# Structural vector autoregression with time varying transition probabilities

Wenjuan Chen* Aleksei Netšunajev[§][¶]

November 3, 2017

**Abstract**

Vector autoregressive models with regime-switching variances have been exploited to test structural assumptions in vector autoregressions. However, these models normally assume the transition probabilities to be constant over time. In reality these probabilities could depend on certain economic fundamentals that help predicting turning points. This paper is the first to introduce time varying probabilities into structural vector autoregressive models. A generalized EM algorithm is updated for estimation of the model. For empirical illustration, we apply the model to test whether macroeconomics variables can immediately respond to uncertainty shocks or not. Results are as follows. First, the hypothesis of constant transition probability is rejected. Second, the time-varying probability model outperforms the constant probability model. Third, a formal test strongly rejects the hypothesis that macroeconomics variables cannot immediately respond to uncertainty shocks.

**Keywords**: Structural vector autoregression, Markov switching, time varying transition probabilities, identification via heteroskedasticity, uncertainty shocks, unemployment dynamics.
**JEL classification**: C32, D80, E24.

*Freie Universitat Berlin, Boltzmannstr 20, 14195 Berlin, Germany, e-mail: wenjuan.chen@fu-berlin.de.

[§]corresponding author, Tallinn University of Technology and Eesti Pank, Estonia pst 13, 15095 Tallinn, Estonia, e-mail: aleksei.netsunajev@eestipank.ee.

[¶]The views and opinions expressed in this paper are those of the authors and do not represent an official position of Eesti Pank or the Eurosystem.

# 1 Introduction

In structural vector autoregressive (SVAR) models, it is critical to come up with convincing identification of structural shocks, since impulse response analysis could be sensitive to various restrictions assumed for identification purposes. Besides, it is often hard to find enough economic theories to justify the identifying restrictions. Therefore, in recent years a large number of papers have used statistical properties of the data, for instance heteroskedasticity, for identification purposes (Rigobon and Sack (2003), Lanne and Lütkepohl (2008), and Lütkepohl and Netšunajev (2017b)). In particular, VAR models with Markov regime switching in variances (see Lanne, Lütkepohl and Maciejowska (2010), and Herwartz and Lütkepohl (2014)) are widely applied and can be employed to test different types of structural identification schemes.

However, in the strand of literature that uses Markov-switching in variances for identification, it is assumed that the transition probabilities are constant over time. While in reality these probabilities can actually vary over time and depend on some underlying economic fundamentals. Several studies, including Diebold, Lee and Weinbach (1994), Filardo (1994), and Bazzi, Blasques, Koopman and Lucas (2017), have investigated the possibility of relaxing the constant transition probability assumption. Diebold et al. (1994) introduce a univariate process with time varying probabilities where the probabilities evolve as a logistic function of underlying fundamentals. Simulation results show that the smoothed state probabilities generated by that model tracks the true states better compared with the model with constant transition probabilities. Filardo (1994) demonstrates that by incorporating certain economic indicators such as the federal funds rates, the regime switching models with time varying transition probabilities can better capture and predict the expansion and contraction phases in the U.S. output compared with regime switching models with constant probabilities.

Advancing on Herwartz and Lütkepohl (2014), this paper is the first to introduce time-varying transition probabilities into regime-switching structural VAR models so as to identify structural shocks. For estimation purpose, we develop a generalized expectation maximization (EM) algorithm. We allow the probabilities of a regime shift to change over time and let it depend on economic variables that help predict turning points. Compared with Diebold et al. (1994), the algorithm developed in the paper uses the filter proposed by Kim (1994), which makes the expectation step more efficient. This algorithm is also more general as it is extended from univariate to a multivariate framework with the application to SVAR. Moreover it can be used not only for two Markov regimes, but also for the case with three or more regimes. Based on the algorithm, we further adopt statistical tests to discriminate between competing conventional identification schemes, which is in the spirit of Lanne and Lütkepohl (2008) and Herwartz and Lütkepohl

(2014).

As an empirical illustration of the new model we study the relationship between uncertainty and the U.S. macroeconomic series using a system similar to the one used by Caggiano, Castelnuovo and Groshenny (2014). Starting from the seminal work by Bloom (2009) there is a growing number of papers studying the role of uncertainty in the economy (see Alexopoulos and Cohen (2009), Bachmann, Elstner and Sims (2013), Colombo (2013), Nodari (2014), and Baker, Bloom and Davis (2016)). In particular, linear structural VAR models for identification of uncertainty shocks are used in many empirical papers. In our study a VAR system consisting of economic policy uncertainty, unemployment, inflation and the federal funds rates is considered.

Our estimation results shed new light on several aspect. Most importantly, the hypothesis of constant transition probabilities are strongly rejected against the alternative hypothesis of time-varying transition probabilities. Moreover, according to the information criteria and the likelihood of the model, the Markov switching model with time varying transition probabilities outperform the standard model with constant transition probabilities. The choice of the economic fundamental to govern the transition probabilities also plays a key role in our analysis. Therefore, we estimate models with many alternative candidates, such as lagged unemployment, lagged federal funds rates, and lagged GDP growth rates. Following Auerbach and Gorodnichenko (2012), Bachmann and Sims (2012), Berger and Vavra (2014) and Caggiano et al. (2014), we also consider the moving averages of seven quarter-on-quarter GDP growth rates as a suitable candidate. It turns out that the seven-quarter moving averages of GDP growth is the most preferred transition variable according to the information criteria.

Further, our model allows to test identifying assumptions formally while in the conventional SVAR setup it is impossible to discriminated between different structural restrictions. The VAR studies on uncertainty shocks typically assume recursive zero restrictions, including Alexopoulos and Cohen (2009), Nodari (2014) and Caggiano et al. (2014). In particular, two different types of zero restrictions based on Cholesky decomposition are considered in Caggiano et al. (2014). The difference is the place of uncertainty measure in the vector of variables. The specification with the uncertainty measure ordered first allows for contemporaneous effects of uncertainty on macroeconomic variables. The alternative with uncertainty ordered last forces the uncertainty shocks to have no impact effects on the macroeconomic indicators. Caggiano et al. (2014) show that despite the similarity between the impulse responses based on the two different identifications, one leads to significant impact effects of uncertainty shocks on unemployment while the other does not. In their framework, there is no over-identifying information to differentiate between these restrictions. However, our framework is capable of providing over-identifying information through changes in variances so as to formally test these two types of identifying assumptions. According to the likelihood ratio test, the hypothesis that the macroeconomic variables can not immediately respond to uncertainty shocks is strongly rejected.

The remainder of the paper is organized as follows. Section 2 sets up the SVAR model with time varying transition probabilities and discusses how it can be estimated and used for identification purposes. The empirical example analyzing the relation between economic policy uncertainty and U.S. unemployment is discussed in Section 3. The last section summarizes the conclusions from our study.

# 2 The regime switching model with time varying transition probabilities

## 2.1 The model setup

Consider the standard VAR model of order $p$:

$$y_t = v + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t. \tag{1}$$

where $y_t$ is the $K \times 1$ vector of variables of interest, $v$ is the $K \times 1$ intercept terms, $A_i s$ are the $K \times K$ coefficient matrices, and $u_t$ is the vector of reduced form residuals which has zero mean and covariance matrix $\Sigma_u$. In order to obtain economically meaningful structural residuals $\varepsilon_t$ with zero mean and identity covariance matrix, a linear transformation is commonly used: $u_t = B\varepsilon_t$ or $Au_t = \varepsilon_t$. In the conventional case the identifying restrictions are usually imposed on the matrix $B$ or on its inverse $A = B^{-1}$.

Now let the distribution of $u_t$ depend on a Markov process $s_t$ with $M$ discrete states, $s_t \in 1 \ldots M$. The transition probabilities are usually assumed to be constant over time: $p_{ij} = Pr(s_t = j|s_{t-1} = i)$. However here we allow them to be time varying. In particular, we follow Diebold et al. (1994) and let the transition probabilities to depend on a vector of economic fundamentals $x_t$ and assume that they evolve according to a logistic function. In a simple two-regime case the matrix of transition probabilities $P_t$ is:

$$P_t = \begin{pmatrix} p_t^{11} = \frac{e^{x'_{t-1}\beta_{11}}}{1+e^{x'_{t-1}\beta_{11}}} & p_t^{21} = 1 - p_t^{22} \\ p_t^{12} = 1 - p_t^{11} & p_t^{22} = \frac{e^{x'_{t-1}\beta_{22}}}{1+e^{x'_{t-1}\beta_{22}}} \end{pmatrix}.$$

The superscripts in $p_t^{ij}$ indicate that switch from regime $i$ to regime $j$ takes place and $\beta_{ij}$ is a vector of parameters to be estimated. For the case of three regimes, the transition probability matrix is:

$$\begin{pmatrix} p_t^{11} = \frac{e^{x'_{t-1}\beta_{11}}}{1+e^{x'_{t-1}\beta_{11}}+e^{x'_{t-1}\beta_{12}}} & p_t^{21} = \frac{e^{x'_{t-1}\beta_{21}}}{1+e^{x'_{t-1}\beta_{21}}+e^{x'_{t-1}\beta_{22}}} & p_t^{31} = 1 - p_t^{32} - p_t^{33} \\ p_t^{12} = \frac{e^{x'_{t-1}\beta_{12}}}{1+e^{x'_{t-1}\beta_{11}}+e^{x'_{t-1}\beta_{12}}} & p_t^{22} = \frac{e^{x'_{t-1}\beta_{22}}}{1+e^{x'_{t-1}\beta_{21}}+e^{x'_{t-1}\beta_{22}}} & p_t^{32} = \frac{e^{x'_{t-1}\beta_{32}}}{1+e^{x'_{t-1}\beta_{32}}+e^{x'_{t-1}\beta_{33}}} \\ p_t^{13} = 1 - p_t^{11} - p_t^{12} & p_t^{23} = 1 - p_t^{21} - p_t^{22} & p_t^{33} = \frac{e^{x'_{t-1}\beta_{33}}}{1+e^{x'_{t-1}\beta_{32}}+e^{x'_{t-1}\beta_{33}}} \end{pmatrix}$$

Identification of structural shocks in the model can be achieved by the assumption that only the variances of the shocks change across states, while impulse responses are not affected, meaning that instantaneous effects are the same across the states. If there are just two regimes with positive definite covariance matrices $\Sigma_1$ $\Sigma_2$, it is known that there exists a matrix $B$ that satisfies $\Sigma_1 = BB'$ and $\Sigma_2 = B\Lambda_2 B'$ where $\Lambda_2$ is a diagonal matrix with positive diagonal elements $\Lambda_2 = diag(\lambda_{21}, ...\lambda_{2K})$. Lanne et al. (2010) prove that the matrix $B$ is unique up to changes in sign, given that the diagonal elements of $\Lambda_2$ are distinct and ordered in a certain way. Therefore, any restrictions set upon $B$ in a conventional VAR model become over-identifying in our framework.

For the case of more than two regimes the covariance matrices are decomposed in the same way: $\Sigma_1 = BB'$, $\Sigma_i = B\Lambda_i B'$, $i = 2, ..., M$, where $\Lambda_i$ are diagonal matrices. The condition for the B to be unique is that if one pair of diagonal elements from $\Lambda_2$ are the same, there must be another pair of distinct diagonal elements from some other $\Lambda_i$. For example, if $\lambda_{2k} = \lambda_{2l}$, then there must be a pair $\lambda_{ik}, \lambda_{il}$ so that $\lambda_{ik} \neq \lambda_{il}$ from $i = 3, ..., M$. Unfortunately there are no formal tests to see whether the pairwise inequality $\lambda_{ik} \neq \lambda_{il}$ holds in the estimated model. Testing a null hypothesis of no identification $H_0 : \lambda_{21} = \lambda_{22}$ implies that some parameters are not identified under $H_0$ and standard $\chi^2$ asymptotic properties are not valid (Lütkepohl and Netšunajev, 2017a).

## 2.2 The estimation

We use maximum likelihood estimation based on log-likelihood function derived from conditional normality: given the state, the distribution of $u_t$ is assumed to be normal: $u_t|s_t \sim N(0, \Sigma_{s_t})$. The log likelihood function is highly nonlinear so that numerical optimization techniques are required. Therefore we adopt the expectation maximization (EM) algorithm of Herwartz and Lütkepohl (2014) that advances on Diebold et al. (1994) for the actual likelihood optimization task. The iterative algorithm consists of expectation step where the estimates of the unobserved regime probabilities are obtained and the maximization step where the transition parameters, structural parameters and VAR parameters are estimated.

The expectation step of the algorithm follows closely Kim (1994), Krolzig (1997) and Herwartz and Lütkepohl (2014). In the smoothing part of the expectation step we introduce the filter of Kim (1994) that is not part of the algorithm of Diebold et al. (1994). By doing this we economise on the iterations needed to compute the smoothed regime probabilities that incorporate the information from the full sample.

In the maximization step the transition parameters, the structural parameters and the VAR parameters are estimated. We add an additional step to the maximization part of the algorithm of Herwartz and Lütkepohl (2014) to estimate the transition parameters $\beta_{ij}$. As the first order conditions of the likelihood function are nonlinear in $\beta_{ij}$, we use linear approximation of $p_t^{ij}$ around $\beta_{ij}^{n-1}$ that comes from from the previous iteration. Consider $\beta_{11}$ as an example:

$$p_t^{11}(\beta_{11}^n) \approx p_t^{11}(\beta_{11}^{n-1}) + \frac{\partial p_t^{11}(\beta_{11})}{\partial \beta_{11}}\bigg|_{\beta_{11}=\beta_{11}^{n-1}} (\beta_{11} - \beta_{11}^{n-1}).$$

When one further substitutes the linear approximations for the probabilities into the first order conditions, the conditions become linear and may be rearranged to obtain the closed form solution for $\beta_{ij}$.

Even though we obtain closed form solutions to estimate the transition probabilities, the structural parameters $B$ and $\Lambda_m, m = 2, ..., M$ still have to be estimated by numerical methods. The objective function is nonlinear and can have several local optima, hence we run estimation over various initial values. With those estimates in hand the VAR parameters of the model are obtained by generalised least squares as in Herwartz and Lütkepohl (2014). The detailed procedure of our algorithm is given in the Appendix.

The classical residual based bootstrapping method is problematic for generating confidence intervals of impulse responses derived from our model because of the difficulties in the optimization of the nonlinear likelihood function. Therefore, we obtain the confidence intervals through a fixed design wild bootstrap following Gonçalves and Kilian (2004). In that procedure the bootstrap samples are constructed conditionally on the maximum likelihood estimates as:

$$y_t^* = \hat{v} + \hat{A}_1 y_{t-1} + \cdots + \hat{A}_p y_{t-p} + u_t^*,$$

where $u_t^* = \eta_t \hat{u}_t$ and $\eta_t$ is a Rademacher distributed random variable that takes value $-1$ and $1$ with probability $0.5$. We bootstrap parameter estimates conditionally on the initially estimated transition probabilities and transition parameters to preserve the pattern of volatility. Computing the bootstrap impulse responses in such a way requires nonlinear optimization of the likelihood function and, hence, is computationally demanding. We use the ML estimates as starting values in the bootstrap cycle.

At this point there are no results that confirm the reliability of this procedure in the framework of the models set up in Herwartz and Lütkepohl (2014) or Lütkepohl and Netšunajev (2017b). In a similar setup Brüggemann, Jentsch and Trenkler (2016) show that the wild bootstrap does not properly capture the higher-order moments of the distribution of interest and, hence, may not produce reliable confidence intervals for the functions of the parameters. They propose a moving block bootstrap with more appealing theoretical properties. Unfortunately, it does not preserve the volatility pattern and in a simulation they find this method to be very unreliable in small samples.

# 3 Macroeconomic impact of uncertainty shocks

## 3.1 The data

We apply our method on studying the effects of uncertainty shocks on unemployment dynamics, following closely Caggiano et al. (2014). Since the seminal paper by Bloom (2009), a growing number of research papers have studied the impact of uncertainty shocks on macroeconomic variables. One strand of the literature has studied the role of uncertainty shocks in dynamic stochastic general equilibrium models. Another strand employed VAR models for identifying uncertainty shocks and study their effects. This paper contribute to the second strand of literature by extracting information from heteroskedasticity for identification of uncertainty shocks.

There exist various uncertainty measures. Most widely used are the CBOE's Volatility Index (VIX) index, which is an index of 30-day option-implied volatility in the S&P 500 stock index, and the economic policy uncertainty index (EPU) which is based on newspaper coverage frequency. Baker et al. (2016) show that the two measures move closely together. However, the EPU index shows stronger responses to the political events such as an election of a new president, the September 11 attack, the political debates over taxes and government spending, while the VIX has a stronger connection to the financial market events such as the Asian financial crisis. Besides, the VIX covers only publicly traded firms, which accounts for around a third of private employment (see Davis, Haltiwanger, Jarmin and Miranda (2007)). But the EPU index not only reflects the stock market volatility but also major political events that affects employment on a national level. Even though Caggiano et al. (2014) use the VIX index in their paper we stick to the EPU measure. Given data availability and the arguments brought above we believe that the EPU index is very suitable as the proxy for uncertainty for our purposes.

The VAR contains the following vector of variables $y_t = (EPU_t, \pi_t, U_t, FFR_t)'$. The variables are defined in the following way:

- $EPU_t$ stands for the economic policy uncertainty index developed by Baker et al. (2016), a proxy for uncertainty;

- the inflation rate $\pi_t$ is calculated as the quarter-on-quarter percentage growth rate of the implicit GDP deflator;

- $U_t$ stands for the civilian unemployment rate.
- $FFR_t$ is the federal funds rate.

Quarterly observations of monthly data are constructed via quarterly averaging. The sample covers 1962Q3 to 2012Q3 as in Caggiano et al. (2014). The source of the EPU index is the webpage `http://www.policyuncertainty.com/` while the other time series are obtained from the FRED database offered by the Federal Reserve Bank of St. Louis.

The uncertainty shocks in the work by Caggiano et al. (2014) are identified through the Cholesky decomposition, which is a common approach in the empirical literature. Their main results are based on the ordering with EPU first: $y_t = [EPU_t, \pi_t, U_t, FFR_t]'$, which assumes that no macroeconomic shock can affect the uncertainty on impact. Alternative Cholesky ordering $y_t = [\pi_t, U_t, FFR_t, EPU_t]'$ is also considered in the paper. In the alternative ordering, the uncertainty shock is ordered last implying that all the macroeconomic variables are forced to have zero impact reaction to uncertainty shocks. These two different ways of ordering are based on assumptions, and thus are not judgment-free.

Our model has the advantage of not relying on these assumptions, instead we may use the information on the changes in volatility to achieve identification of uncertainty shocks. If our system is identified through changes in volatility, then any of above mentioned restrictions become over-identifying and testable. Our method can also be used to test whether any of these two Cholesky orderings is supported by data.

## 3.2 Model comparison

The summary statistics for the estimated models are shown in Table 1. We report the log likelihood and Akaike information criterion (AIC) for various models including a linear VAR model, the Markov switching model with constant transition probabilities, the Markov switching model with time-varying transition probabilities, and the ones restricted with two types of Cholesky ordering as in Caggiano et al. (2014). The lag order is selected to be three for all models based on AIC of the linear VAR.

The transition variable plays a key role in our analysis. Therefore we estimate and compare models with various transition variables, such as lags of inflation, lags of unemployment, lags of the federal funds rate, and lags of GDP growth rates. We also consider the moving average involving seven realizations of quarter-on-quarter GDP growth rates, following Auerbach and Gorodnichenko (2012), Bachmann and Sims (2012), Berger and Vavra (2014) and Caggiano et al. (2014). Out of all the above mentioned transition variables, the moving average of GDP growth rates, denoted as $\overline{\Delta GDP}_{t-1}$, perform the best according to the AIC and the log likelihood. We focus on the models with this transition variable in what follows.

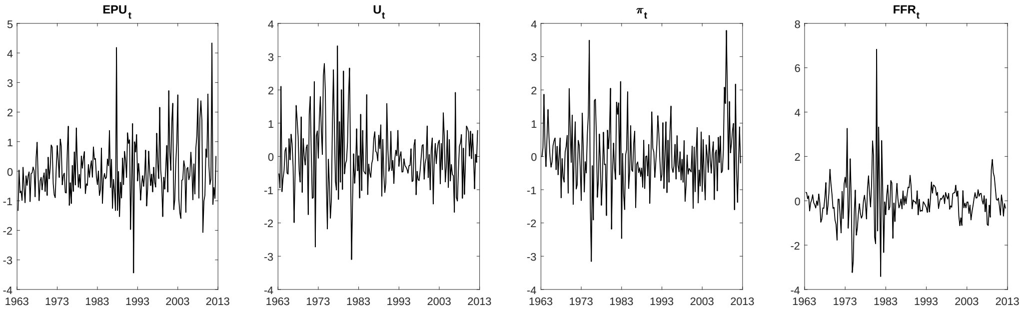
Results shown in Table 1 show that the models with time varying transition probabilities outperform the standard MS models according to the information criteria and the maximum log-likelihood. Following Diebold et al. (1994), we also employ the likelihood ratio (LR) test on whether the constant transition probabilities can be rejected or not. For the 2-regime specification, the null hypothesis of the constant transition probabilities is rejected with a $p$-value $3.98\%$. For the 3-regime case, we could also reject the null with a $p$-value $0.72\%$. These LR test results suggest that the assumption of constant transition probabilities is too restrictive. It is advantageous to allow the transition probabilities to vary over time.

Table 1: Comparison of VAR(3) Models for $y_t = (EPU_t, \pi_t, U_t, FFR_t)'$, sample period: 1962Q3 - 2012Q3

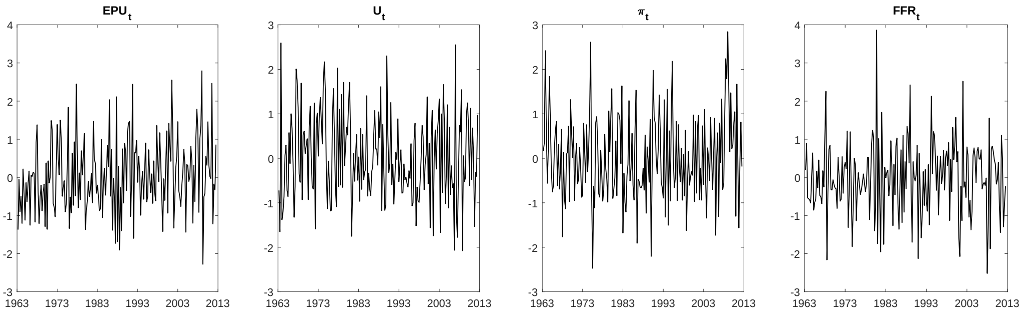
| Model | $x_{t-1}$ | $\log L_T$ | AIC |
|---|---|---|---|
| VAR(3), linear | | -1408.72 | 2921.44 |
| MS-SVAR with constant $P$ | | | |
| MS(2) | | -1273.59 | 2695.18 |
| MS(3) | | -1247.74 | 2659.49 |
| MS-SVAR with time varying $P_t$ | | | |
| MS(2), | $(1; \overline{\Delta GDP}_{t-1})$ | -1268.57 | 2689.14 |
| MS(3), | $(1; \overline{\Delta GDP}_{t-1})$ | -1234.14 | 2656.29 |
| MS(3), state invariant $B$ | $(1; \overline{\Delta GDP}_{t-1})$ | -1236.67 | 2649.34 |
| MS-SVAR with time varying $P_t$ and Cholesky restrictions | | | |
| MS(3), $y_t = (EPU_t, \pi_t, U_t, FFR_t)'$ | | -1239.91 | **2643.82** |
| MS(3), $y_t = (\pi_t, U_t, FFR_t, EPU_t)'$ | | -1244.97 | 2653.93 |

Note: $L_T$ – likelihood function, AIC $= -2\log L_T + 2\times$no of free parameters, SC $= -2\log L_T + \log T\times$no of free parameters. $\overline{\Delta GDP}_{t-1}$stands for the seven quarter moving average of GDP growth rates.

Moreover, choosing the number of regimes is also critical for our analysis. Following Psaradakis and Spagnolo (2006), as well as Herwartz and Lütkepohl (2014), we use the information criteria as the tool to select the number of regimes. If judged upon this criteria, the model without any regime switching perform the worst, while the model with time varying transition probabilities with three regimes perform the best. The standardised residuals of the VAR(3) model and of the MS(3) model with time varying $P_t$ are shown in Figure 1. The residuals of the model that takes into account changing volatility are much more regular than the ones of the standard model. Estimated parameters of the transition function




(a) Residual of the VAR(3) model




(b) Residuals of the MS(3) model with time varying transition probabilities

Figure 1: Residuals of various models

for the model are reported in Table 2. They are estimated rather precisely with the standard errors being mostly smaller than the estimates. Admittedly this is not the case for the the parameters describing changes from regime two.

Figure 2 compares the estimated smoothed regime probabilities from two different models. The subfigure on the top displays the regime probabilities estimated from Model 1 that let the transition probabilities depend on the moving averages of GDP growth rates. The first regime, which is the most volatile one, covers some period in the beginning of the 1970s, the beginning of the 1980s, as well as the recent financial crisis. The second regime includes most of the first half of 2000s, and the recovery period following the financial crisis. The third regime, which represent the least volatile regime, covers almost the whole period between 1984 till 2008 with a few exceptions. The timing of this low volatility regimes corresponds to the well known phenomenon named the Great Moderation. The graph on the bottom shows the estimated regime probabilities from Model 2 that assumes constant transition probabilities.

There are noticeable differences when the assumption of constant transition probabilities is imposed. For example, the period of the 2008 financial crisis,

Table 2: Estimated transition parameters and their standard errors

| | $\beta_{11}$ | $\beta_{12}$ | $\beta_{21}$ | $\beta_{22}$ | $\beta_{32}$ | $\beta_{33}$ |
|---|---|---|---|---|---|---|
| Estimate(intercept) | 5.40 | -1.95 | 61.22 | -69.10 | 19.60 | -10.92 |
| Estimate(slope) | 4.65 | -3.85 | 63.67 | -67.85 | 14.65 | -6.12 |
| Std.err.(intercept) | 3.92 | 1.68 | 305.87 | 321.31 | 9.76 | 5.58 |
| Std.err.(slope) | 5.77 | 6.78 | 307.39 | 323.70 | 7.10 | 3.17 |

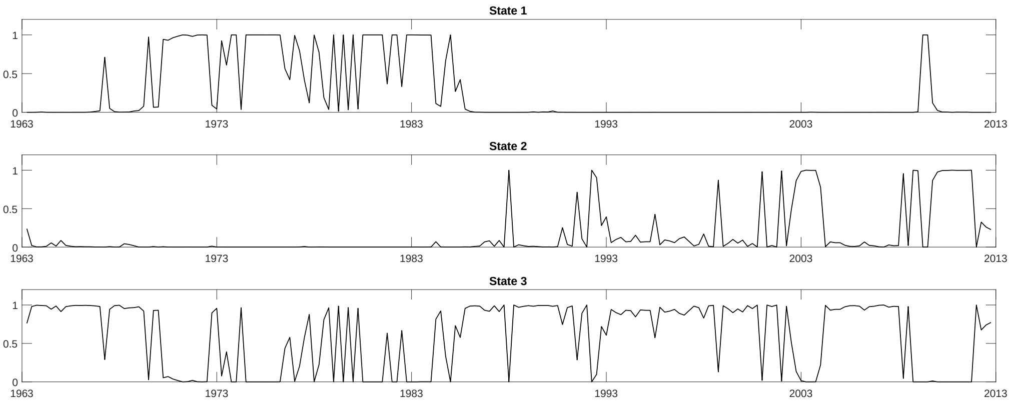
is estimated to be in the most volatile state according to Model 1. In contrast, according to Model 2, the 2008 financial crisis is estimated to be in a relatively calm state. Otherwise, the whole period from the end of 1960s to the beginning of the 1980s are estimated to be constantly in the most volatile regime according to Model 2. However, according to Model 1, certain intervals out of this period are also relatively calm. These results are in line with the economic narratives and with the simulation study of a univariate Markov switching process by Diebold et al. (1994), which demonstrate that the estimated regime probabilities of the model with time varying transition probabilities capture the state better than the model with constant transition probabilities.

The estimated time varying transition probabilities in Figure 3 provide more evidence for the importance of relaxing the constant transition-probability assumption. Take for example the top-right subfigure, this figure displays the probability of switching from the most calm regime to the most volatile regime, i.e., $p_t^{31}$. From 1973, this probability starts to rise above 50% to almost 100%, falls shortly back to almost zero around 1975, and then rise again and remain high till the beginning of 1980s. Take again for example $p_t^{21}$, which represents the probability of switching from the medium regime to the most volatile regime. This probability $p_t^{21}$ takes value of zero for most of the sample, however, towards the end of sample for the period around the 2008 financial crisis, this probability rises to almost one, which suggests that it is highly likely for the underlying state to switch from the medium regime to the most volatile regime. If the constant transition probabilities are assumed, these probabilities would look like a straight horizontal line, and the leading information contained in the moving averages of the GDP growth rates for the transition probabilities would have been lost.

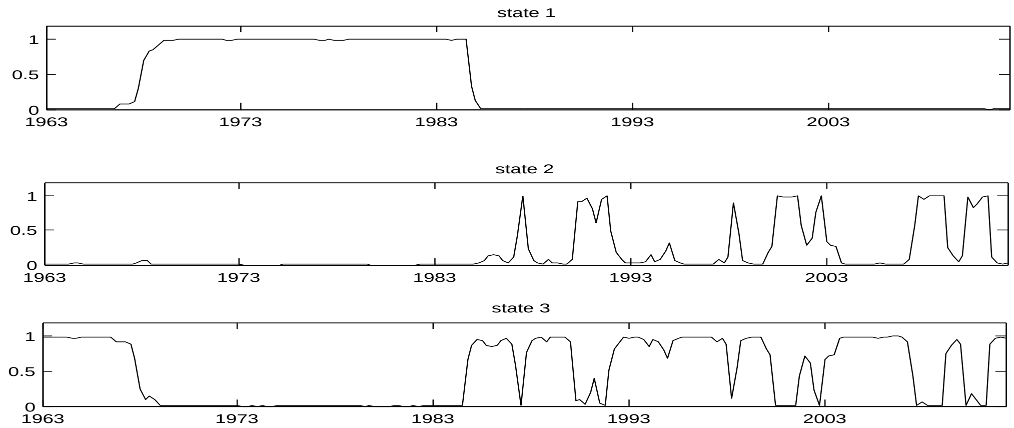
## 3.3 Identification analysis

We next proceed with the analysis of structural shocks identified using the model proposed. It is important to check whether the estimated model is identified at least by comparing the pairs of the relatives variances. The estimates of these

Figure 2: A comparison of estimated smoothed regime probabilities




(a) Model 1: time-varying transition probabilities




(b) Model 2: constant transition probabilities

Note: This graph compares the estimated regime probabilities from two different models. Model 1 is the three regime MS-VAR model in which transition probabilities are governed by the moving averages of seven quarter-on quarter GDP growth rates. Model 2 is the three regime MS-VAR model assuming constant transition probabilities. State 1 stands for the state with highest volatility. State 2 is the state with medium volatility and State 3 is the one with lowest volatility.

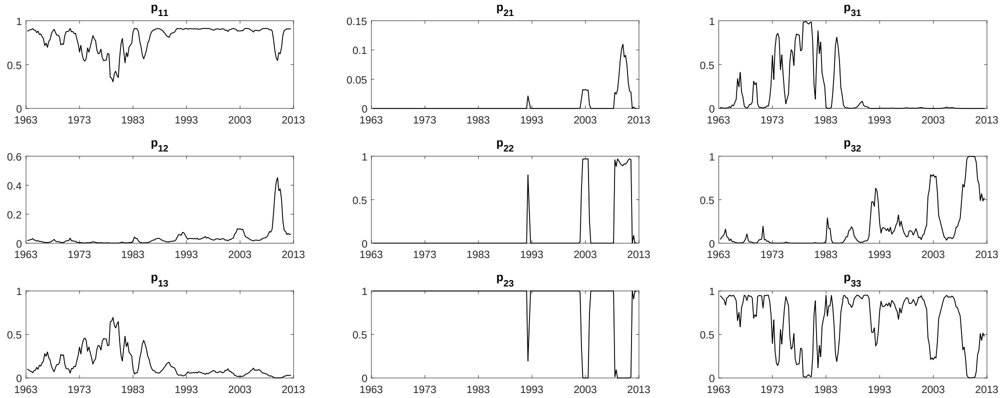


Figure 3: Time varying transition probabilities

Table 3: Estimated relative variances of the MS(3) model with state invariant $B$

| Parameter | Estimate | Standard error |
|---|---|---|
| $\lambda_{21}$ | 0.07 | 0.06 |
| $\lambda_{22}$ | 0.41 | 0.27 |
| $\lambda_{23}$ | 8.96 | 5.99 |
| $\lambda_{24}$ | 0.03 | 0.02 |
| $\lambda_{31}$ | 0.20 | 0.08 |
| $\lambda_{32}$ | 0.14 | 0.05 |
| $\lambda_{33}$ | 1.04 | 0.55 |
| $\lambda_{34}$ | 0.03 | $8.60 \times 10^{-3}$ |

parameters along with their standard errors of our preferred model are shown in Table 3. In the situation where no formal tests for identification exist the standard errors of the variances have to be examined (Lütkepohl and Netšunajev, 2017a). In the case of the preferred three state model with time varying transition probabilities, the estimates of the $\Lambda_2$ and $\Lambda_3$ matrices are quite precise and heterogeneous with standard errors being much lower than corresponding point estimates. Thus there are good reasons to believe that the we have additional information in volatilities and the structural matrix $B$ is well identified and the tests for restrictions have power.

The two types of Cholesky orderings as in Caggiano et al. (2014) are to be tested next. The first ordering, $y_t = (EPU_t, \pi_t, U_t, FFR_t)'$, which assumes that the macroeconomic variables have no contemporaneous effects on uncertainty, fits the economic intuition well. However, even though the second ordering counterintuitively assumes that uncertainty shocks can not have impact effects on macroeconomic indicators, it is shown by Caggiano et al. (2014) that the responses of the unemployment to an uncertainty shocks under the second ordering look very similar to those under the first ordering. Now that we have additional information from changes in volatility, we could test whether any of these two Cholesky orderings can be confirmed by the data.

Results of testing these two different restrictions are show in Table 4. Under the assumption of 2 regimes, both Cholesky orderings can not be rejected. However, this is not the case for the preferred three regime specification. For the three regime model we first test the structural covariance matrix decomposition where $H_0$: $\Sigma_1 = BB'$, $\Sigma_2 = B\Lambda_2 B'$, $\Sigma_3 = B\Lambda_3 B'$ and under alternative the covariance matrices are fully unrestricted. We do not reject the decomposition using the LR test with $df = 6$, value of $5.06$ and $p = 0.53$. Thus we proceed with testing the Chelesky orderings of interest. The identification scheme that assumes that the macroeconomic shocks have no contemporaneous effects on uncertainty can not be rejected with the $p = 0.38$. On the contrary, the Cholesky ordering that assumes counterintuitive effects of uncertainty shocks is rejected with the $p$-value of $0.01$. These results indicate that the model with only two regimes may lack information for identification. The model with three regimes exploits the pattern of volatility better and thus the estimated variances of structural shocks are much more distinct making the LR test more powerful.

Table 4: Tests for different restrictions for identifying uncertainty shocks

| $H_0$ | $H_1$ | df | LR statistic | $p$-value |
|---|---|---|---|---|
| $B_1$ | MS(2), state invariant $B$ | 6 | 6.34 | 0.39 |
| $B_2$ | MS(2), state invariant $B$ | 6 | 10.02 | 0.12 |
| $B_1$ | MS(3), state invariant $B$ | 6 | 6.48 | 0.38 |
| $B_2$ | MS(3), state invariant $B$ | 6 | 16.60 | 0.01 |

Note: The null hypothesis $B_1$ stands for the six zero restrictions on the upper triangular of $B$ matrix in the spirit of Cholesky decomposition with the variables ordered as $(EPU_t, \pi_t, U_t, FFR_t)'$. The null hypothesis $B_2$ represents the six zero restrictions on $B$ when the EPU index is ordered as the last variable in the VAR system, i.e., $(\pi_t, U_t, FFR_t, EPU_t)'$. The $MS(2)$, state invariant $B$ models under the alternative hypothesis are the ones with regime-switching variances and time-varying probabilities under 2 regimes. The $MS(3)$, state invariant $B$ models are the ones under 3 regimes. The models under $H_1$ impose no restrictions on the $B$.

Can uncertainty shock have contemporaneous effects on macroeconomic variables? Caggiano et al. (2014) are ambiguous about the validity of the two ways of identification of uncertainty shocks, and believe that they are robust to each other.

However, our model exploits information from changes in variances of shocks and allows to reject strongly the Cholesky ordering $(y_t = \pi_t, U_t, FFR_t, EPU_t)'$. In other words, the hypothesis that macroeconomic variables can not respond immediately to uncertainty shocks is rejected.

# 4 Conclusions

In the paper we propose a structural vector autoregressive model where the changes in volatility are governed by a Markov process with time varying transition probabilities. Time varying transition probabilities are assumed to depend on fundamental economic variables. The structural parameters of the model are identified with changes in volatility of shocks. Additional information that comes from the time variation in the variances of structural shocks allows to test conventional identifying restrictions. We estimate the model using maximum likelihood and a flexible EM algorithm.

In the empirical illustration of our model, we apply this method to identify uncertainty shocks and evaluate their effects on macroeconomic variables following the study by Caggiano et al. (2014). Compared with a standard Markov switching models as in Lanne et al. (2010), our model with time varying transition probabilities fits the example data better than ones assuming constant transition probabilities according to the information criteria and the maximum log likelihood. This is most likely due the leading information contained in the transition variable, in our case, the moving averages of seven quarter-on-quarter GDP growth rates.

With extra information extracted from changes in variances we test two different types of Cholesky orderings used in Caggiano et al. (2014). Using the preferred three regime MS-VAR model with moving averages of GDP growth rates as the transition variable governing the transition probabilities, we strongly reject the Cholesky ordering restricting that uncertainty shocks not to have contemporaneous effects on macroeconomic variables. Interestingly, we do not reject the Cholesky ordering that allows for the impact effects of uncertainty shocks on the macroeconomic variables. This shows the power of our method to differentiate between economic assumptions that are used for identification purposes.

# Acknowledgments

We thank participants in the 2017 Nordic Econometric Meeting, and participants in the Research Seminar at Free University Economics Department for their helpful comments.

# Appendix. Estimation of the MS-SVAR model with time-varying transition probabilities

The section describes in detail the expectation maximization (EM) algorithm based on Krolzig (1997), Herwartz and Lütkepohl (2014) and Diebold et al. (1994), and presents the estimation procedure for structural VAR model with changes in volatility of shocks where the transition probability matrix is also allowed to vary over time.

## Definitions

The baseline model is the VAR(p) of the form:

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t.$$

Denote:
$M$ - number of states, assumed to be three in this appendix,
$K$ - number of variables in the vector $y$,
$p$ - number of lags.

Let the matrix $X = [x_0, x_1, ..., x_T]$ contain transition variables up to $T$ with the entries for specific $t$ given by a $(J+1) \times 1$ vector $x_t$ of $J$ economic variables that affect the transition probabilities and a leading one for a constant.

Define $\xi_t = \begin{bmatrix} I(s_t = 1) \\ \vdots \\ I(s_t = M) \end{bmatrix}$, then $E(\xi_t) = \begin{bmatrix} Pr(s_t = 1) \\ \vdots \\ Pr(s_t = M) \end{bmatrix}$, where $I()$ is an indicator function which takes value 1 if statement in the argument is true and 0 otherwise.

Further define

$\xi_{t|s} = E(\xi_t | Y_s, X_s) = \begin{bmatrix} Pr(s_t = 1 | Y_s, X_s) \\ \vdots \\ Pr(s_t = M | Y_s, X_s) \end{bmatrix}$, where $Y_s = (y_1, ..., y_s), X_s = (x_0, ..., x_s)$

Next:

$$\xi_{t|T}^{(2)} = \begin{bmatrix} Pr(s_t = 1|s_{t-1} = 1, Y_T, X) \\ \vdots \\ Pr(s_t = 1|s_{t-1} = M, Y_T, X) \\ \vdots \\ Pr(s_t = M|s_{t-1} = 1, Y_T, X) \\ \vdots \\ Pr(s_t = M|s_{t-1} = M, Y_T, X) \end{bmatrix}.$$

We let the transition probability matrix to be time varying for $M$ state Markov process. Define $P_t$ as the time varying transition matrix, which yields $\xi_{t+1|t} = P_t \xi_{t|t}$, for $t = 0, 1, ..., T-1$. Advancing on Diebold et al. (1994) we shows the closed form solutions for estimating models with $M = 2$ and $M = 3$ as these appear to be the most important in practice. Expressions for models with $M > 3$ may be derived analogously. The individual elements of the $P_t$ matrix evolve as logistic functions of $x'_{t-1}\beta_{ij}$. Then $\beta_{ij}$ is the $(J+1) \times 1$ vector of parameters. It is convenient to collect the individual $\beta_{ij}$ vectors into a matrix $\beta = [\beta_{11}\ \beta_{22}]$ for 2 regimes and $\beta = \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{32} \\ \beta_{12} & \beta_{22} & \beta_{33} \end{bmatrix}$ for 3 regimes. The matrix $\beta_0$ denotes the initial values for the transition parameters. The matrix $P_t$ is defined as:

$$P_t = \begin{pmatrix} Pr(s_{t+1} = 1|s_t = 1, \beta, x_{t-1}) & \dots & Pr(s_{t+1} = 1|s_t = M, \beta, x_{t-1}) \\ \vdots & \ddots & \vdots \\ Pr(s_{t+1} = M|s_t = 1, \beta, x_{t-1}) & \dots & Pr(s_{t+1} = M|s_t = M, \beta, x_{t-1}) \end{pmatrix}$$

The following matrices illustrate the details. Note that the subscripts for $\beta_{ij}$ and superscripts for $p_t^{ij}$ denote transition from state $i$ to state $j$. Transition probability matrix for $M = 2$:

$$\begin{pmatrix} p_t^{11} = \frac{e^{x'_{t-1}\beta_{11}}}{1+e^{x'_{t-1}\beta_{11}}} & p_t^{21} = 1 - p_t^{22} \\ p_t^{12} = 1 - p_t^{11} & p_t^{22} = \frac{e^{x'_{t-1}\beta_{22}}}{1+e^{x'_{t-1}\beta_{22}}} \end{pmatrix}$$

Transition probability matrix for $M = 3$:

$$\begin{pmatrix} p_t^{11} = \frac{e^{x'_{t-1}\beta_{11}}}{1+e^{x'_{t-1}\beta_{11}}+e^{x'_{t-1}\beta_{12}}} & p_t^{21} = \frac{e^{x'_{t-1}\beta_{21}}}{1+e^{x'_{t-1}\beta_{21}}+e^{x'_{t-1}\beta_{22}}} & p_t^{31} = 1 - p_t^{32} - p_t^{33} \\ p_t^{12} = \frac{e^{x'_{t-1}\beta_{21}}}{1+e^{x'_{t-1}\beta_{11}}+e^{x'_{t-1}\beta_{12}}} & p_t^{22} = \frac{e^{x'_{t-1}\beta_{22}}}{1+e^{x'_{t-1}\beta_{21}}+e^{x'_{t-1}\beta_{22}}} & p_t^{32} = \frac{e^{x'_{t-1}\beta_{32}}}{1+e^{x'_{t-1}\beta_{32}}+e^{x'_{t-1}\beta_{33}}} \\ p_t^{13} = 1 - p_t^{11} - p_t^{12} & p_t^{23} = 1 - p_t^{21} - p_t^{22} & p_t^{33} = \frac{e^{x'_{t-1}\beta_{33}}}{1+e^{x'_{t-1}\beta_{32}}+e^{x'_{t-1}\beta_{33}}} \end{pmatrix}$$

Next define $\eta_t = \begin{bmatrix} f(y_t|s_t = 1, Y_{t-1}, X_{t-1}) \\ \vdots \\ f(y_t|s_t = M, Y_{t-1}, X_{t-1}) \end{bmatrix}$,

where $f()$ is conditional distribution function:

$$f(y_t|s_t = m, Y_{t-1}, X_{t-1}) = (2\pi)^{-K/2} \det(\Sigma_m)^{-1/2} \exp(-0.5u_t'\Sigma_m^{-1}u_t).$$

Covariance matrices have decomposition as previously described: $\Sigma_1 = BB', \Sigma_m = B\Lambda_m B'$ for $m = 2, ..., M$

Further the following notation is used:
$\odot$ elementwise multiplication,
$\oslash$ elementwise devision,
$\otimes$ Kronecker product,
$I_K$ is a $K \times K$ dimensional identity matrix,
$1_M = (1, ..., 1)'$ is a $M \times 1$ dimensional vector of ones,
$\theta = \text{vec}(v, A_1, A_2, ..., A_P)$ is the vector of VAR coefficients
$Z_{t-1}' = (1, y_{t-1}', y_{t-2}', ..., y_{t-p}')$ is the matrix of ones and lagged observations.

# Initial values

The following starting values are used for the iterations:
$P_t$ is calculated for given $X$ and $\beta_0$ for $1, ..., T$.
$\hat{\theta} = \text{vec}(\hat{v}, \hat{A}_1, ..., \hat{A}_p) = \left[\sum_{t=1}^{T} Z_{t-1}Z_{t-1}' \otimes I_K\right]^{-1} \sum_{t=1}^{T} (Z_{t-1} \otimes I_K)y_t$
$B = T^{-1}(\sum_{t=1}^{T} \hat{u}_t\hat{u}_t')^{1/2} + B_0$ , where $\hat{u}_t = y_t - (Z_{t-1}' \otimes I_K)\hat{\theta}$
and $B_0$ is a matrix of random numbers coming form standard normal distribution and scaled by a factor of $10^{-5}$.

$\Lambda_1 = I_K, \Lambda_m = c_m I_K, m = 2, ..., M$

$\xi_{0|0} = M^{-1}1_M$

# Expectation step

For given $P_t, \theta, \Sigma_m, m = 1, 2, ..., M$ and $\xi_0 = \xi_{0|0}$ the following parameters are computed:

$\eta_t$ for $t = 1, 2, ..., T$,

$\xi_{t|t} = \frac{\eta_t \odot P_t \xi_{t-1|t-1}}{1'_M(\eta_t \odot P_t \xi_{t-1|t-1})}$, for $t = 1, 2, ..., T$.

$\xi_{t|T} = (P'_t(\xi_{t+1|T} \oslash P_t \xi_{t|t})) \odot \xi_{t|t}$, for $t = T-1, ..., 0$.

$\xi^{(2)}_{t|T} = \text{vec}(P'_t) \odot ((\xi_{t+1|T} \oslash P_t \xi_{t|t}) \otimes \xi_{t|t})$, for $t = 1, ..., T-1$.

# Maximization step

## Estimation of transition parameters $\beta$

Given the smoothed probabilities, the expected complete-data log likelihood are non-linear in the $\beta$ transition parameters. Taking into account the logistic transition function, the first order conditions for $\beta$ are given as follows:

M = 2:
$\sum_{t=2}^{T} x_{t-1} \{Pr(s_t = 1|s_{t-1} = 1, Y_T, X) - p_t^{11} Pr(s_{t-1} = 1|Y_T, X)\} = 0,$
$\sum_{t=2}^{T} x_{t-1} \{Pr(s_t = 2|s_{t-1} = 2, Y_T, X) - p_t^{22} Pr(s_{t-1} = 2|Y_T, X)\} = 0,$

M = 3:
$\sum_{t=2}^{T} x_{t-1} \{Pr(s_t = 1|s_{t-1} = 1, Y_T, X) - p_t^{11} Pr(s_{t-1} = 1|Y_T, X)\} = 0,$
$\sum_{t=2}^{T} x_{t-1} \{Pr(s_t = 2|s_{t-1} = 1, Y_T, X) - p_t^{12} Pr(s_{t-1} = 1|Y_T, X)\} = 0,$
$\sum_{t=2}^{T} x_{t-1} \{Pr(s_t = 1|s_{t-1} = 2, Y_T, X) - p_t^{21} Pr(s_{t-1} = 2|Y_T, X)\} = 0,$
$\sum_{t=2}^{T} x_{t-1} \{Pr(s_t = 2|s_{t-1} = 2, Y_T, X) - p_t^{22} Pr(s_{t-1} = 2|Y_T, X)\} = 0,$
$\sum_{t=2}^{T} x_{t-1} \{Pr(s_t = 3|s_{t-1} = 2, Y_T, X) - p_t^{32} Pr(s_{t-1} = 3|Y_T, X)\} = 0,$
$\sum_{t=2}^{T} x_{t-1} \{Pr(s_t = 3|s_{t-1} = 3, Y_T, X) - p_t^{33} Pr(s_{t-1} = 3|Y_T, X)\} = 0.$

Using the following Taylor approximation of the elements of $P_t$ matrix, we find the closed-form solution for all $\beta$ vectors. Consider $\beta_{11}$ as an example:
$p_t^{11}(\beta_{11}^{n}) \approx p_t^{11}(\beta_{11}^{n-1}) + \frac{\partial p_t^{11}(\beta_{11})}{\partial \beta_{11}}\Big|_{\beta_{11}=\beta_{11}^{n-1}} (\beta_{11} - \beta_{11}^{n-1})$

where $\beta_{11}^{n-1}$ is the $\beta_{11}$ coming from the previous iteration of the algorithm. The closed-form solutions for $\beta$ are given as follows.

$\underline{M=2}$:

$$\beta_{11} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=1|Y_T,X)p_{1t}^{11} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=1|Y_T,X)p_{Jt}^{11} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=1|Y_T,X)p_{1t}^{11} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=1|Y_T,X)p_{Jt}^{11} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} \left\{ Pr(s_t=1|s_{t-1}=1,Y_T,X) - Pr(s_{t-1}=1|Y_T,X)[p_t^{11} - \beta_{11}^{j-1} \frac{\partial P_t^{11}}{\partial \beta_{11}}] \right\} \\ \vdots \\ \sum_{t=2}^{T} x_{J,t-1} \left\{ Pr(s_t=1|s_{t-1}=1,Y_T,X) - Pr(s_{t-1}=1|Y_T,X)[p_t^{11} - \beta_{11}^{j-1} \frac{\partial P_t^{11}}{\partial \beta_{11}}] \right\} \end{pmatrix}$$

$$\beta_{22} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=2|Y_T,X)p_{1t}^{22} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=2|Y_T,X)p_{Jt}^{22} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=2|Y_T,X)p_{1t}^{22} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=2|Y_T,X)p_{Jt}^{22} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} \left\{ Pr(s_t=2|s_{t-1}=2,Y_T,X) - Pr(s_{t-1}=2|Y_T,X)[p_t^{22} - \beta_{22}^{j-1} \frac{\partial P_t^{22}}{\partial \beta_{22}}] \right\} \\ \vdots \\ \sum_{t=2}^{T} x_{J,t-1} \left\{ Pr(s_t=2|s_{t-1}=2,Y_T,X) - Pr(s_{t-1}=2|Y_T,X)[p_t^{22} - \beta_{22}^{j-1} \frac{\partial P_t^{22}}{\partial \beta_{22}}] \right\} \end{pmatrix}$$

$\underline{M=3}$:

$$\beta_{11} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=1|Y_T,X)p_{1t}^{11} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=1|Y_T,X)p_{Jt}^{11} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=1|Y_T,X)p_{1t}^{11} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=1|Y_T,X)p_{Jt}^{11} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} \left\{ Pr(s_t=1|s_{t-1}=1,Y_T,X) - Pr(s_{t-1}=1|Y_T,X)[p_t^{11} - \beta_{11}^{j-1} \frac{\partial P_t^{11}}{\partial \beta_{11}}] \right\} \\ \vdots \\ \sum_{t=2}^{T} x_{J,t-1} \left\{ Pr(s_t=1|s_{t-1}=1,Y_T,X) - Pr(s_{t-1}=1|Y_T,X)[p_t^{11} - \beta_{11}^{j-1} \frac{\partial P_t^{11}}{\partial \beta_{11}}] \right\} \end{pmatrix}$$

$$\beta_{12} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=1|Y_T,X)p_{1t}^{12} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=1|Y_T,X)p_{Jt}^{12} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=1|Y_T,X)p_{1t}^{12} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=1|Y_T,X)p_{Jt}^{12} \end{pmatrix}^{-1}$$

$$\times \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} \left\{ Pr(s_t=2|s_{t-1}=1,Y_T,X) - Pr(s_{t-1}=1|Y_T,X)[p_t^{12} - \beta_{12}^{j-1} \frac{\partial P_t^{12}}{\partial \beta_{12}}] \right\} \\ \vdots \\ \sum_{t=2}^{T} x_{J,t-1} \left\{ Pr(s_t=2|s_{t-1}=1,Y_T,X) - Pr(s_{t-1}=1|Y_T,X)[p_t^{12} - \beta_{12}^{j-1} \frac{\partial P_t^{12}}{\partial \beta_{12}}] \right\} \end{pmatrix}$$

$$\beta_{21} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=2|Y_T,X) p_{1t}^{21} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=2|Y_T,X) p_{Jt}^{21} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=2|Y_T,X) p_{1t}^{21} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=2|Y_T,X) p_{Jt}^{21} \end{pmatrix}^{-1}$$
$$\times \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} \left\{ Pr(s_t=1|s_{t-1}=2,Y_T,X) - Pr(s_{t-1}=2|Y_T,X)[p_t^{21} - \beta_{21}^{j-1} \frac{\partial P_t^{21}}{\partial \beta_{21}}] \right\} \\ \vdots \\ \sum_{t=2}^{T} x_{J,t-1} \left\{ Pr(s_t=1|s_{t-1}=2,Y_T,X) - Pr(s_{t-1}=2|Y_T,X)[p_t^{21} - \beta_{21}^{j-1} \frac{\partial P_t^{21}}{\partial \beta_{21}}] \right\} \end{pmatrix}$$

$$\beta_{22} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=2|Y_T,X) p_{1t}^{22} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=2|Y_T,X) p_{Jt}^{22} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=2|Y_T,X) p_{1t}^{22} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=2|Y_T,X) p_{Jt}^{22} \end{pmatrix}^{-1}$$
$$\times \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} \left\{ Pr(s_t=2|s_{t-1}=2,Y_T,X) - Pr(s_{t-1}=2|Y_T,X)[p_t^{22} - \beta_{22}^{j-1} \frac{\partial P_t^{22}}{\partial \beta_{22}}] \right\} \\ \vdots \\ \sum_{t=2}^{T} x_{J,t-1} \left\{ Pr(s_t=2|s_{t-1}=2,Y_T,X) - Pr(s_{t-1}=2|Y_T,X)[p_t^{22} - \beta_{22}^{j-1} \frac{\partial P_t^{22}}{\partial \beta_{22}}] \right\} \end{pmatrix}$$

$$\beta_{32} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=3|Y_T,X) p_{1t}^{32} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=3|Y_T,X) p_{Jt}^{32} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=3|Y_T,X) p_{1t}^{32} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=3|Y_T,X) p_{Jt}^{32} \end{pmatrix}^{-1}$$
$$\times \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} \left\{ Pr(s_t=2|s_{t-1}=3,Y_T,X) - Pr(s_{t-1}=3|Y_T,X)[p_t^{32} - \beta_{32}^{j-1} \frac{\partial P_t^{32}}{\partial \beta_{32}}] \right\} \\ \vdots \\ \sum_{t=2}^{T} x_{J,t-1} \left\{ Pr(s_t=2|s_{t-1}=3,Y_T,X) - Pr(s_{t-1}=3|Y_T,X)[p_t^{32} - \beta_{32}^{j-1} \frac{\partial P_t^{32}}{\partial \beta_{32}}] \right\} \end{pmatrix}$$

$$\beta_{33} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=3|Y_T,X) p_{1t}^{33} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=3|Y_T,X) p_{Jt}^{33} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=3|Y_T,X) p_{1t}^{33} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=3|Y_T,X) p_{Jt}^{33} \end{pmatrix}^{-1}$$
$$\times \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} \left\{ Pr(s_t=3|s_{t-1}=3,Y_T,X) - Pr(s_{t-1}=3|Y_T,X)[p_t^{33} - \beta_{33}^{j-1} \frac{\partial P_t^{33}}{\partial \beta_{33}}] \right\} \\ \vdots \\ \sum_{t=2}^{T} x_{J,t-1} \left\{ Pr(s_t=3|s_{t-1}=3,Y_T,X) - Pr(s_{t-1}=3|Y_T,X)[p_t^{33} - \beta_{33}^{j-1} \frac{\partial P_t^{33}}{\partial \beta_{33}}] \right\} \end{pmatrix}$$

where $p_{1t}^{ii}, \cdots, p_{Jt}^{ii}$ are denoting the elements in the vector of partial derivatives as used in the Taylor approximation.

### Estimation of structural parameters $B$ and $\Lambda_m$ :

Define $T_m = \sum_{t=1}^{T} \xi_{mt|T}$, where $\xi_{mt|T}$ denotes the $m$-th element of the vector $\xi_{t|T}$. Estimation of $B$ and $\Lambda_m$ is done by minimizing the likelihood function:

$$l(B, \Lambda_2, ..., , \Lambda_M) = T \log\det(B) + \frac{1}{2}\left(B'^{-1}B^{-1}\sum_{t=1}^{T}\xi_{1t|T}\hat{u}_t\hat{u}_t'\right)$$
$$+ \sum_{m=2}^{M}\left[\frac{T_m}{2}\log\det(\Lambda_M) + \frac{1}{2}tr\left(B'^{-1}\Lambda_M^{-1}B^{-1}\sum_{t=1}^{T}\xi_{mt|T}\hat{u}_t\hat{u}_t'\right)\right].$$

Then compute:
$\hat{\Sigma}_1 = \hat{B}\hat{B}', \hat{\Sigma}_m = \hat{B}\hat{\Lambda}_m\hat{B}'$ for $m = 2, ..., M$

### Estimation of VAR parameters:

Estimates of the parameter vector $\theta$ are given by:

$$\hat{\theta} = \left[\sum_{m=1}^{M}\left(\sum_{t=1}^{T}\xi_{mt|T}Z_{t-1}Z_{t-1}'\right)\otimes\hat{\Sigma}_m^{-1}\right]^{-1}\sum_{t=1}^{T}\left(\sum_{m=1}^{M}\xi_{mt|T}Z_{t-1}\otimes\hat{\Sigma}_t^{-1}\right)y_t$$

Initial regime probabilities are updated according to:
$\xi_{0|0} = \xi_{0|T}$

# Convergence Criteria

Relative change in the value of the log-likelihood function is used as convergence criteria. The log-likelihood is evaluated for given $P_t, \theta, \Sigma_m, m = 1, 2, ..., M$ and $\xi_{0|0}$ in the end of the expectation step. Given:

$\eta_t$ for $t = 1, 2, ..., T$,

$\xi_{t|t-1} = P_t\xi_{t-1|t-1}$, for $t = 1, 2, ..., T$,

$\xi_{t|t} = \frac{\eta_t \odot P_t\xi_{t|t-1}}{1_M'(\eta_t \odot P_t\xi_{t|t-1})}$, for $t = 1, 2, ..., T$.

The log likelihood is:
$\log L_T = \sum_{t=1}^{T}\log f(y_t|Y_{t-1})$,

$f(y_t|Y_{t-1}) = \xi_{t|t-1}'\eta_t$.

Estimation of $\beta, B$, $\Lambda_m$ and $\theta$ are iterated until convergence, i.e. relative change $\Delta$ in the log-likelihood is negligibly small (does not exceed tolerance value $\alpha = 10^{-9}$) for $k$-th and $(k-1)$-th rounds of iterations:

$$\Delta = \frac{\log L_T(k) - \log L_T(k-1)}{\log L_T(k-1)} < \alpha \; .$$

# Bootstrapping confidence bands for impulse responses

Herwartz and Lütkepohl (2014) discuss a fixed design wild bootstrap procedure for constructing confidence intervals for impulse responses in the structural VARs with Markov switching in variances. As long as there are no better alternatives, we adopt their idea for the model developed in the present paper. The bootstrap samples are constructed as:

$$y_t^* = \hat{v} + \hat{A}_1 y_{t-1} + \cdots + \hat{A}_p y_{t-p} + u_t^*$$

where $u_t^* = \zeta_t \hat{u}_t$ and $\zeta_t$ is a random variable taking values 1 and $-1$, each with probability 0.5. We bootstrap parameter estimates $\theta^*$, $B^*$ and $\Lambda^*$ conditionally on the initially estimated transition probabilities.

# References


Alexopoulos, M. and Cohen, J. (2009). Uncertainty times, uncertainty measures, *Technical report*, University of Toronto, Department of Economics, Working Paper No. 325.

Auerbach, A. and Gorodnichenko, Y. (2012). Measuring the output responses o fiscal policy, *American Econmic Journal: Economic Policy* **4**(2): 1–27.

Bachmann, R., Elstner, S. and Sims, E. (2013). Uncertainty and economic activity: evidence from business survey data, *American Economic Journal: Economic Policy* **4**(2): 1–27.

Bachmann, R. and Sims, E. (2012). Confidence and the transmission of government spending shocks, *Journal of Monetary Economics* **59**: 235–249.

Baker, S., Bloom, N. and Davis, S. (2016). Measuring economic policy uncertainty, *Technical report*, Northwestern University, Standford University, and University of Chicago.

Bazzi, M., Blasques, F., Koopman, S. J. and Lucas, A. (2017). Time varying transition probabilities forMarkov regime switching models, *Journal of Time Series Analysis* **38**: 458–478.

Berger, D. and Vavra, J. (2014). Measuring how fiscal shocks affect durable spending in recessions and expansions, *American Economic Review Papers and Proceedings* **104**(5): 112–115.

Bloom, N. (2009). The impact of uncertainty shocks, *Econometrica* **77**(3): 623–685.

Brüggemann, R., Jentsch, C. and Trenkler, C. (2016). Inference in {VARs} with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **191**(1): 69 – 85.

Caggiano, G., Castelnuovo, E. and Groshenny, N. (2014). Uncertainty shocks and unemployment dynamics in u.s. recessions, *Journal of Monetary Economics* **67**: 78–92.

Colombo, V. (2013). Economic policy uncertainty in the u.s.: does it matter for the euro area?, *Economics Letters* **121**(1): 39–42.

Davis, S., Haltiwanger, J., Jarmin, R. and Miranda, J. (2007). Volatility and dispersion in business growth rates: Publicly traded versus privately held firms, *NBER Macroeconomics Annual* **21**: 107–180.

Diebold, F., Lee, J.-H. and Weinbach, G. (1994). Regime switching with time varying transition probabilities, *Business Cycles: Durations, Dynamics, and Forecasting* pp. 144–165.

Filardo, A. (1994). Business-cycle phases and their transitional dynamics, *Journal of Business and Economic Statistics* **12**(3): 299–308.

Gonçalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **123**(1): 89–120.

Herwartz, H. and Lütkepohl, H. (2014). Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks, *Journal of Econometrics* **183**(1): 104–116.

Kim, C.-J. (1994). Dynamic linear models with markov-switching, *Journal of Econometrics* **60**(1): 1 – 22.

Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*, Springer-Verlag, Berlin.

Lanne, M. and Lütkepohl, H. (2008). Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking* **40**(6): 1131–1149.

Lanne, M., Lütkepohl, H. and Maciejowska, K. (2010). Structural vector autoregressions with Markov switching, *Journal of Economic Dynamics and Control* **34**: 121–131.

Lütkepohl, H. and Netšunajev, A. (2017a). Structural vector autoregressions with heteroskedasticity: A review of different volatility models, *Econometrics and Statistics* **1**: 2 – 18.

Lütkepohl, H. and Netšunajev, A. (2017b). Structural Vector Autoregressions with Smooth Transition in Variances - The Interaction Between U.S. Monetary Policy and the Stock Market, *Journal of Economic Dynamics and Control* **forthcoming**.

Nodari, G. (2014). Financial regulation policy uncertainty and credit spreads in the u.s., *Journal of Macroeconomics.* **41**: 122–132.

Psaradakis, Z. and Spagnolo, N. (2006). Joint determination of the state dimension and autoregressive order for models with markov regime switching, *Journal of Time Series Analysis* **27**(5): 753–766.

Rigobon, R. and Sack, B. (2003). Measuring the reaction of monetary policy to the stock market., *Quarterly Journal of Economics* **118**(2).