986): "The stochastic equation $y_{n+1} = a_n y_n + b_n$ with stationary coefficients," *Advances in Applied Probability*, 18, 211–220.

Canova, F. and M. Ciccarelli (2009): "Estimating multicountry VAR models," *International Economic Review*, 50, 929–959.

Carriero, A., T. E. Clark, and M. Marcellino (2016): "Common Drifting Volatility in Large Bayesian VARs," *Journal of Business & Economic Statistics*, 34, 375–390.

Carter, C. K. and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553.

Chan, J. and I. Jeliazkov (2009): "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1, 101–120.

Cogley, T. and T. J. Sargent (2005): "Drifts and volatilities: Monetary policies and outcomes in the post WWII US," *Review of Economic Dynamics*, 8, 262–302.

de Wind, J. and L. Gambetti (2014): "Reduced-rank time-varying vector autoregressions," CPB Discussion Paper 270, CPB Netherlands Bureau for Economic Policy Analysis.

Eisenstat, E., J. C. C. Chan, and R. W. Strachan (2018): "Reducing Dimensions in a Large TVP-VAR," *mimeo*.

Elaydi, S. (2013): *An Introduction to Difference Equations*, Undergraduate Texts in Mathematics, Springer New York.

Francq, C. and J.-M. Zakoïan (2001): "Stationarity of multivariate Markovswitching ARMA models," *Journal of Econometrics*, 102, 339 – 364.

Galli, A. (2017): "Which indicators matter? Analyzing the Swiss business cycle using a large-scale mixed frequency dynamic factor model," *SNB Working Paper*, 8/2017.

George, E. I., D. Sun, and S. Ni (2008): "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, 142, 553–580.

Griffin, J. E. and P. J. Brown (2010): "Inference with normal-gamma prior distributions in regression problems," *Bayesian Analysis*, 5, 171–188.

Huber, F. and M. Feldkircher (2017): "Adaptive Shrinkage in Bayesian Vector Autoregressive Models," *Journal of Business & Economic Statistics*, 2, 1–13.

Kastner, G. (2016): "Sparse Bayesian time-varying covariance estimation in many dimensions," *ArXiv e-prints*.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models Inference and Comparison with ARCH Models," *The Review of Economic Studies*, 65, 361–393.

Lucas, R. E. J. (1976): "Econometric policy evaluation: A critique," *Carnegie-Rochester Conference Series on Public Policy*, 1, 19–46.

Nakajima, J. (2011): "Time-Varying Parameter VAR Model with Stochastic Volatility: An Overview of Methodology and Empirical Applications," *Monetary and Economic Studies*, 29, 107–142.

Neusser, K. (2018): "Time-Varying Rational Expectations Models: Solutions, Stability, Numerical Implementation," .

Park, T. and G. Casella (2008): "The Bayesian Lasso," *Journal of the American Statistical Association*, 103, 681–686.

Primiceri, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821–852.

Stevanovic, D. (2016): "Common time variation of parameters in reduced-form macroeconomic models," *Studies in Nonlinear Dynamics and Econometrics*, 20, 159–183.

Stock, J. H. and M. W. Watson (1996): "Evidence on structural instability in macroeconomic time series relations," *Journal of Business & Economic Statistics*, 14, 11–30.

A Precision sampler

Stack all observations to obtain the matrix representation

$$\tilde{\mathbf{Y}} = \mathbf{\Lambda}^f F + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(0, \mathbf{\Sigma}) \tag{24}$$

$$\mathbf{\Phi}^f F = \boldsymbol{\eta}^f \quad \boldsymbol{\eta}^f \sim N(0, \boldsymbol{Q}) \tag{25}$$

where $\tilde{\mathbf{Y}} = \left(y_{p+1}^{f\prime}, \ldots, y_T^{f\prime}\right)'$ contains all data, $F = ({f_{1-p}}', \ldots, {f_{p+1}}', \ldots, {f_T}')'$ stacks all unobserved factors, including initial states. The matrices $\mathbf{\Lambda}^f$ and $\mathbf{\Phi}^f$ are respectively of dimension $(T-1)N \times (T+p)\,k$ and square $(T+p)k$. Typically, these matrices are sparse and banded around the main diagonal Chan and Jeliazkov (2009)

$$\mathbf{\Lambda}^f = \left[\begin{array}{c|cccc} & \lambda^{*f} & 0 & \cdots & 0 \\ \mathbf{0}_{(T)N\times k} & 0 & \ddots & \ddots & \vdots \\ & & & \ldots & \lambda^{*f} \end{array}\right]$$

$$\mathbf{\Phi}^f = \left[\begin{array}{cccccc} I_k & & 0 & \ldots & & \\ \hline -I_k & I_k & 0 & \ldots & & \\ & & \ddots & & & \\ & \ldots & 0 & & \ldots & -I_k \quad I_k \end{array}\right],$$

$$\mathbf{\Sigma} = \begin{bmatrix} \Sigma_0 & 0 & \ldots & 0 \\ 0 & \Sigma_1 & & \vdots \\ \vdots & & \ddots & \\ 0 & \ldots & & \Sigma_T \end{bmatrix}, \quad \boldsymbol{Q} = \begin{bmatrix} Q & 0 & \ldots & 0 \\ 0 & Q & & \vdots \\ \vdots & & \ddots & \\ 0 & \ldots & & Q \end{bmatrix}$$

where Σ_0 represents the variance of the initial states of the unobserved factors. Combining the prior with the likelihood $\pi(\tilde{\boldsymbol{Y}}|F,\theta) \sim N(\mathbf{\Lambda}^f F, \mathbf{\Omega})$ we obtain the posterior distribution

$$\bar{F}|\tilde{\boldsymbol{Y}}, \theta \sim N(\boldsymbol{\mu}_f, \mathbf{F}) \tag{26}$$

$$\mathbf{F}^{-1} = \mathbf{F}_0^{-1} + \mathbf{\Lambda}^{f\prime}(\boldsymbol{Q}^{-1})\mathbf{\Lambda}^f \tag{27}$$

$$\boldsymbol{\mu}_{\bar{f}} = \mathbf{F}\mathbf{\Lambda}^{f\prime}(\boldsymbol{Q}^{-1})\tilde{\boldsymbol{Y}} \tag{28}$$

In order to avoid the full inversion of $\mathbf{F}$ we take the Cholesky decomposition, $\mathbf{F}^{-1} = L'L$, then $\mathbf{F} = L^{-1}L^{-1'}$. We obtain a draw F by setting $F = \boldsymbol{\mu}_f + L^{-1}\boldsymbol{\nu}$, where $\boldsymbol{\nu}$ is a $(T+1)k$ vector of independent draws from the standard normal

distribution.

B Estimating prior hyperparameters

In contrast to the specification of the innovation variances for the log-volatilities in the main text, W is assumed to be a full matrix here. The prior distribution is then set to

$$W \sim IW\left(k_W^2 df_W V_W, df_W\right) \tag{29}$$

where IW stands for the inverted Wishart distribution, and $df_\bullet, V_\bullet$ denotes the degrees of freedom and prior scaling matrices respectively. The standard in the literature for setting the hyperparameters $V_\bullet$, $k_\bullet$ is to use a training sample to pin down the scaling matrices and to set $k_\bullet$ to fixed values, similar to those proposed in Primiceri (2005). An alternative to setting the low-dimensional hyperparameters of the prior variances for the innovations that govern the time variation of the volatilities and the covariance states to fixed values, is to estimate them using the procedure of Amir-Ahmadi et al. (2016). Their approach augments the traditional Gibbs sampler used for these models with an additional Metropolis step for each hyperparameter to be estimated. In the context of the present model this results in one additional step for the scaling factor k_W. Conditional on the model structure and assumptions (3)-(4) stated in Amir-Ahmadi et al. (2016), the acceptance probability of k_W is given by

$$\alpha_W^i = \min\left(1, \frac{p(W|k_W^*)p(k_W^*)q(h_W^{i-1}|k_W^*)}{p(W|k_W^{i-1})p(k_W^{i-1})q(h_W^*|k_W^{i-1})}\right)$$

Amir-Ahmadi et al. (2016) set the prior of each of the hyperparameters to be estimated to an inverse gamma distribution which they parametrize such that the mode is equal to 0.05 while the variance is chosen to be infinite, which leads to $p(k_\bullet) = IG(2, 0.05)$.

B.1 The scaling factors for the prior variances

Instead of setting the low-dimensional hyperparameters of the prior variances for the innovations that govern the time variation of the volatilities and the covariance states to fixed values, as is standard in the literature, they are estimated following the procedure of Amir-Ahmadi et al. (2016). Their approach augments the traditional

Gibbs sampler used for these models with an additional Metropolis step for each hyperparameter to be estimated. In the context of the present model this results in two additional steps, one for each scaling factor k_W, k_S. Conditional on the model structure and assumptions (3)-(4) stated in Amir-Ahmadi et al. (2016), the acceptance probability of $k_X, X \in \{W, S\}$ is given by

$$\alpha_X^g = \min\left(1, \frac{p(X|k_X^*)p(k_X^*)q(h_X^{g-1}|k_X^*)}{p(X|k_X^{g-1})p(k_X^{g-1})q(h_X^*|k_X^{g-1})}\right)$$

where g denotes the iteration of the Gibbs-sampler.

1. Get a candidate draw k_X^* from $N(k_X^{g-1}, \sigma_{k_X}^2)$.
2. Compute the acceptance probability $\alpha_{k_X}^g$.
3. Accept the candidate draw with probability $\alpha_{k_X}^g$.

During the first half of the burn-in phase an additional automatic tuning step for the proposal variance $\sigma_{k_X}^*$ is used to achieve a target acceptance rate α^*. The target acceptance rate is set to $\alpha^* = 0.5$ as in Amir-Ahmadi et al. (2016).
The posterior sampler presented in Section 3 is then extended by this additional step.

C Convergence of sampler

C.1 Model for the Swiss economy

Figures 14 to 15 show the Markov chain of the 10,000 draws that are stored for some selected parameters of the model. All of the chains appear to be stationary and there is no evidence for jumps or breaks.

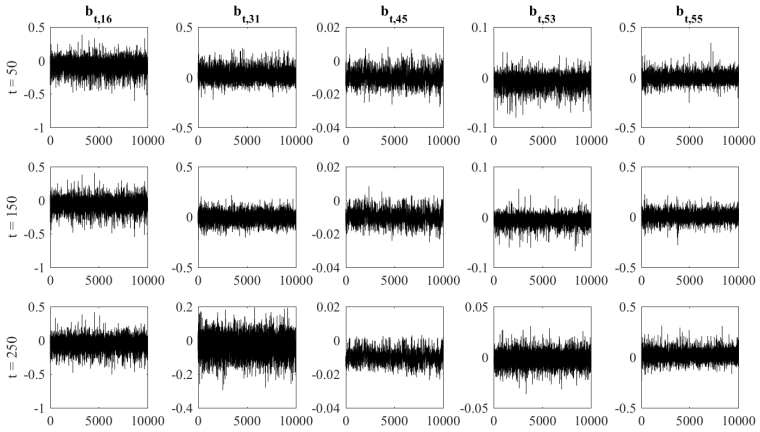

Figure 14: Markov chain for randomly selected elements of the coefficient vector b_t for different time periods.

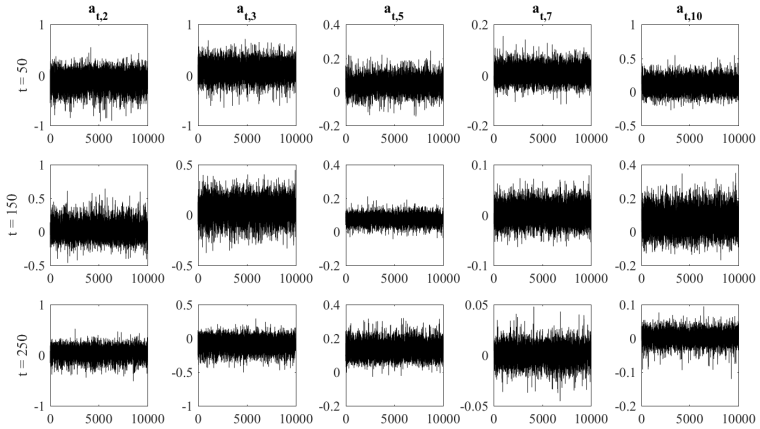

Figure 15: Markov chain for randomly selected elements of the vector a_t for different time periods.

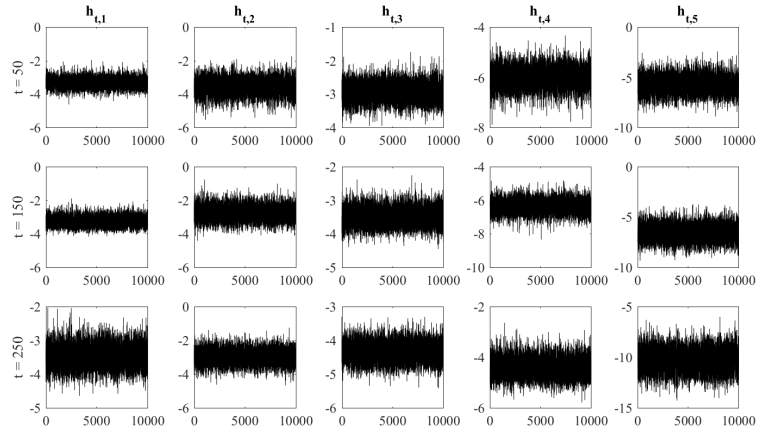

Figure 16: Markov chain for the log-volatilities h_t for different time periods.

C.2 Model with historical inflation data

Figures 14 to 15 show the Markov chain of the 10,000 draws that are stored for some selected parameters of the model. All of the chains appear to be stationary and there is no evidence for jumps or breaks.

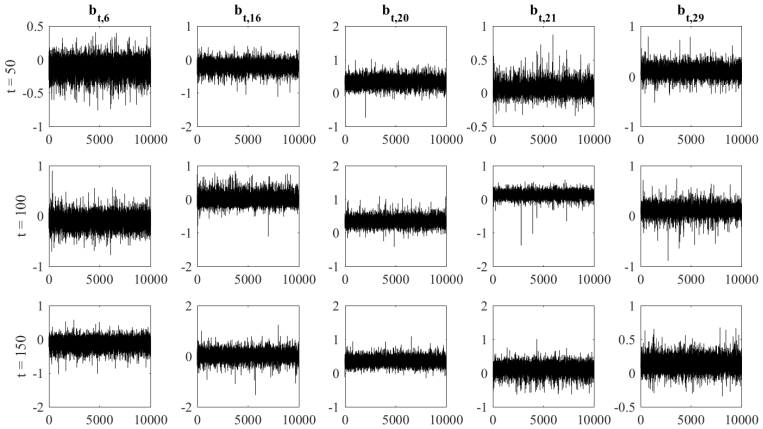

Figure 17: Markov chain for randomly selected elements of the coefficient vector b_t for different time periods.

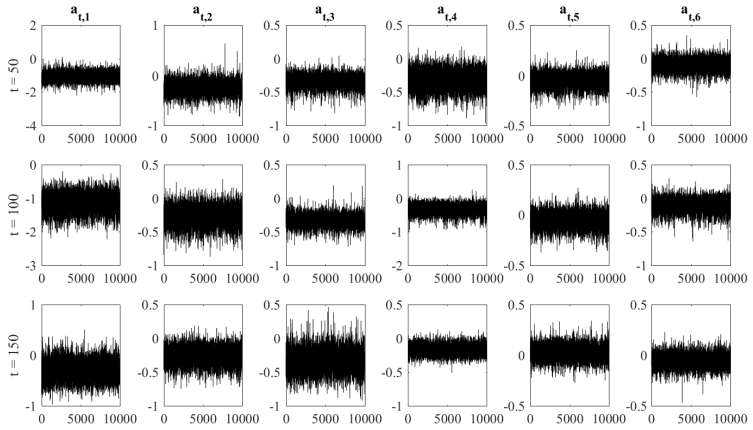

Figure 18: Markov chain for the elements of the vector a_t for different time periods.

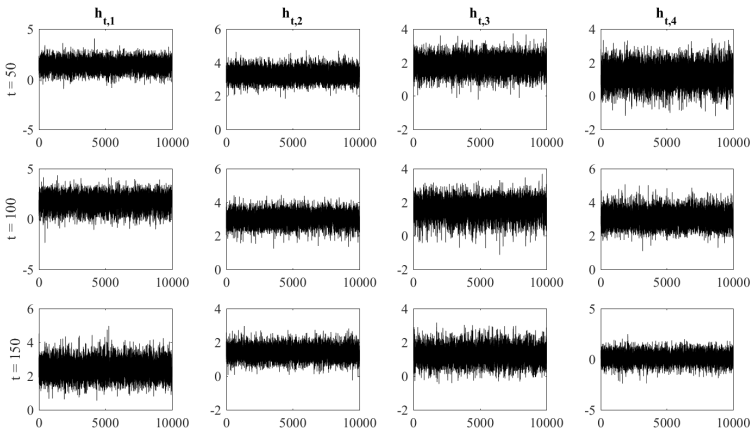

Figure 19: Markov chain for the log-volatilities h_t for different time periods.