Structural vector autoregression with time varying transition probabilities: identification via heteroskedasticity

Wenjuan Chen* Aleksei Netšunajev§¶

February 7, 2017

Abstract

Vector autoregression models with Markov regime switching variances have been exploited to test for structural identifications. However, these Markov regime switching models normally assume the transition probabilities to be constant over time while in reality, these probabilities could depend on certain economic fundamentals. Therefore, in this paper, we propose to include a new feature,i.e., the time varying transition probabilities, into Markov regime switching mechanism so as to test for validity of structural identifications. A generalized EM algorithm is developed for estimation of the model. As an illustration of the model we analyse the interaction of the US monetary policy and the stock market in a system including money aggregates. We find that the stock price index and money aggregates play a significant role in the Taylor rule. Furthermore we document strong impact effects of the monetary policy shocks on prices and output as well as a lagged decline of the stock prices after the shock.

Keywords: Structural vector autoregression, Markov switching, time varying transition probabilities, identification via heteroskedasticity.
JEL classification: C32.

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

§corresponding author, 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

Vector autoregressive models (VAR) with regime switching have been widely used because they are capable of capturing occasional but recurrent regime shifts. The most well-known applications include Hamilton (1989) on U.S. aggregate output and Sims and Zha (2006) on U.S. monetary policy. More recently, Lanne, Lütkepohl and Maciejowska (2010) develop tests on structural identification strategies exploiting information from regime-switching variances. However, in these papers it is assumed that the transition probabilities are constant over time. While in reality these probabilities can actually be time varying and depend on some underlying economic fundamentals.

Several papers have relaxed the constant transition probability assumption, including Diebold, Lee and Weinbach (1994), Filardo (1994), and Bazzi, Blasques, Koopman and Lucas (2014). 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 state 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 of the U.S. output compared with regime switching models with constant transition probabilities.

In this paper advancing on Diebold et al. (1994) and Herwartz and Lütkepohl (2014), we develop a new expectation maximization (EM) algorithm for estimation of the Markov switching structural VAR model with time-varying transition probabilities. We allow the probabilities of a regime shift to be changing over time and let it depend on an economic variable. Compared with Diebold et al. (1994), the algorithm developed in the paper uses the filter developed 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 this new 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 consider a system of six variables for the US on monthly frequency aiming to study the interaction between the monetary policy and the stock market. We include in our model not only the federal funds rate, but also the Divisia M2 aggregates and its opportunity cost so as to help identify monetary policy shocks. As demonstrated in Belongia and Ireland (2015) and Belongia and Ireland (2016), Divisia monetary aggregates are highly correlated with real activities, and monetary policy easing is associated with both a lower interest rate and a strongly accelerating rates of Divisia money growth. Therefore, we identify the monetary policy shocks through an extended Taylor rule that includes this liquidity measure. Furthermore, our framework allows us to test the hypothesis whether the financial market plays a role in the reaction function of policy makers, which is studied in Rigobon and Sack (2003) and Bjørnland and Leitemo (2009).

Our estimation results show that according to information criteria, the model with time varying transition probabilities outperform the standard linear model and the Markov switching models with changes in variances but with constant transition probabilities. We also estimate models with several alternative candidates as the transition variables. Based on information criteria the lagged policy rate governing the changes in volatility is the most preferred transition variable.

Regarding the identification schemes, the likelihood ratio tests provide no support for the popular Cholesky decomposition, or the non-recursive scheme by Belongia and Ireland (2015). However, further tests on individual zero restriction demonstrate that excluding the stock market from the monetary policy reaction function is strongly rejected at 1% significant level, which is in line with those of Rigobon and Sack (2003) and Bjørnland and Leitemo (2009). The most plausible monetary policy function is the one taking into consideration of output, price, monetary aggregates as well as stock market movements, which is accepted at 1% significant level. Furthermore, the monetary policy shocks identified through heterskedasticity generates reasonable impulse responses, such as negative impact effects on monetary aggregates, and significant negative effects on both output and stock prices.

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 US monetary policy and the stock market 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

First 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 is the $K \times K$ coefficient matrices, and u_t is the vector of reduced form residuals which has zero mean and covariance matrix Σ_u. In order to obtain economically meaningful structural residuals ε_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 assume that the distribution of u_t depends 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 assume that the transition probabilities depend on a vector of economic fundamentals x_t and evolve as 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 β_{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. In particular, instantaneous effects are the same across the states. If there are just two regimes with positive definite covariance matrices Σ_1 Σ_2, it is well known that there exists a matrix B that satisfies $\Sigma_1 = BB'$ and $\Sigma_2 = B\Lambda_2 B'$ where Λ_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 Λ_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.

As for the case of more than two regimes, $\Sigma_1 = BB'$, $\Sigma_i = B\Lambda_i B'$, $i = 2, ..., M$, where Λ_i are diagonal matrices. The condition for B to be unique is that, if one pair of diagonal elements from Λ_2 are the same, there must be another pair of distinct diagonal elements from one of the other Λ_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, $H_0 : \lambda_{21} = \lambda_{22}$ for example, implies that some parameters are not identified under H_0 and standard χ^2 asymptotic properties are not valid (Lütkepohl and Netšunajev, 2017).

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 develop a new expectation maximization (EM) algorithm that advances on the algorithms by Diebold et al. (1994) and Herwartz and Lütkepohl (2014) for the actual likelihood maximization task. The iterative algorithm consists of expectation step where the estimates of the unobserved regime probabilities are obtained and the maximisation 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 in the maximization part of the algorithm of Herwartz and Lütkepohl (2014) to estimate the transition parameters β_{ij}. As the first order conditions of the likelihood function are nonlinear in β_{ij}, we use linear approximation of p_t^{ij} around β_{ij}^{n-1} that comes from from the previous iteration. Consider β_{11} as an example:

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

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 β_{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. The diagonal elements of the $\Lambda_m, m = 2, ..., M$ are bounded away from zero. 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 η_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 this framework of in the models set up in Herwartz and Lütkepohl (2014) or Lütkepohl and Netšunajev (2014a). 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 Empirical example

3.1 The data

In this section we illustrate our approach with a structural model similar to Belongia and Ireland (2015). The vector of variables takes the form: $y_t = (p_t \; ip_t \; \Delta sp_t \; r_t \; m_t \; uc_t)'$. The series enter the VAR in the following form:

- p_t is the first difference of the log of consumer price index;
- ip_t is the first difference of the log of the industrial production;
- sp_t is the log of the nominal S&P500 stock price index. The series is first differenced to represent monthly returns (Δsp_t);

- r_t denotes the federal funds rate;
- m_t is the Divisia M2 aggregates.
- uc_t is the corresponding opportunity cost of the Divisia M2.

We use monthly U.S. data for the period 1967M2 - 2008M12 which excludes the period of the recent financial crisis. The lag order is chosen to be three according to the Akaike information criterion.

3.2 Various identification schemes

The interaction between monetary policy and the stock market is studied in many papers that use SVAR models. As discussed in Belongia and Ireland (2015), there is no consensus on the identifying scheme that serves to pin down the monetary policy shocks. The most popular way of identification is the Cholesky decomposition used, for instance, by Millard and Wells (2003), Thorbecke (1997), and Cheng and Jin (2013). Another way to identify monetary policy shocks, which is used in Leeper and Roush (2003) and Belongia and Ireland (2015), is to use a modified Taylor rule that includes monetary aggregates.

The following matrices shows the identification scheme on B matrix via the Cholesky decomposition, and the non-recursive identification strategy on A matrix in the spirit of Belongia and Ireland (2015):

$$B = \begin{bmatrix} b_{11} & 0 & 0 & 0 & 0 & 0 \\ b_{21} & b_{22} & 0 & 0 & 0 & 0 \\ b_{31} & b_{32} & b_{33} & 0 & 0 & 0 \\ b_{41} & b_{42} & b_{43} & b_{44} & 0 & 0 \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} & 0 \\ b_{61} & b_{62} & b_{63} & b_{64} & b_{65} & b_{66} \end{bmatrix} \text{ or } A = \begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & 0 & a_{44} & a_{45} & 0 \\ -a_{55} & a_{52} & 0 & 0 & a_{55} & a_{56} \\ -a_{65} & 0 & 0 & a_{64} & a_{65} & a_{66} \end{bmatrix}$$

The zero restrictions on the B matrix are based on assumptions of impact effects of structural shocks, for example, the monetary policy shocks are typically assumed to have no impact effects on real output. The most identifying restrictions on the A matrix follow the non-recursive scheme and are motivated by economic theories. For example, the fourth row of the matrix is based on the monetary policy reaction function: $a_{41}p_t + a_{42}ip_t + a_{43}sp_t + a_{44}r_t + a_{45}m_t = \varepsilon_t^{mp}$. The fifth row is motivated by the money demand function, i.e., $a_{52}ip_t + a_{55}(m_t - p_t) + a_{56}uc_t = \varepsilon_t^{md}$.

Even though different data and sometimes controversial identifying assumptions are used, most studies suggest that monetary policy shocks affect stock prices in an important way. Several studies find that stock market is affected by monetary policy on impact. Thorbecke (1997) documents a 0.8% decrease in stock prices after an increase in the federal funds rate of one standard deviation. Similarly, Li, Iscan and Xu (2010) observe an immediate and relatively prolonged negative response of stock prices to a monetary policy shock. Bjørnland and Leitemo (2009) use a combination of short- and long-run restrictions and estimate a very strong and persistent decline of around 12% in stock prices after a contractionary monetary policy shock of 100 basis points. All various identification schemes share a drawback that the restrictions just-identify the shocks and cannot be tested against the data.

Table 1: Various identification restrictions assumed on B or A matrices

B1	The B is a lower triangular matrix.
B2	The monetary policy shock has no instantaneous effects on output, prices and stock returns: $b_{14} = b_{24} = b_{34} = 0$.
B3	The monetary policy shock has no instantaneous effects on stock returns: $b_{34} = 0$.
A1	The matrix A has all the zero restrictions following Belongia and Ireland (2015).
A2	It is restricted that opportunity cost of holding money is excluded from the monetary policy reaction function: $a_{46} = 0$.
A3	It is restricted that stock prices is excluded from the monetary policy reaction function: $a_{43} = 0$.
A4	The element for the money demand setting are restricted as: $a_{51} = -a_{55}$, $a_{53} = a_{54} = 0$.

It is also highly debated that whether the monetary policy makers takes into consideration the movements in the stock market. Most of the literature on monetary policy analysis ignores the role of financial market, and only consider the output and inflation in the policy reaction function. However, instability in financial market can cause significant fluctuations in the real sector, as observed in the recent financial crisis. Moreover, studies such as Rigobon and Sack (2003) and Bjørnland and Leitemo (2009) document supporting evidence of significant responses of short term interest rates to stock market movements.

If our system is identified through changes in volatility, then any of above mentioned restrictions become over-identifying. Therefore we could test not only a set of restrictions such as Cholesky decomposition, but also any individual restriction or a subset of restrictions that we are interested in. Table 1 summarizes seven different sets of restrictions where some are based on the impact effect matrix B and some are based on its inverse $A = B^{-1}$. We denote by the b_{ij} and a_{ij}

the specific entries in row i and column j of matrices B and A. We go on and test the restrictions of Table 1 using our model.

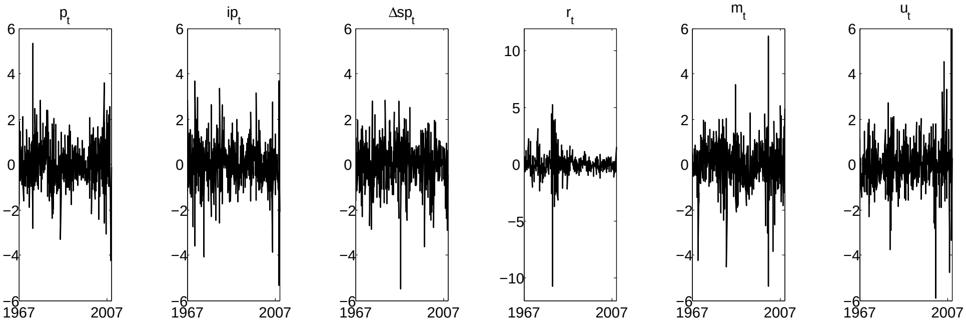
3.3 Statistical analysis

The summary statistics for the estimated models are shown in Table 2. The models with time varying transition probabilities outperform the standard MS models if we judge according to information criteria. We estimate models with two transition variables π_{t-1} being the lag of CPI inflation and r_{t-1} being the lag of the federal funds rate. The choice of transition variables expresses our belief that the evolution of volatility in this system is linked to monetary policy directly trough interest rate or indirectly through inflation. It is worth noting that Filardo (1994) uses the policy rate as the transition variable and finds it to be very suitable.

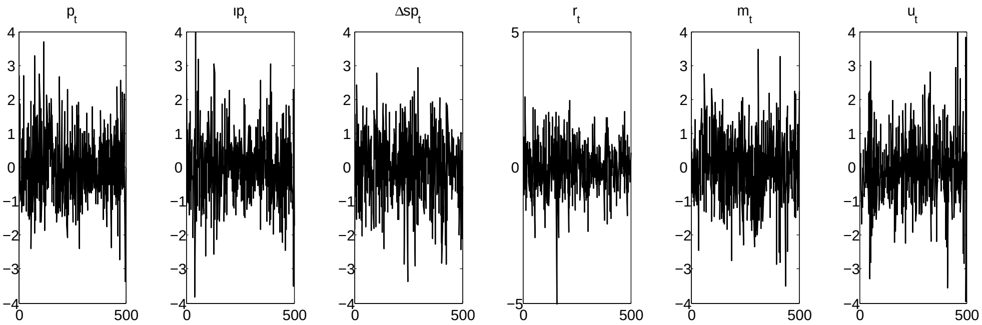
Table 2: Comparison of VAR(3) Models for $y_t = (p_t, ip_t, \Delta sp_t, r_t, m_t, u_t)'$, Sample Period: 1967M2 - 2008M12

Model	x_{t-1}	$\log L_T$	AIC	SC
VAR(3), linear		8076.690	-15883.380	-15314.408
		MS-VAR with constant P		
MS(2)		8538.568	-16761.136	-16095.228
MS(3)		8599.803	-16833.606	-16062.333
		MS-VAR with time varying P		
MS(2)	$(1;\pi_{t-1})$	8526.068	-16732.137	-16057.799
MS(3)	$(1;\pi_{t-1})$	8585.261	-16792.521	-15995.960
MS(3), state invariant B	$(1;\pi_{t-1})$	8571.171	-16794.344	-16061.002
MS(2)	$(1;r_{t-1})$	8532.030	-16744.060	-16069.723
MS(3)	$(1;r_{t-1})$	8613.496	-16848.993	-16052.432
MS(3), state invariant B	$(1;r_{t-1})$	8613.182	**-16878.365**	**-16145.023**
		MS-VAR with time varying P under various restrictions		
MS(3), B1	$(1;r_{t-1})$	8599.091	-16880.184	-16210.061
MS(3), B2	$(1;r_{t-1})$	8600.906	-16859.812	-16139.114
MS(3), B3	$(1;r_{t-1})$	8613.156	-16880.312	-16151.185
MS(3), A1	$(1;r_{t-1})$	8593.477	-16868.954	-16198.831
MS(3), A2	$(1;r_{t-1})$	8611.015	-16876.030	-16146.903
MS(3), A3	$(1;r_{t-1})$	8603.772	-16861.544	-16132.417
MS(3), A4	$(1;r_{t-1})$	8599.055	-16856.110	-16135.412

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.



(a) Residual of the VAR(3) model



(b) Residuals of the MS(3) model with state invariant B, $x_{t-1} = (1; r_{t-1})$

Figure 1: Residuals of various models

Choosing the number of regimes is critical for our analysis. Following Psaradakis and Spagnolo (2006) and Herwartz and Lütkepohl (2014) we use the information criteria as the tool to select the number of regimes as well as the tool to motivate the choices of transition variables. If judged upon these criteria, the preferred model is the one with three regimes and lagged federal funds rate as the transition variable. 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 that the ones of the standard model. Estimated parameters of the transition function for the model are shown in Table 3. These are estimated rather precisely with the standard errors being mostly smaller than the estimates.

The estimated time varying transition probabilities are shown on Figure 2. Take for example p_{21}, which represents the probability of switching from Regime 2 to Regime 1. This probability p_t^{21} takes value of zero for most of the sample, however, from the end of 1970s till the mid-1980s, this probability switches to

Table 3: Estimated transition parameters and their standard errors

	β_{11}	β_{12}	β_{21}	β_{22}	β_{32}	β_{33}
Estimate(intercept)	-2.729	-3.571	1.059	10.248	-10.919	2.224
Estimate(slope)	0.855	0.836	0.018	-0.771	1.613	0.245
Std.err.(intercept)	3.743	4.123	4.948	5.239	12.729	3.382
Std.err.(slope)	0.772	0.796	0.414	0.485	1.477	0.963

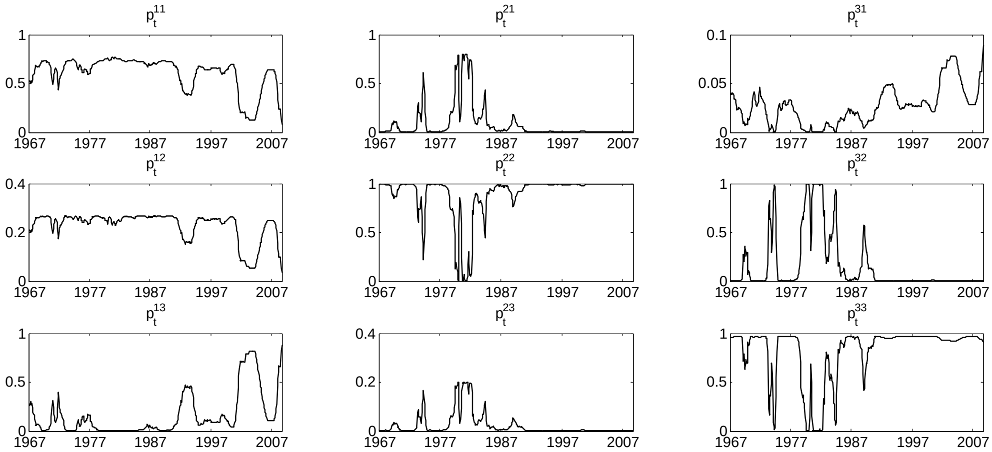


Figure 2: Time varying transition probabilities

almost one. Restricting this probability to be constant over time could have led to misspecification. The estimated smoothed regime probabilities are displayed on Figure 3. The first regime, which is the most volatile one, covers some period in the beginning of the 1970s, the beginning of the 1980s and several short periods ahead of the recent financial crisis. The second regime includes most of the periods in the 1970s, 1980s till the end of 1990s. The third regime covers almost the whole period between 1998 till 2008 except the Dot com bubble period and the periods ahead of the financial crisis. The timing of the low-volatility regimes corresponds to the well known period of the Great Moderation.

It is important to check whether the estimated model is identified by comparing the pairs of the relatives variances. The estimates of these parameters along with their standard errors are shown in Table 4. In the situation where no formal tests for identification exist the standard errors of the variances have to be examined (Lütkepohl and Netšunajev, 2017). In the case of the preferred three state

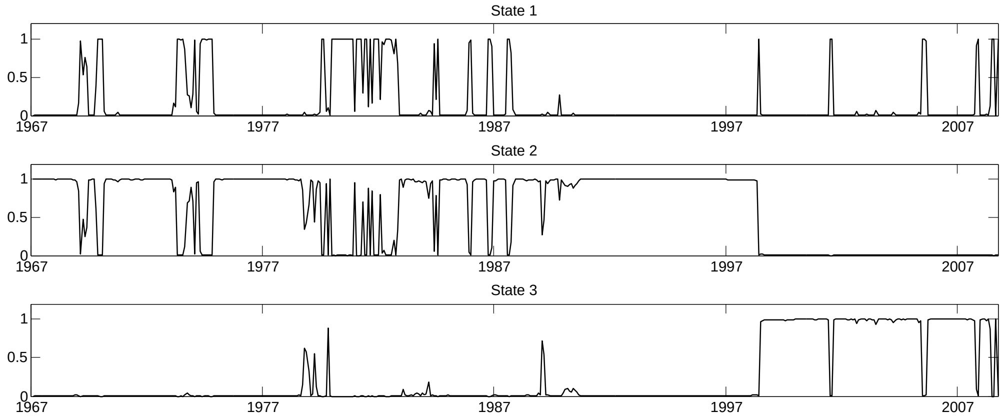


Figure 3: Estimated smoothed regime probabilities

model the estimates of the Λ_2 and Λ_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.

Results of testing the restrictions are show in Table 5. It becomes clear that neither the Cholesky identification scheme nor the non-recursive scheme in line with Belongia and Ireland (2015) are compatible with the data. The model $B3$ gets accepted, which suggests that it is fine to set the zero restriction on the impact effect of the monetary policy shock on the stock returns. This finding contrasts to the claims of Thorbecke (1997), Li et al. (2010), and Bjørnland and Leitemo (2009) who document pronounced impact effects of monetary shocks on the stock market.

Considering the restrictions on the A matrix, the model $A2$ is rejected at a 1% significance level but accepted at a 5% significance level. In contrast, the model $A3$ is rejected with the p-value of 1.4×10^{-5}. These results demonstrate that there is evidence against excluding the stock prices from the policy reaction function, which confirms the finding by Rigobon and Sack (2003). However the opportunity cost for holding the monetary aggregates uc_t is not that important component of the Taylor rule. Rejection of restrictions $A4$ that are based on the money demand function shows that stock market and policy rate are important for money demand.

It is interesting to note that restrictions $B1$ and $A1$ that would fully identify the conventional VAR are rejected by the data. Similarly several sets of identifying restrictions in a monetary system are rejected in the study of Lütkepohl and Netšunajev (2017). In that paper the system analyzed consisted of the monthly

Table 4: Estimated relative variances of the MS(3) model with state invariant B, $x_{t-1} = (1; r_{t-1})$

Parameter	Estimate	Standard error
λ_{21}	0.302	0.067
λ_{22}	0.228	0.053
λ_{23}	0.205	0.060
λ_{24}	0.027	0.005
λ_{25}	0.103	0.023
λ_{26}	0.188	0.043
λ_{31}	0.137	0.040
λ_{32}	0.316	0.089
λ_{33}	0.456	0.156
λ_{34}	0.004	0.001
λ_{35}	0.282	0.070
λ_{36}	0.628	0.206

data as well. That may be an indication that monthly data contains much more interactions and the restrictions based on economic theory may not be applied equally to data at various frequencies.

Table 5: Tests for Identifying Restrictions in MS(3) models

H_0	H_1	df	LR statistic	p-value
B1	MS(3), state invariant B	15	28.181	0.020
B2	MS(3), state invariant B	3	24.553	1.9×10^{-5}
B3	MS(3), state invariant B	1	0.052	0.819
A1	MS(3), state invariant B	15	39.410	5.5×10^{-4}
A2	MS(3), state invariant B	1	4.334	0.037
A3	MS(3), state invariant B	1	18.821	1.4×10^{-5}
A4	MS(3), state invariant B	3	28.255	3.2×10^{-6}

Impulse responses for the model identified via changes in volatility and with state invariant B are shown in Figure 4. The fourth shock is the only shock that has a pronounced effect on the federal funds rate. Thus it is the only candidate for the monetary policy shock. The impact effect of the shock on prices and industrial production is quite pronounced with the price puzzle being rather substantial. It should be mentioned however that the study of Belongia and Ireland (2015) reports the price puzzle as well without showing confidence bands around impulse

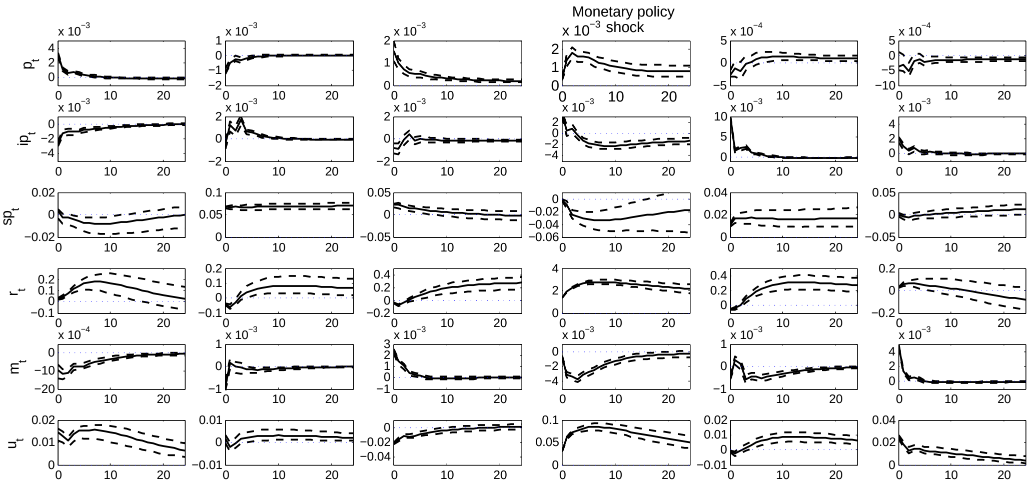


Figure 4: Impulse responses for the MS(3) model identified via changes in volatility

responses. Lütkepohl and Netšunajev (2014b) document price puzzle in a similar context. The stock returns go down by nearly 3% during half a year after a shocks which is lower than the Bjørnland and Leitemo (2009) suggest. Lütkepohl and Netšunajev (2017) study impulse responses for various models with heteroscedasticity. Even though we have a different model, different data, and different sample period the overall reaction of the variables to the monetary policy shocks is quite similar to those in Lütkepohl and Netšunajev (2017): there is a pronounced price puzzle, lagged negative effect on output and only lagged effect on the stock returns.

4 Conclusions

In the paper we propose a structural vector autoregressive model where the 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 the model we analyze the interaction of the

US monetary policy and the stock market using a monetary model that includes monetary aggregates. Inflation and federal funds rate are chosen to govern the variation in the transition probabilities. The model with three Markov regimes and federal funds rate as the transition variable has the best fit. The structural parameters are estimated precisely and this forms good ground to believe that the model is identified.

Seven economic theory based restrictions are tested using likelihood ratio tests. The model with a full set of restrictions that identify shocks in the conventional case get rejected. On the contrary, we find strong support that the stock market does not contemporaneously react to monetary policy shock and some support that the user cost of M2 may be excluded from the Taylor rule. The impulse responses of the model identified via heteroskedasticity indicate strong impact effect of monetary policy shocks on prices and output and a delayed decline in the stock prices.

Acknowledgments

We thank... for their 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:

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

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 β_{ij} is the $(J+1) \times 1$ vector of parameters. It is convenient to collect the individual β_{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 β_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 β_{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:

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

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 β_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:

η_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 $t = T-1, ..., 0$.

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

Maximization step

Estimation of transition parameters β

Given the smoothed probabilities, the expected complete-data log likelihood are non-linear in the β transition parameters. Taking into account the logistic transition function, the first order conditions for β 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 β vectors. Consider β_{11} as an example:

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

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

<u>$M = 2$</u>:

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

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

<u>$M = 3$</u>:

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

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

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

$$\beta_{21} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=2|Y_T,X) p_{1t}^{21} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=2|Y_T,X) p_{Jt}^{21} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=2|Y_T,X) p_{1t}^{21} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=2|Y_T,X) p_{Jt}^{21} \end{pmatrix}^{-1}$$

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

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

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

$$\beta_{32} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=3|Y_T,X) p_{1t}^{32} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=3|Y_T,X) p_{Jt}^{32} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=3|Y_T,X) p_{1t}^{32} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=3|Y_T,X) p_{Jt}^{32} \end{pmatrix}^{-1}$$

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

$$\beta_{33} = \begin{pmatrix} \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=3|Y_T,X) p_{1t}^{33} & \dots & \sum_{t=2}^{T} x_{1,t-1} Pr(s_{t-1}=3|Y_T,X) p_{Jt}^{33} \\ \vdots & \ddots & \vdots \\ \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=3|Y_T,X) p_{1t}^{33} & \dots & \sum_{t=2}^{T} x_{J,t-1} Pr(s_{t-1}=3|Y_T,X) p_{Jt}^{33} \end{pmatrix}^{-1}$$

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

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 Λ_m :

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

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

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 θ 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:

η_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 β, B, Λ_m and θ are iterated until convergence, i.e. relative change Δ 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 ζ_t is a random variable taking values 1 and -1, each with probability 0.5. We bootstrap parameter estimates θ^*, B^* and Λ^* conditionally on the initially estimated transition probabilities.

References

Bazzi, M., Blasques, F., Koopman, S. J. and Lucas, A. (2014). Time varying transition probabilities forMarkov regime switching models, *Tinbergen institute discussion papers*.

Belongia, M. T. and Ireland, P. N. (2015). Interest rates and money in the measurement of monetary policy, *Journal of Business & Economic Statistics* **33**(2): 255–269.

Belongia, M. T. and Ireland, P. N. (2016). Money and output: Friedman and schwartz revisited, *Journal of Money, Credit and Banking* **48**: 1223–1266.

Bjørnland, H. C. and Leitemo, K. (2009). Identifying the interdependence between us monetary policy and the stock market, *Journal of Monetary Economics* **56**(2): 275–282.

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

Cheng, L. and Jin, Y. (2013). Asset prices, monetary policy, and aggregate fluctuations: An empirical investigation, *Economics Letters* **119**: 24–27.

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.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**: 357–84.

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.

Leeper, E. M. and Roush, J. E. (2003). Putting'm'back in monetary policy, *Technical report*, National Bureau of Economic Research.

Li, Y. D., Iscan, T. B. and Xu, K. (2010). The impact of monetary policy shocks on stock prices: Evidence from Canada and the United States, *Journal of International Money and Finance* **29**: 876–896.

Lütkepohl, H. and Netšunajev, A. (2014a). Structural Vector Autoregressions with Smooth Transition in Variances - The Interaction Between U.S. Monetary Policy and the Stock Market, *Sfb 649 discussion papers*, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Lütkepohl, H. and Netšunajev, A. (2014b). Structural Vector Autoregressions with Smooth Transition in Variances - The Interaction Between U.S. Monetary Policy and the Stock Market, *SFB 649 Discussion Papers SFB649DP2014-031*, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

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

Millard, S. P. and Wells, S. J. (2003). The role of asset prices in transmitting monetary and other shocks, *Bank of England Working Papers 188*, Bank of England.

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).

Sims, C. A. and Zha, T. (2006). Were there regime switches in u.s. monetary policy?, *American Economic Review* **96**: 54–81.

Thorbecke, W. (1997). On stock market returns and monetary policy, *Journal of Finance* **52**: 635–654.