

Time Varying Transition Probabilities for Markov Regime Switching Models*

Marco Bazzi^(a), Francisco Blasques^(b)
Siem Jan Koopman^(b,c), André Lucas^(b)

^(a) *University of Padova, Italy*

^(b) *VU University Amsterdam and Tinbergen Institute, The Netherlands*

^(c) *CREATES, Aarhus University, Denmark*

Abstract

We propose a new Markov switching model with time varying probabilities for the transitions. The novelty of our model is that the transition probabilities evolve over time by means of an observation driven model. The innovation of the time varying probability is generated by the score of the predictive likelihood function. We show how the model dynamics can be readily interpreted. We investigate the performance of the model in a Monte Carlo study and show that the model is successful in estimating a range of different dynamic patterns for unobserved regime switching probabilities. We also illustrate the new methodology in an empirical setting by studying the dynamic mean and variance behaviour of U.S. Industrial Production growth. We find empirical evidence of changes in the regime switching probabilities, with more persistence for high volatility regimes in the earlier part of the sample, and more persistence for low volatility regimes in the later part of the sample.

Some key words: Hidden Markov Models; observation driven models; generalized autoregressive score dynamics.

JEL classification: C22, C32.

*The authors thank participants of the “2014 Workshop on Dynamic Models driven by the Score of Predictive Likelihoods”, La Laguna, and seminar participants and VU University Amsterdam for useful comments and discussions. Blasques and Lucas thank the Dutch Science Foundation (NWO, grant VICI453-09-005) for financial support. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

1 Introduction

Markov regime switching models have been widely applied in economics and finance. Since the seminal application of Hamilton (1989) to U.S. real Gross National Product growth and the well-known NBER business cycle classification, the model has been adopted in numerous other applications. Examples are switches in the level of a time series, switches in the (autoregressive) dynamics of vector time series, switches in volatilities, and switches in the correlation or dependence structure between time series; see Hamilton and Raj (2002) for a partial survey. The key attractive feature of Markov switching models is that the conditional distribution of a time series depends on an underlying latent state or regime, which can take only a finite number of values. The discrete state evolves through time as a discrete Markov chain and we can summarize its statistical properties by a transition probability matrix.

Diebold et al. (1994) and Filardo (1994) argue that the assumption of a constant transition probability matrix for a Markov switching model is too restrictive for many empirical settings. They extend the basic Markov switching model to allow the transition probabilities to vary over time using observable covariates, including strictly exogenous explanatory variables and lagged values of the dependent variable. Although this approach can be useful and effective, it is not always clear what variables or which functional specification we should use for describing the dynamics in the transition probabilities.

Our main contribution in this paper is to propose a new, dynamic approach to model time variation in transition probabilities in Markov switching models. We let the transition probabilities vary over time as specific transformations of the lagged observations. Hence we adopt an observation driven approach to time varying parameter models; see Cox (1981) for a detailed discussion. Observation driven models have the advantage that the likelihood is typically available in closed form using a prediction error decomposition. Our main challenge is to specify a suitable functional form to link past observations to future transition probabilities. For this purpose, we use the scores of the predictive likelihood function. Such score driven dynamics have been introduced by Creal et al. (2011, 2013) and Harvey (2013). Score driven models encompass many well-known time series models in economics and finance, including the ARCH model of Engle (1982), the generalized ARCH (GARCH) model of Bollerslev (1986), the exponential GARCH (EGARCH) model of Nelson (1991), the autoregressive conditional duration (ACD) model of Engle and Russell (1998), and many more. In addition, various successful applications of score models have appeared in the recent literature. For example, Creal et al. (2011) and Lucas et al. (2014) study dynamic volatilities and correlations under fat-tails and possible skewness; Harvey and Luati (2014) introduce new models for dynamic changes in levels under fat tails; Creal et al. (2014) investigate score-based mixed measurement dynamic factor models; Oh and Patton (2013) and De Lira Salvatierra and Patton (2013) investigate factor copulas based on score dynamics; and Koopman et al. (2012) show that score driven time series models have a similar fore-

casting performance as correctly specified nonlinear non-Gaussian state space models over a range of model specifications.

We show that the score function in our Markov switching model has a highly intuitive form. The score combines all relevant innovative information from the separate models associated with the latent states. The updates of the time varying parameters are therefore based on the probabilities of the states, given all information up to time $t - 1$. In our simulation experiments, the new model performs well and succeeds in capturing a range of time varying patterns for the unobserved transition probabilities.

We apply our model to study the monthly evolution of U.S. Industrial Production growth from January 1919 to October 2013. We uncover three regimes for the mean and two regimes for the variance over the sample period considered. The corresponding transition probabilities are time varying. In particular, the high volatility regime appears to be much more persistent in the earlier part of the sample compared to the later part. The converse holds for the low volatility regime. Such changes in the dynamics of the time series are captured in a straightforward way within our model. Moreover, the fit of the new model outperforms the fit of several competing models.

As a final contribution, it is worthwhile mentioning that our model also presents an interesting mix of parameter driven (Markov switching) dynamics with observation driven score dynamics for the corresponding (transition probability) parameters. In particular, it is interesting to see that score driven models can still be adopted when an additional filtering step (for the unobserved discrete states) is required to compute the score of the resulting conditional observation density. This feature of the new dynamic switching model is interesting in its own right. Similar developments for a linear Gaussian state space model have been reported by Creal et al. (2008) and Delle Monache and Petrella (2014)

The remainder of the paper is organized as follows. In Section 2 we briefly discuss the main set-up of the Markov switching model and its residual diagnostics. In Section 3 we introduce the new Markov switching model with time varying transition probabilities based on the score of the predictive likelihood function. In Section 4 we discuss some of the statistical properties of the model. In Section 5 we report the results of a Monte Carlo study. In Section 6 we present the results of our empirical study into the dynamic salient features of U.S. Industrial Production growth. Section 7 concludes.

2 Markov switching models

Markov switching models are well-known and widely used in applied econometric studies. We refer to the textbook of Frühwirth-Schnatter (2006) for an extensive introduction and discussion. The treatment below establishes the notation and discusses some basic notions of Markov switching models.

Let $\{y_t, t = 1, \dots, T\}$ denote a time series of T univariate observations. We consider the time series $\{y_t, t = 1, \dots, T\}$ as a subset of a stochastic process $\{y_t\}$. The probability distribution of the stochastic process y_t depends on the realizations of a hidden discrete stochastic process z_t . The stochastic process y_t is directly observable, whereas z_t is a latent random variable that is observable only indirectly through its effect on the realizations of y_t . The hidden process $\{z_t\}$ is assumed to be an irreducible and aperiodic Markov chain with finite state space $\{0, \dots, K-1\}$. Its stochastic properties are sufficiently described by the $K \times K$ transition matrix, Π , where π_{ij} is the $(i+1, j+1)$ element of Π and is equal to the transition probability from state i to state j . All elements of Π are nonnegative and the elements of each row sum to 1, that is

$$\pi_{ij} = \mathbb{P}[z_t = j | z_{t-1} = i], \quad \sum_{j=0}^{K-1} \pi_{ij} = 1, \quad \pi_{ij} \geq 0, \quad \forall i, j \in \{0, \dots, K-1\}. \quad (1)$$

Let $p(\cdot | \theta_i, \psi)$ be a parametric conditional density indexed by parameters $\theta_i \in \Theta$ and $\psi \in \Psi$, where θ_i is a regime dependent parameter and ψ is not regime-specific. We assume that the random variables y_1, \dots, y_T are conditionally independent given z_1, \dots, z_T , with densities

$$y_t | (z_t = i) \sim p(\cdot | \theta_i, \psi). \quad (2)$$

For the joint stochastic process $\{z_t, y_t\}$, the conditional density of y_t is

$$p(y_t | \psi, I_{t-1}) = \sum_{i=0}^{K-1} p(y_t | \theta_i, \psi) \mathbb{P}(z_t = i | \psi, I_{t-1}), \quad (3)$$

where $I_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$ is the observed information available at time $t-1$. All parameters ψ and $\theta_0, \dots, \theta_{K-1}$ are unknown and need to be estimated.

The conditional mean of y_t given z_t and I_{t-1} may contain lags of y_t itself. Francq and Roussignol (1998) and Francq and Zakoian (2001) derive the conditions for the existence of an ergodic and stationary solution for the general class of Markov switching ARMA models. In particular, they show that global stationarity of y_t does not require the stationarity conditions within each regime separately.

As an example, consider the case $K = 2$ for a continuous variable y_t with conditional density

$$p(\cdot | z_t) = N((1 - z_t)\mu_0 + z_t\mu_1, \sigma^2), \quad (4)$$

where μ_0 and μ_1 are static regime-dependent means, and σ^2 is the common variance. The latent two-state process $\{z_t\}$ is driven by the transition probability matrix Π

$$\Pi = \begin{pmatrix} \pi_{00} & 1 - \pi_{00} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix}, \quad (5)$$

where the transition probabilities satisfy $0 < \pi_{00}, \pi_{11} < 1$. We have $\theta_i = \mu_i$ for $i = 0, 1$, and $\psi = (\sigma^2, \pi_{00}, \pi_{11})'$.

To evaluate equation (3), we require the quantities $\mathbb{P}(z_t = i|\psi, I_{t-1})$ for all t . We can compute these efficiently using the recursive filtering approach of Hamilton (1989). Assuming we have an expression for the filtered probability $\mathbb{P}(z_{t-1} = i|\psi, I_{t-1})$, we can obtain the predictive probabilities $\mathbb{P}(z_t = i|\psi, I_{t-1})$ as

$$\mathbb{P}(z_t = i|\psi, I_{t-1}) = \sum_{k=0}^{K-1} \pi_{ki} \cdot \mathbb{P}(z_{t-1} = k|\psi, I_{t-1}). \quad (6)$$

Hence, the conditional density of y_t given I_{t-1} is given by

$$p(y_t|\psi, I_{t-1}) = \sum_{i=0}^{K-1} \sum_{k=0}^{K-1} p(y_t|\theta_i, \psi) \cdot \pi_{ki} \cdot \mathbb{P}(z_{t-1} = k|\psi, I_{t-1}). \quad (7)$$

We can rewrite this expression more compactly in matrix notation. Define ξ_{t-1} as the K -dimensional vector containing the filtered probabilities $\mathbb{P}(z_{t-1} = i|\psi, I_{t-1})$ at time $t-1$ and let η_t be the K -dimensional vector collecting the densities $p(y_t|\theta_i, \psi)$ at time t for $i = 0, \dots, K-1$. It follows that (7) reduces to

$$p(y_t|\psi, I_{t-1}) = \xi_{t-1}' \Pi \eta_t. \quad (8)$$

The filtered probabilities ξ_t can be updated by the Hamilton recursion

$$\xi_t = \frac{(\Pi' \xi_{t-1}) \odot \eta_t}{\xi_{t-1}' \Pi \eta_t}, \quad (9)$$

where \odot denotes the Hadamard element by element product. The filter needs to be started from an appropriate set of initial probabilities $\mathbb{P}(z_0 = i|\psi, I_0)$. The smoothed estimates of the regime probabilities $\mathbb{P}(z_t = i|\psi, I_T)$ can be obtained from the algorithm of Kim (1994). The Hamilton filter in (9) is implemented for the evaluation of the the log-likelihood function which is numerically maximized with respect to the parameter vector $(\theta'_0, \dots, \theta'_{K-1}, \psi')'$ using a quasi-Newton optimization algorithm. To avoid local maxima, we consider different starting values for the numerical optimization.

Diagnostic checking in Markov regime switching models is somewhat more complicated when compared to other time series models because the true residuals depend on the latent variable z_t . Hence the residuals are unobserved. A standard solution is the use of generalized residuals which have been introduced by Gouriéroux et al. (1987) in the context of latent variable models. They have been used in the context of Markov regime switching models by Turner et al. (1989), Gray (1996), Maheu and McCurdy (2000), and Kim et al. (2004). Given the filtered regime probabilities $\mathbb{P}(z_t = i|\psi, I_{t-1})$, for $i = 0, \dots, K-1$, let μ_i and σ_i^2

denote the conditional mean and the conditional variance of y_t in regime i . The standardized generalized residual e_t is defined as

$$e_t = \sum_{i=0}^{K-1} \frac{y_t - \mu_i}{\sigma_i} \mathbb{P}(z_t = i | \psi, I_{t-1}), \quad t = 1, \dots, T. \quad (10)$$

Also in the context of switching models, Smith (2008) adopts the transformation proposed by Rosenblatt (1952) and defines the Rosenblatt residual \tilde{e}_t as

$$\tilde{e}_t = \Phi^{-1} \left(\sum_{i=0}^{K-1} \mathbb{P}(z_t = i | \psi, I_{t-1}) \Phi(\sigma_i^{-1} (y_t - \mu_i)) \right), \quad (11)$$

where Φ denotes the cumulative distribution function of a standard normal with the corresponding inverse function Φ^{-1} . If y_t is generated by the distribution implied by the Markov switching model, then the Rosenblatt residual \tilde{e}_t is standard normally distributed. Furthermore, Smith (2008) shows in an extensive Monte Carlo study that Ljung-Box tests based on the Rosenblatt transformation have good finite-sample properties for the diagnostic checking of serial correlation in the context of Markov regime switching models.

3 Time varying transition probabilities

In the previous section we considered the transition probability matrix Π to be constant over time. Diebold et al. (1994) and Filardo (1994) argue for having time varying transition probabilities Π_t . They propose to let the elements of Π_t be functions of past values of the dependent variable y_t and of exogenous variables. The Hamilton filter and Kim smoother can easily be generalized to handle such cases of time varying Π_t . A key challenge is to specify an appropriate and parsimonious function that links the lagged dependent variables to future transition probabilities. For the specification of the dynamics of Π_t , we adopt the generalized autoregressive score dynamics of Creal et al. (2013); similar dynamic score models have been proposed by Creal et al. (2011) and Harvey (2013). We provide the details of the score driven model for time varying transition probabilities in the Markov regime switching model. The new dynamic model is parsimonious and the updating mechanism is highly intuitive. Each probability update is based on the weighting of the likelihood information $p(\cdot | \theta_i, \psi)$ in (2) for each separate regime i .

3.1 Dynamics driven by the score of predictive likelihood

The parameter vector ψ contains both the transition probabilities as well as other parameters capturing the shape of the conditional distributions $p(y_t | \psi, I_{t-1})$. With a slight abuse of notation, we split ψ into a dynamic parameter f_t that we use to capture the dynamic

transition probabilities, and a new static parameter ψ^* that gathers all remaining static parameters in the model, as well as some new static parameters that govern the transition dynamics of f_t . For example, in the two-state example of Section 2 we may choose $f_t = (f_{00,t}, f_{11,t})'$ with $f_{00,t} = \text{logit}(\pi_{00,t})$ and $f_{11,t} = \text{logit}(\pi_{11,t})$, where $\text{logit}(\pi_{00,t}) = \log(\pi_{00,t}) - \log(1 - \pi_{00,t})$, and $\log(\cdot)$ refers to the natural logarithm. At the same time, we set $\psi^* = (\sigma^2, \omega, A, B)$, where ω , A , and B are defined below in equation (12). For the remainder of this paper, we denote the conditional observation density by $p(y_t|f_t, \psi^*, I_{t-1})$.

In the framework of Creal et al. (2013), the dynamic processes for the parameters are driven by information contained in the score of the conditional observation density $p(y_t|f_t, \psi^*, I_{t-1})$ with respect to f_t . The main challenge in the context of Markov switching models is that the conditional observation density is itself a mixture of densities using the latent mixing variable z_t . Therefore, the shape of our conditional observation density as given by equation (3) is somewhat involved.

The updating equation for the time varying parameter f_t based on the score of the predictive density is given by

$$f_{t+1} = \omega + A s_t + B f_t, \quad s_t = S_t \cdot \nabla_t, \quad \nabla_t = \frac{\partial}{\partial f_t} \log p(y_t|f_t, \psi^*, I_{t-1}), \quad (12)$$

where ω is a vector of constants, A and B are coefficient matrices, and s_t is the scaled score of the predictive observation density with respect to f_t using the scaling matrix S_t . The updating equation (12) can be viewed as a steepest ascent or Newton step for f_t using the log conditional density at time t as its criterion function. An interesting choice for S_t , as recognized by Creal et al. (2013), is the square root matrix of the inverse Fisher information matrix. This particular choice of S_t accounts for the curvature of ∇_t as a function of f_t . Also, for this choice of S_t and under correct model specification, the scaled score function s_t has a unit variance; see also Section 4.

3.2 Time varying transition probabilities: the case of 2 states

We first consider the two-state Markov regime switching model, $K = 2$. We let the transition probabilities $\pi_{00,t}$ and $\pi_{11,t}$ vary over time while the two remaining probabilities are set to $\pi_{01,t} = 1 - \pi_{00,t}$ and $\pi_{10,t} = 1 - \pi_{11,t}$ as in (5). We specify the transition probabilities as

$$\pi_{ii,t} = \delta_{ii} + (1 - 2\delta_{ii}) \exp(-f_{ii,t}) / (1 + \exp(-f_{ii,t})), \quad i = 0, 1,$$

where $f_{00,t}$ and $f_{11,t}$ are the only two elements in the time varying parameter vector f_t , and where the two parameters $0 \leq \delta_{ii} \leq 0.5$, for $i = 0, 1$, can be set by the econometrician to limit the range over which $\pi_{ii,t}$ can vary. In the application in Section 6, we set $\delta_{ii} = 0$, for $i = 0, 1$, such that $\pi_{ii,t}$ can take any value in the interval $(0, 1)$.

We prefer to work with a parsimonious model specification and therefore we typically have diagonal matrices for A and B in (12). The updating equations for the time varying

parameter f_t is given by equation (12) where the scaling is set to $S_t = \mathcal{I}_{t-1}^{-0.5}$ where \mathcal{I}_{t-1} is the 2×2 Fisher information matrix corresponding to the 2×1 score vector ∇_t defined in (12). The score vector for the conditional density in (7) takes the form

$$\nabla_t = \frac{p(y_t|\theta_0, \psi^*) - p(y_t|\theta_1, \psi^*)}{p(y_t|\psi^*, I_{t-1})} g(f_t, \psi^*, I_{t-1}), \quad (13)$$

$$g(f_t, \psi^*, I_{t-1}) = \begin{pmatrix} \mathbb{P}[z_{t-1} = 0|\psi^*, I_{t-1}] \cdot (1 - 2\delta_{00})\pi_{00,t}(1 - \pi_{00,t}) \\ -\mathbb{P}[z_{t-1} = 1|\psi^*, I_{t-1}] \cdot (1 - 2\delta_{11})\pi_{11,t}(1 - \pi_{11,t}) \end{pmatrix}. \quad (14)$$

This expression has a highly intuitive form. The first factor in (13) is the difference in the likelihood of y_t given $z_t = 0$ versus $z_t = 1$. The difference is scaled by the total likelihood of the observation given all the static parameters. If the likelihood of y_t given $z_t = 0$ is relatively large compared to that for $z_t = 1$, we expect $f_{00,t}$ to rise and $f_{11,t}$ to decrease. This is precisely what happens in equations (13) and (14). The magnitudes of the steps are determined by the conditional probabilities of being in regime $z_{t-1} = 0$ or $z_{t-1} = 1$, respectively, at time $t - 1$. The remaining factors $(1 - 2\delta_{ii})\pi_{ii,t}(1 - \pi_{ii,t})$, for $i = 0, 1$, are due to the logit parameterization. In particular, if we are almost certain of being in regime $z_{t-1} = 0$ at time $t - 1$, that is $\mathbb{P}[z_{t-1} = 0|\psi^*, I_{t-1}] \approx 1$, then we take a large step with $f_{00,t}$ but we do not move $f_{11,t}$ by much. Obviously, if we are almost certain of being in regime $z_{t-1} = 0$, y_t can only learn us something about $\pi_{00,t}$. We do not learn much about $\pi_{11,t}$ in this case. The converse holds if we are almost certain of being in regime $z_{t-1} = 1$ at time $t - 1$, in which case we can only learn about $f_{11,t} = \text{logit}(\pi_{11,t})$. The weights for the filtered probabilities in the vector $g(f_t, \psi^*, I_{t-1})$ in (13) takes account of this.

The conditional Fisher information matrix based on (13) is singular by design. The vector $g(f_t, \psi^*, I_{t-1})$ on the right-hand side of (13) is I_{t-1} -measurable and hence the expectation of its outer product remains of rank 1. Therefore, we scale the score by a square root Moore-Penrose pseudo-inverse¹ of the conditional Fisher information matrix. We have

$$s_t = G_t \frac{[p(y_t|\theta_0, \psi^*) - p(y_t|\theta_1, \psi^*)] / p(y_t|\psi^*, I_{t-1})}{\sqrt{\int_{-\infty}^{\infty} [p(y_t|\theta_0, \psi^*) - p(y_t|\theta_1, \psi^*)]^2 / p(y_t|\psi^*, I_{t-1}) dy}}, \quad (15)$$

with $G_t = g(f_t, \psi^*, I_{t-1}) / \|g(f_t, \psi^*, I_{t-1})\|$, and where the integral has no closed form in general and is computed numerically, for example using Gauss-Hermite quadrature methods. An alternative to the analytic Moore-Penrose pseudo-inverse is a numerical pseudo inverse; for example, we could use the Tikhonov regularized matrix inverse as given by $\mathcal{I}_{t-1}^{*1/2} = (\lambda \mathbf{I} + (1 - \lambda)\mathcal{I}_{t-1})^{-1/2}$, with unit matrix \mathbf{I} and fixed scalar $0 < \lambda < 1$. For $\lambda \rightarrow 0$ the Tikhonov inverse collapses to the Moore-Penrose pseudo-inverse.

¹If $x \in \mathbb{R}^n$ is a vector, then the Moore-Penrose pseudo-inverse of xx' is given by $\|x\|^{-4}xx'$, and its square root by $\|x\|^{-3}xx'$, as $\|x\|^{-3}xx'\|x\|^{-3}xx' = \|x\|^{-4}xx'$. As $g(f_t, \psi^*, I_{t-1})$ is I_{t-1} -measurable, scaling the score by the square root Moore-Penrose pseudo-inverse of the conditional Fisher information matrix yields an expression proportional to $\|g(f_t, \psi^*, I_{t-1})\|^{-3}g(f_t, \psi^*, I_{t-1})g(f_t, \psi^*, I_{t-1})'g(f_t, \psi^*, I_{t-1}) = \|g(f_t, \psi^*, I_{t-1})\|^{-1}g(f_t, \psi^*, I_{t-1})$.

3.3 Time varying transition probabilities: the case of K states

We can easily generalize the two-regime model to K regimes. To enforce that all transition probabilities are non-negative and sum to one (row-wise), we use the multinomial logit specification. Given a set of values for $0 \leq \delta_{ij} \leq 0.5$, we set

$$\pi_{ij,t} = \delta_{ij} + (1 - 2\delta_{ij}) \exp(f_{ij,t}) \left[1 + \sum_{j=1}^{K-1} \exp(f_{ij,t}) \right]^{-1}, \quad \pi_{i,K-1,t} = 1 - \sum_{j=1}^{K-1} \pi_{ij,t}(\delta_{ij}), \quad (16)$$

for $i = 0, \dots, K-1$ and $j = 0, \dots, K-2$. The time varying parameters $f_{ij,t}$, corresponding to the time varying transition probabilities $\pi_{ij,t}$, are collected in the $K(K-1) \times 1$ vector f_t . The vector f_t is subject to the updating equation (12). The ingredients for the scaled score vector in the updating equation (12) are given by

$$\begin{aligned} \nabla_t &= J_t' \nabla_t^\Pi, & \nabla_t^\Pi &= \frac{\partial \log p(y_t | \psi^*, I_{t-1})}{\partial \text{vec}(\Pi)'} = \frac{\eta_t \otimes \xi_{t-1}}{p(y_t | \psi^*, I_{t-1})}, \\ \mathcal{I}_{t-1} &= \mathbb{E}[J_t' \nabla_t^\Pi \nabla_t^{\Pi'} J_t], \end{aligned}$$

where \otimes is the Kronecker product and the elements of $J_t = \partial \text{vec}(\Pi_t) / \partial f_t'$ are given by

$$\frac{\partial \pi_{ij,t}}{\partial f_{i'j',t}} = \begin{cases} (1 - 2\delta_{ij})\pi_{ij,t}(1 - \pi_{ij,t}), & \text{for } i = i' \wedge j = j', \\ -(1 - 2\delta_{ij})\pi_{ij,t} \pi_{i'j',t}, & \text{for } i = i' \wedge j \neq j', \\ 0, & \text{otherwise.} \end{cases},$$

for $i, i' = 0, \dots, K-1$ and $j, j' = 0, \dots, K-2$.

4 Statistical properties

In this section we study the stochastic properties of the estimated dynamic transition probabilities in our score driven Markov switching model. In particular, we analyze the behavior of the estimated time varying parameter as a function of past observations y_1, \dots, y_{t-1} , parameter vector ψ^* , and initial point \bar{f}_1 . We write the process as $\{\tilde{f}_t\}$ with $\tilde{f}_t := \tilde{f}_t(\psi^*, \bar{f}_1)$, for $t = 1, \dots, T$. We follow Blasques et al. (2012), who use the stationarity and ergodicity (SE) conditions formulated by Bougerol (1993) and Straumann and Mikosch (2006) for general stochastic recurrence equations. Define $\tilde{X}_t = (\tilde{f}_t', \tilde{\xi}_t')$ as the stacked vector of filtered time varying parameters \tilde{f}_t and filtered probabilities $\tilde{\xi}_t$ as defined in (9). We define $\tilde{\xi}_t = \tilde{\xi}_t(\psi^*, \bar{\xi}_1)$ for some initial point $\bar{\xi}_1$. Our stochastic recurrence equation for the filtered process $\{\tilde{X}_t\}$ now takes the form $\tilde{X}_{t+1} = H(\tilde{X}_t, y_t; \psi^*)$, where

$$\tilde{X}_{t+1} = \begin{bmatrix} \tilde{\xi}_{t+1} \\ \tilde{f}_{t+1} \end{bmatrix} = H(\tilde{X}_t, y_t; \psi^*) := \begin{bmatrix} \alpha(\tilde{\xi}_t, \tilde{f}_t, y_t; \psi^*) \\ \omega + A s(\tilde{\xi}_t, \tilde{f}_t, y_t; \psi^*) + B \tilde{f}_t \end{bmatrix},$$

where $s(\tilde{\xi}_t, \tilde{f}_t, y_t; \psi^*)$ is the scaled score defined in (15) and $\alpha(\tilde{\xi}_t, \tilde{f}_t, y_t; \psi^*)$ is the fraction defined in (9) for the recursion $\tilde{\xi}_{t+1} = \alpha(\tilde{f}_t, \tilde{\xi}_t, y_t; \psi^*) := (\Pi' \tilde{\xi}_{t-1}) \odot \eta_t / (\tilde{\xi}_{t-1}' \Pi \eta_t)$. The

following proposition states sufficient conditions for the filtered process $\{\tilde{X}_t(\psi^*, \bar{X}_1)\}$ with initialization at $\bar{X}_1 := (\bar{\xi}_1', \bar{f}_1')$ to converge almost surely and exponentially fast (e.a.s) to a unique limit SE process $\{\tilde{X}_t(\psi^*)\}$.²

PROPOSITION 1. *Let $\{y_t\}$ be SE, with δ_{ij} in (16) satisfying $\delta_{ij} > 0$ for all pairs (i, j) , and assume that for every $\psi^* \in \Psi^*$*

$$(i) \quad \mathbb{E} \log^+ \left\| H(\bar{\xi}_1, \bar{f}_1, y_1, \psi^*) \right\| < \infty;$$

$$(ii) \quad \mathbb{E} \ln \sup_{(f, \xi)} \left\| \dot{H}(f, \xi, y_1, \psi^*) \right\| < 0;$$

where $\dot{H}(X, y_1; \psi^*) = \partial H(X, y_1; \psi^*) / \partial X$ denotes the Jacobian function of H w.r.t. X . Then $\{\tilde{X}_t(\psi^*, \bar{X}_1)\}$ converges e.a.s. to a unique SE process $\{\tilde{X}_t(\psi^*)\}$, for every $\psi^* \in \Psi^*$, that is $\|\tilde{X}_t(\psi^*, \bar{X}_1) - \tilde{X}(\psi^*)\| \xrightarrow{e.a.s.} 0$ as $t \rightarrow \infty \forall \psi^* \in \Psi^*$.

Proof. The assumption that $\{y_t\}$ is SE implies that $\{\eta_t\}$ in (9) is SE. Together with the continuity of H (and the resulting measurability w.r.t. the Borel σ -algebra), it follows that $\{H(\cdot, \cdot, y_t, \psi^*)\}$ is SE for every ψ^* by Krengel (1985, Proposition 4.3). Condition C1 in Bougerol (1993, Theorem 3.1) is immediately implied by assumption (i) for every $\psi^* \in \Psi^*$. Condition C2 in Bougerol (1993, Theorem 3.1) is implied, for every $\psi^* \in \Psi^*$, by condition (ii). As a result, for every $\psi^* \in \Psi^*$, $\{\tilde{X}_t(\psi^*, f_1)\}$ converges almost surely to an SE process $\{\tilde{X}(\psi^*)\}$. Uniqueness and e.a.s. convergence is obtained by Straumann and Mikosch (2006, Theorem 2.8). \square

Proposition 1 effectively defines those combinations of δ , A and B for which we can ensure that the filtered sequence $\{\tilde{X}_t(\psi^*, f_1)\}$ converges e.a.s. to an SE limit for a given SE data sequence $\{y_t\}$. We emphasize that a finite ω is required for condition (i) to hold since ω enters the H function. However, ω plays no role in the contraction condition (ii) as it does not influence the Jacobian \dot{H} . Numerical experiments (not reported here) suggest that stability is achieved by a wide range of combinations of the parameters δ , A and B , where δ is a vector containing all δ_{ij} . In particular, the set of stable combinations (A, B) becomes larger for higher values of δ . This mechanism is intuitive because the entries of the vector δ bound the elements of the vector ξ and Π away from zero and one. As a result, $\delta > 0$ ensures that the denominator in (9) is bounded away from zero and hence the sequence $\{\xi_t\}$ becomes more stable.

Proposition 1 is essential in characterizing the stochastic properties of the filtered time varying parameters. It does not only allow us to have further insights into the nature of the filtered estimates f_t in the Monte Carlo study of Section 5, but it also enables us to interpret the parameter estimates of the GAS model. The SE nature of the filtered sequence is also an important ingredient in obtaining proofs of consistency and asymptotic normality of the maximum likelihood estimator that rely on an application of laws of large numbers and central limit theorems; see Blasques et al. (2014) for more details.

²We say that a random sequence $\{\tilde{X}_t\}$ converges e.a.s. to another random sequence $\{\tilde{X}_t^*\}$ if there is a constant $c > 1$ such that $c^t \|\tilde{X}_t - \tilde{X}_t^*\| \xrightarrow{a.s.} 0$; see Straumann and Mikosch (2006) for further details.

Table 1: Simulation patterns for $\pi_{00,t}$ and $\pi_{11,t}$

	Model	$\pi_{00,t}$	$\pi_{11,t}$
1.	Constant	0.95	0.85
2.	Slow Sine	$0.5 + 0.45 \cos(4\pi t/T)$	$0.5 - 0.45 \cos(4\pi t/T)$
3.	Sine	$0.5 + 0.45 \cos(8\pi t/T)$	$0.5 - 0.45 \cos(8\pi t/T)$
4.	Fast Sine	$0.5 + 0.45 \cos(20\pi t/T)$	$0.5 - 0.45 \cos(20\pi t/T)$
5.	Break	$0.2 \mathbf{1}_{\{T < T/2\}} + 0.8 \mathbf{1}_{\{t \geq T/2\}}$	$0.8 \mathbf{1}_{\{T < T/2\}} + 0.2 \mathbf{1}_{\{t \geq T/2\}}$

5 Monte Carlo study

5.1 Design of the simulation study

To investigate the performance of our estimation procedure for the Markov regime switching model with time varying transition probabilities, we consider a Monte Carlo study for the two regime model (4). The two regimes consist of two normal distributions with common variance $\sigma^2 = 0.5$ and means $\mu_0 = -1$ and $\mu_1 = 1$. We set $\delta_{00} = \delta_{11} = 0$ and consider 5 different forms of time variation for the transition probabilities $\pi_{00,t}$ and $\pi_{11,t}$. The patterns are summarized in Table 1 and range from a constant set of transition probabilities, via slow and fast continuously changing transition probabilities to an incidental structural break in the middle of the sample. We investigate the robustness of our estimation procedure in these different settings. For all of the data generation processes considered, our regime switching model with time varying parameters is clearly misspecified.

In our Monte Carlo study we consider three different sample sizes: $T = 250, 500, 1000$. The number of Monte Carlo replications is set to 100. We adopt the model with $K = 2$ states as described in detail in Sections 3.1 and 3.2. For each data generating process and sample size, we estimate the static parameters ψ^* using the method of maximum likelihood. Given the estimated values for the static parameters, we compute the filtered parameters $\hat{\pi}_{00,t}$ and $\hat{\pi}_{11,t}$ using the updating equations in (12). We compare the results to those for the Markov switching model with time invariant transition probabilities.

5.2 The simulation results

In Table 2, we present Monte Carlo averages of the maximized log-likelihood value, a goodness-of-fit statistic and a forecast precision measure. Our statistic for model fit is the corrected Akaike Information Criterion (AICc). The AICc is the original AIC of Akaike (1973) but with a stronger finite sample penalty; see Hurvich and Tsai (1991). Our measure of forecast precision is the mean squared one-step ahead forecast error (MSFE). The MSFE uses the past observations y_1, \dots, y_{t-1} to make a forecast of y_t , using the static parameter

Table 2: Simulation results I

We have simulated 100 time series from each data generation process (DGP) listed in Table 1 and for sample sizes $T = 250, 500, 1000$. The static parameters are estimated by the method of maximum likelihood, both for the Markov regime switching model with static (Static ψ) and with time varying transition probabilities (TV ψ^*, f_t). In the latter case, the underlying time varying parameters are updated using equation (12). We report the sample averages for the 100 simulated series of the maximized log-likelihood value (LogLik), the corrected Akaike Information Criterion (AICc) and the mean squared error of the one-step ahead forecast of y_t (MSFE). The one-step ahead forecast errors are computed within the sample that is used for parameter estimation.

DGP	T	LogLik		AICc		MSFE	
		Static π	TV ψ^*, f_t	Static ψ	TV ψ^*, f_t	Static ψ	TV ψ^*, f_t
Constant	250	-300.299	-293.037	610.859	604.871	1.695	1.583
	500	-617.392	-611.947	1244.910	1242.271	1.670	1.627
	1000	-1255.957	-1252.046	2521.975	2522.276	1.683	1.668
Slow Sine	250	-355.189	-344.200	720.639	707.196	2.373	2.240
	500	-730.993	-704.058	1472.110	1426.493	2.212	2.100
	1000	-1483.296	-1412.045	2976.652	2842.274	2.219	2.148
Sine	250	-354.946	-346.154	720.153	711.105	2.389	2.237
	500	-732.544	-711.564	1475.213	1441.507	2.243	2.174
	1000	-1485.593	-1426.265	2981.248	2870.714	2.237	2.152
Fast Sine	250	-356.440	-347.479	723.141	713.754	2.602	2.323
	500	-733.108	-720.650	1476.341	1459.678	2.358	2.277
	1000	-1487.584	-1449.217	2985.229	2916.619	2.216	2.212
Break	250	-356.643	-343.721	723.547	706.239	2.438	2.197
	500	-735.293	-706.544	1480.711	1431.467	2.384	2.137
	1000	-1493.767	-1429.812	2997.596	2877.809	2.372	2.104

estimates obtained from the entire sample y_1, \dots, y_T . The criterion therefore measures the forecast precision of the model for the maximum likelihood estimate of ψ in the static model and for the maximum likelihood estimate of ψ^* in the time varying parameter model.

We learn from Table 2 that the average maximized log-likelihood values are uniformly higher for the model with time varying transition probabilities compared to the model with constant transition probabilities. This result is not very surprising since the time varying model is a dynamic generalization of the static model and has more parameters. A more convincing result is that the time varying model produces overall substantially smaller AICc values than the constant model. As expected, the only exception is the data generating process with constant transition probabilities. For time series simulated with time varying transition probabilities change gradually or are of a structural break type, the model with time varying transition probabilities performs substantially better than the static model. Finally, when we focus on the forecast precision measure MSFE, we also conclude that the

Table 3: Simulation results II

We have simulated 100 time series from each data generation process (DGP) listed in Table 1 and for sample sizes $T = 250, 500, 1000$. The static parameters are estimated by the method of maximum likelihood, both for the Markov regime switching model with static (Static) and with time varying transition probabilities (TV). In the latter case, the underlying time varying parameters are updated using equation (12). We report the sample averages for the 100 simulated series of the mean squared error (MSE) and the mean absolute error (MAE) of the one-step ahead forecast of two transition probabilities π_{00} and π_{11} . The one-step ahead forecast errors are computed within the sample that is used for parameter estimation.

DGP	T	MSE		MAE		MSE		MAE	
		π_{00}				π_{11}			
		Static	TV	Static	TV	Static	TV	Static	TV
Constant	250	0.000	0.010	0.015	0.046	0.006	0.037	0.049	0.136
	500	0.000	0.005	0.012	0.032	0.002	0.015	0.032	0.089
	1000	0.000	0.002	0.007	0.019	0.001	0.006	0.021	0.057
Slow Sine	250	0.189	0.087	0.356	0.223	0.197	0.097	0.362	0.238
	500	0.204	0.073	0.369	0.194	0.208	0.085	0.372	0.212
	1000	0.202	0.055	0.367	0.160	0.205	0.055	0.369	0.161
Sine	250	0.184	0.108	0.350	0.255	0.190	0.122	0.355	0.275
	500	0.200	0.084	0.365	0.219	0.204	0.094	0.367	0.234
	1000	0.199	0.065	0.364	0.182	0.201	0.068	0.365	0.190
Fast Sine	250	0.159	0.164	0.329	0.326	0.155	0.167	0.326	0.328
	500	0.182	0.114	0.348	0.265	0.186	0.119	0.350	0.274
	1000	0.199	0.079	0.363	0.216	0.201	0.086	0.365	0.229
Break	250	0.177	0.101	0.341	0.220	0.170	0.094	0.342	0.221
	500	0.175	0.082	0.339	0.181	0.172	0.079	0.340	0.188
	1000	0.174	0.076	0.337	0.162	0.172	0.077	0.338	0.171

Markov switching model with time varying transition probabilities convincingly outperforms its static counterpart.

Next we verify the precision of the filtered transition probability estimates for the static model with π_{00} and π_{11} , and for the time varying model with $\pi_{00,t}$ and $\pi_{11,t}$. In a Monte Carlo study, the transition probabilities are simulated as part of the data generation process. Hence we are able to compare true transition probabilities with their filtered estimates and compute the mean squared error (MSE) and the mean absolute error (MAE) statistics. In Table 3 we present the Monte Carlo averages of these two statistics, for the different data generation processes, the different samples sizes and the two models: the static model (static) and the time varying model (TV).

The results in Table 3 provide strong evidence that the Markov regime switching model with time varying transition probabilities is successful in producing accurate filtered esti-

mates of the probabilities for all different time varying patterns. It is only for the series that are simulated from a model with constant transition probabilities that the MSE and MAE statistics for the static model are smaller than those for the time varying model. For this case, however, the absolute value of all statistics are of an order of magnitude smaller than for all the other data generating processes. For constant transition probabilities, the differences between the static and time varying model can thus be qualified as small. We also observe that for increasing sample sizes T , the MSE and MAE statistics mostly decrease for the time varying model, while this does not occur for the static model. To put these observations in some perspective, we notice that the sinusoid patterns have the same number of swings over the entire sample for different sample sizes. Therefore, the change in the transition probabilities gets smaller per unit of time as T increases. It follows that the updating equation (12) for the time varying model can track the true transition probability more accurately as T increases. This mechanism does not affect the inaccuracy of the estimates obtained from the static model.

6 An empirical study of U.S. Industrial Production

Markov regime switching models are popular in empirical studies of macroeconomic time series. We therefore illustrate our new methodology for time varying transition probabilities in an empirical study concerning a key variable for macroeconomic policy, U.S. Industrial Production (IP). The time series for IP is obtained from the Federal Reserve Bank of St. Louis economic database (FRED); we have monthly seasonally adjusted observations from January 1919 to October 2013, $T = 1137$. We analyze the percentage growth of IP (log-differences $\times 100$) and consider the resulting series as our y_t variable. Figure 1 presents both the IP index and the IP percentage log differences y_t .

6.1 Three model specifications

The time series of IP growth is relatively long. Inspecting the time series plot of y_t , we anticipate that there may be possible changes in both the mean and variance of the series over time. To remain within a parsimonious modelling framework, we adopt the model of Doornik (2013). Doornik analyzes a quarterly time series of post-war U.S. gross domestic product growth by means of a Markov switching mean-variance component model. In our specific setting, we consider a model with *three* regimes for the mean ($m = 0, 1, 2$) and *two* regimes for the variance ($v = 0, 1$). The three regimes for the mean may represent recession, stable and growth periods in U.S. production. Each regime for the mean consists of a constant and p_m lagged dependent variables for y_t , with $m = 0, 1, 2$. The constant and the p_m autoregressive coefficients are collectively subject to the regime to which they belong.

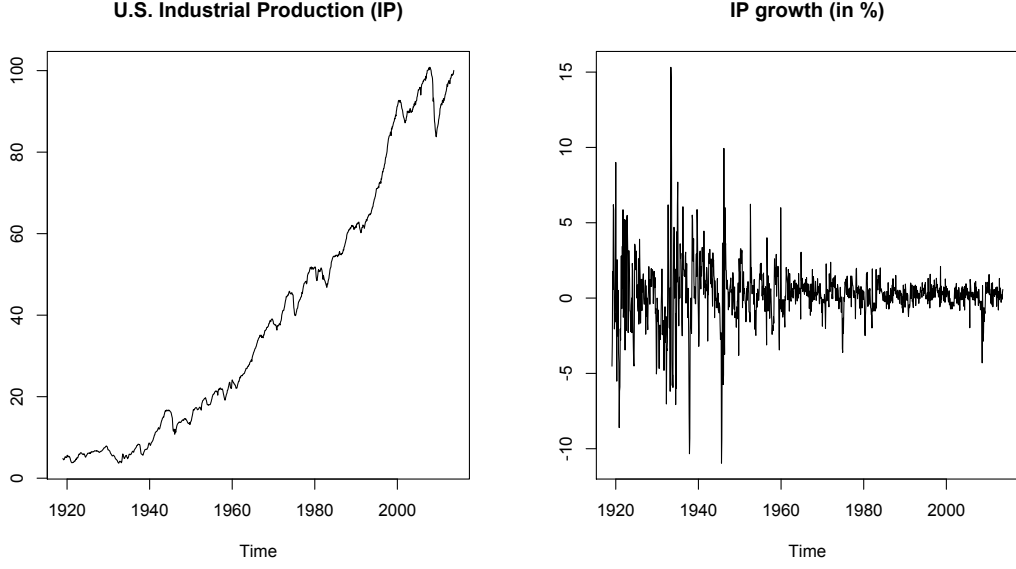


Figure 1: U.S. Industrial Production (monthly, seasonally adjusted) and IP percentage growth (log-differences $\times 100$)

Hence, in case of three regimes for the mean, we have $3 + \sum_{m=1}^3 p_m$ coefficients. The two regimes for the variance may simply distinguish periods of low and high volatility. Hence the number of coefficients for the variance part equals 2.

Model I: static specification

Let $\{z_t^\mu\}$ and $\{z_t^\sigma\}$ denote the hidden processes that determine the mean and the variance for the density of y_t , respectively. We have

$$y_t \mid (z_t^\mu = m, z_t^\sigma = v, I_{t-1}) \sim N(\mu_{m,t}, \sigma_v^2), \quad m = 0, 1, 2, \quad v = 0, 1, \quad (17)$$

with the three mean equations

$$\mu_{m,t} = \phi_{0,m} + \phi_{1,m}y_{t-1} + \dots + \phi_{p_m,m}y_{t-p_m}, \quad m = 0, 1, 2, \quad (18)$$

where $\phi_{0,m}$ is an intercept, $\phi_{1,m}, \dots, \phi_{p_m,m}$ are autoregressive coefficients, for $p_m \in \mathbb{N}^+$, and σ_1^2 and σ_2^2 are the two variances. The transition probabilities for the mean and variance are collected in the matrices Π^μ and Π^σ , respectively, which are given by

$$\Pi^\mu = \begin{pmatrix} \pi_{00}^\mu & \pi_{01}^\mu & 1 - \pi_{00}^\mu - \pi_{01}^\mu \\ \pi_{10}^\mu & \pi_{11}^\mu & 1 - \pi_{10}^\mu - \pi_{11}^\mu \\ \pi_{20}^\mu & \pi_{21}^\mu & 1 - \pi_{20}^\mu - \pi_{21}^\mu \end{pmatrix}, \quad \Pi^\sigma = \begin{pmatrix} \pi_{00}^\sigma & 1 - \pi_{00}^\sigma \\ 1 - \pi_{11}^\sigma & \pi_{11}^\sigma \end{pmatrix}. \quad (19)$$

We follow Doornik (2013) in specifying the transition probability matrix for the $3 \times 2 = 6$ regimes as

$$\Pi = \Pi^\sigma \otimes \Pi^\mu,$$

where the 36 probabilities in Π are a function of 6 mean and 2 variance probabilities. The conditional density of y_t given I_{t-1} can be expressed in terms of the filtered probabilities as in (8). We have

$$\begin{aligned} p(y_t|\psi, I_{t-1}) &= \begin{bmatrix} \mathbb{P}[z_{t-1}^\mu = 0, z_{t-1}^\sigma = 0|\psi, I_{t-1}] \\ \mathbb{P}[z_{t-1}^\mu = 1, z_{t-1}^\sigma = 0|\psi, I_{t-1}] \\ \vdots \\ \mathbb{P}[z_{t-1}^\mu = 2, z_{t-1}^\sigma = 1|\psi, I_{t-1}] \end{bmatrix}' (\Pi^\sigma \otimes \Pi^\mu) \begin{bmatrix} p(y_t; \mu_{0,t}, \sigma_0^2, I_{t-1}) \\ p(y_t; \mu_{1,t}, \sigma_0^2, I_{t-1}) \\ \vdots \\ p(y_t; \mu_{2,t}, \sigma_1^2, I_{t-1}) \end{bmatrix} \\ &= \xi'_{t-1} (\Pi^\sigma \otimes \Pi^\mu) \eta_t. \end{aligned} \quad (20)$$

Next we consider two time varying extensions of the variance regimes in the static model. After experimentation with time varying parameters for the mean regimes, we have concluded that in our empirical setting only the introduction of time varying transition probabilities for the variance regimes can improve the fit of the static model. We notice that once the probabilities in Π^σ are time varying, all probabilities in $\Pi = \Pi^\sigma \otimes \Pi^\mu$ are time varying.

Model II: time varying variance probabilities as a function of $|y_{t-1}|$

In model II, we construct a benchmark in the spirit of Diebold et al. (1994) and Filardo (1994). We specify the time varying transition probabilities for the variance regimes in Π_t^σ as a logistic transformation of the lagged dependent variable, that is

$$\pi_{vv,t}^\sigma = \frac{\exp(g_{vv,t})}{1 + \exp(g_{vv,t})}, \quad g_{vv,t} = c_{0,v} + c_{1,v}|y_{t-1}|, \quad v = 0, 1, \quad (21)$$

where $c_{0,v}$ is an intercept and $c_{1,v}$ is a fixed coefficient, for $v = 0, 1$. The four c coefficients are estimated by the method of maximum likelihood, jointly with the other coefficients.

Model III: time varying variance probabilities as a function of the score

In this specification of time varying transition probabilities for the variance regimes, we adopt the framework of Section 3.2. Empirically, it turns out there is no need to shrink the range of $\pi_{ii,t}$ ex ante. We therefore set $\delta_{ii} = 0$ and specify the time varying matrix Π_t^σ as

$$\pi_{vv,t}^\sigma = \frac{\exp(f_{vv,t})}{1 + \exp(f_{vv,t})}, \quad v = 0, 1, \quad f_t = (f_{00,t}, f_{11,t})',$$

where f_t is updated over time as in (12). The resulting conditional density for y_t is given by $p(y_t|\psi^*, I_{t-1}) = \xi'_{t-1} (\Pi_t^\sigma \otimes \Pi^\mu) \eta_t$. The regime probability structure of Doornik (2013) is

more restricted than the one for the general K -regime Markov switching model in Section 3.3. We therefore have different expressions for the score vector and scaling matrix. The score vector is given by $\nabla_t = (\nabla_{00,t}, \nabla_{11,t})'$ where

$$\nabla_{vv,t} = \frac{\pi_{vv,t}^\sigma(1 - \pi_{vv,t}^\sigma)}{p(y_t|\psi^*, I_{t-1})} \times \xi_{t-1}' \left(\frac{\partial \Pi_t^\sigma}{\partial \pi_{vv,t}^\sigma} \otimes \Pi^\mu \right) \eta_t, \quad v = 0, 1,$$

with ξ_{t-1} and η_t defined implicitly in (20). To compute the conditional Fisher Information $\mathbb{E}[\nabla_t \nabla_t']$, where \mathbb{E} is with respect to $p(y_t|\psi^*, I_{t-1})$, we evaluate 4 numerical integrals, for every time t and for each value for the parameter vector ψ^* , by a Gauss-Hermite method.

6.2 Parameter estimates, model fit and residual diagnostics

Table 4 presents the parameter estimates for the three model specifications. For all models we have taken $p_m = 3$ in (18), with $m = 0, 1, 2$. We have experimented with other values of p_m but this choice provides an adequate fit. The parameter estimates are obtained by numerically maximizing the log-likelihood function with respect to the static parameter vector ψ or ψ^* . The associated standard errors are obtained by numerically inverting the Hessian matrix at the maximized log-likelihood value. The sets of estimated coefficients for the three mean regimes are very similar across the three model specifications. The introduction of time varying variance transition probabilities does not appear to affect the mean specification much. The coefficient $\phi_{0,m}$ determines the interpretation of a regime. We can learn from our estimation results that regime $m = 0$ corresponds to low IP growth, $m = 1$ represents recession and $m = 2$ identifies high growth in IP. The autoregressive coefficients $\phi_{1,m}, \dots, \phi_{3,m}$ show that in "normal years" IP growth is persistent, while during recessions IP growth is subject to persistent cyclical dynamics. Periods of high IP growth are very short lived given the strong negatively estimated autoregressive coefficients for $m = 2$. The estimated transition probabilities reveal the typical situation in regime switching models that once we are in a recession or low growth regime, it is most likely that we remain in this state. It is only for the high growth regime that it is more likely to move to a low growth regime while the probability to stay, $\pi_{22}^\mu = 1 - \pi_{20}^\mu - \pi_{21}^\mu$, is estimated around 0.35.

For the variance regimes, the models clearly distinguish between a low (approximately 0.3) and a high (approximately 5.5 to 6.0) variance regime. The magnitudes of these variances are again comparable across the different models. For Model I, both variance regimes are highly persistent with probabilities π_{00}^σ and π_{11}^σ both estimated close to 1. We also learn from Table 4 that the model fit has improved after introducing time varying variance transition probabilities, both for Model II and Model III. The maximized log-likelihood values have increased by 8 and 17 basis points at the cost of an additional 2 and 4 parameters for the respective Models II and III. The corrected Akaike information criterion clearly points to Model III as the best compromise to model fit and a parsimonious model specification.

The time varying transition probabilities for the low volatility ($v = 0$) and high volatility ($v = 1$) regimes in Model III are highly persistent processes over time: the estimates for the diagonal elements of B are very close to unity. From Model II we may conclude that only the transition probability of a low volatility regime is time varying: the estimate of $c_{1,1}$ is not significant. For Model III we find somewhat stronger evidence that both transition probabilities are time varying: the estimates of both diagonal elements of A are significant at the usual level of 5%.

Finally, Table 4 presents diagnostic test statistics for the generalized and Rosenblatt residuals which we have discussed in Section 2. The p -values for the well-known Jarque-Bera χ^2 normality test and the Ljung-Box χ^2 serial correlation test, for the residuals and the squared residuals, indicate that all models are capable of describing the salient features in IP growth. There are some differences between the statistics for the generalized and Rosenblatt residuals, but they are small and have no bearings on the main conclusions. The Jarque-Bera test may indicate that the IP growth time series is subject to a few outlying observations.

6.3 Signal extraction: regime and transition probabilities

In Figure 2 we present the smoothed estimates of probabilities for mean and variance regimes and the filtered estimates of (time varying) transition probabilities for Models I, II and III. Model II appears unable to capture the dynamics in the transition probabilities. We have learned from Table 4 that the estimate of coefficient $c_{1,1}$ is not significant; it is also reflected in Figure 2 with the time series plot of the filtered probability estimates for the high variance regime that is almost constant over time. On the other hand, the filtered probability estimates for the low volatility regime are highly erratic.

The filtered probabilities for Model III show an entirely different pattern. Both the low and high volatility transition probabilities evolve gradually over time. In particular, the persistence of the low volatility regime appears to have gone up over time, with values around 0.7 in the early part of the sample, and values close to 1 in the second half of the sample. The converse holds for the high volatility regime. The persistence probability π_{11}^σ is close to 1 up to the 1940s. After that, the probability decreases substantially to values around 0.5, and slowly rises towards the end of the sample again. The pattern for the filtered probabilities is consistent with the empirical pattern in the data in Figure 1. In the earlier part of the sample, high volatility levels are predominant. Towards the middle of the sample, large volatilities are incidental and short-lived, whereas towards the end of the sample during the years of the financial crisis, U.S. debt ceiling crisis, and the European sovereign debt crisis, higher volatility levels appear to cluster again more.

The empirical patterns are also corroborated by the parameter estimates in Table 4. In particular, the parameter estimates for the diagonal elements of B are both close to 1; it suggests that the dynamic transition probabilities evolve gradually over time. The estimates

Table 4: Parameter estimates, model fit and residual diagnostics

In the first two panels we report the maximum likelihood estimates with standard errors in parantheses below, for Models I, II and III. In the first panel the parameter estimates for the mean $\mu_{m,t}$ in (17) are reported for each regime $m = 0, 1, 2$: the intercept $\phi_{0,m}$, the autoregressive coefficients $\phi_{1,m}, \dots, \phi_{3,m}$, and the transition probabilities π_{mj}^μ , for $j = 0, 1$, in Π^μ of (19). In the second panel the two regime variance estimates for σ_v^2 are reported. The variance transition probability for Model I π_{vv}^σ is estimated directly while we have $\pi_{vv}^\sigma = \text{logit}^{-1}(x_v)$ for Model II ($x_v = c_{0,v}$) and for Model III ($x_v = \omega_v/(1 - B_{vv})$), for $v = 0, 1$, where ω_v and B_{vv} are the $(v + 1)^{\text{th}}$ elements of vector ω and diagonal matrix B in (12), respectively. The time varying variance probabilities are determined in Model II by $c_{1,v}$, and in Model III by A_{vv} and B_{vv} which are the $(v + 1)^{\text{th}}$ diagonal elements of A and B in (12), respectively, for $v = 0, 1$. In the third panel we report model fit statistics: Fit(1) is number of static parameters; Fit(2) is maximized log-likelihood value; Fit(3) is AICc, see Section 5.2. We further report the p -values of the residual diagnostic (RD) test statistics for the generalized (e_t) and Rosenblatt's residuals (\tilde{e}_t): RD(1) is Jarque-Bera normality $\chi^2(2)$ test; RD(2) is Ljung-Box serial correlation $\chi^2(6)$ test; RD(3) is as RD(2) for squared residuals.

	Model I			Model II			Model III		
	$m = 0$	$m = 1$	$m = 2$	$m = 0$	$m = 1$	$m = 2$	$m = 0$	$m = 1$	$m = 2$
$\phi_{0,m}$	0.076 (0.038)	-0.212 (0.182)	0.846 (0.168)	0.068 (0.037)	-0.277 (0.140)	0.837 (0.172)	0.043 (0.033)	-0.128 (0.095)	0.737 (0.128)
$\phi_{1,m}$	0.316 (0.050)	1.121 (0.096)	-0.609 (0.135)	0.327 (0.050)	1.126 (0.088)	-0.621 (0.130)	0.351 (0.040)	1.087 (0.079)	-0.479 (0.107)
$\phi_{2,m}$	0.212 (0.050)	-0.569 (0.146)	-0.395 (0.123)	0.220 (0.040)	-0.526 (0.139)	-0.484 (0.086)	0.234 (0.037)	-0.537 (0.108)	-0.221 (0.086)
$\phi_{3,m}$	0.105 (0.037)	0.076 (0.107)	0.039 (0.124)	0.113 (0.035)	0.011 (0.115)	0.133 (0.101)	0.112 (0.030)	0.036 (0.065)	0.103 (0.092)
π_{m0}^μ	0.909 (0.041)	0.111 (0.103)	0.577 (0.140)	0.909 (0.032)	0.113 (0.072)	0.592 (0.142)	0.864 (0.036)	0.145 (0.070)	0.576 (0.175)
π_{m1}^μ	0.016 (0.019)	0.858 (0.092)	0.055 (0.052)	0.014 (0.012)	0.858 (0.064)	0.073 (0.062)	0.021 (0.015)	0.842 (0.059)	0.048 (0.055)
σ_v^2	$v = 0$ 0.336 (0.025)	$v = 1$ 5.579 (0.629)		$v = 0$ 0.351 (0.026)	$v = 1$ 5.866 (0.691)		$v = 0$ 0.317 (0.023)	$v = 1$ 5.920 (0.541)	
π_{vv}^σ	0.980 (0.007)	0.947 (0.018)		0.996 (0.003)	0.883 (0.053)		0.886 (0.080)	0.702 (0.200)	
$c_{1,v}$				-1.899 (0.419)	0.108 (0.209)				
A_{vv}							0.132 (0.058)	0.148 (0.074)	
B_{vv}							0.998 (0.003)	0.989 (0.011)	
	$i = 1$	$i = 2$	$i = 3$	$i = 1$	$i = 2$	$i = 3$	$i = 1$	$i = 2$	$i = 3$
Fit(i)	22	-1642	3330	24	-1634	3317	26	-1625	3302
RD(i) e_t	0.065	0.772	0.556	0.032	0.968	0.659	0.019	0.792	0.924
RD(i) \tilde{e}_t	0.011	0.409	0.648	0.012	0.830	0.738	0.007	0.632	0.676

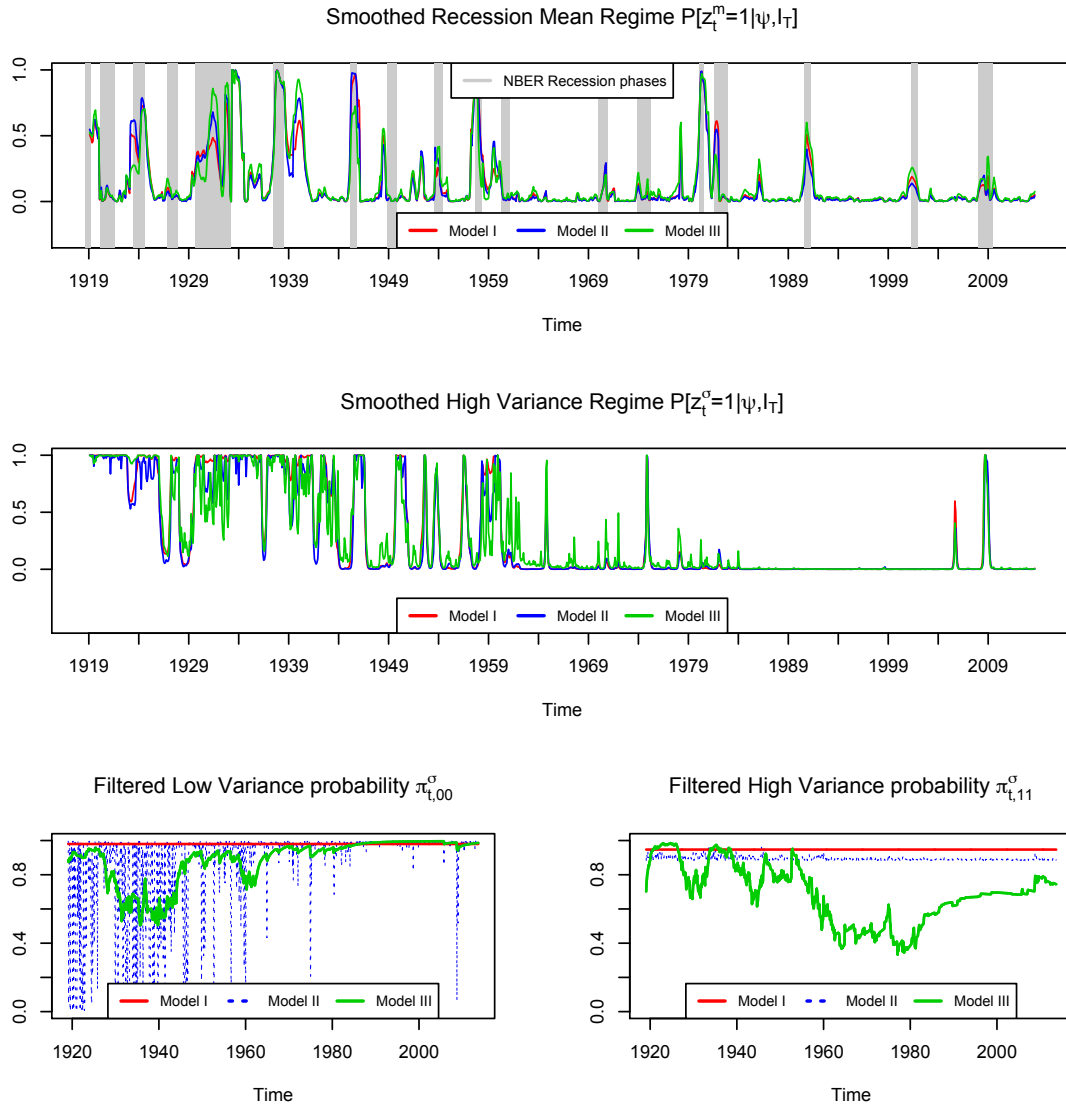


Figure 2: Smoothed probability estimates for the recession regime in the mean and for the high variance regime. Filtered transition probability estimates for the low and high variance regimes. In the first graph, the vertical gray areas indicate recessions according to the NBER business cycle classifications.

of both diagonal elements of A have the correct sign and lead to parameter changes that increase the local fit of the model in terms of log-likelihood.

Finally, we present the smoothed estimates of z_t in the top panels of Figure 2, together with the NBER business cycle classifications. We may conclude that all models result in higher smoothed recession probabilities in the NBER classified periods. The model fit for a model with time varying transition probabilities for the variance regimes (Model II or III) is typically higher than the static Model I. From the smoothed probabilities for the high variance regime, most of the high variance regime is located in the first half of the sample. The second episode of high variance is during the financial crisis, with the intermediate period having predominantly a low level of volatility. We notice that some, but not all, NBER recessions correspond to periods of high volatility. This supports the use of our current framework with separate regimes for the (conditional) means and for the variances.

7 Conclusion

We have introduced a new methodology for time varying transition probabilities in Markov switching models. We have shown that the use of the score of the predictive likelihood and the generalized autoregressive score (GAS) modelling framework of Creal et al. (2013) can drive the dynamics of the transition probabilities effectively over time. The corresponding dynamics can easily be interpreted while the information embedded in the conditional observation densities are fully incorporated. We have formulated conditions for the estimated time varying probabilities from our score driven model to converge to stationary and ergodic stochastic processes.

By means of an extensive Monte Carlo study, we have shown that the our proposed observation driven model is able to adequately track the dynamic patterns in transition probabilities, even if the underlying dynamics themselves are possibly misspecified. Both for deterministic structural breaks and deterministic sinusoid patterns, our model yields a large improvement in model fit compared to a model with constant transition probabilities only.

In our empirical study for Industrial Production growth, we have shown that we can effectively use the model dynamic features in the mean and variance simultaneously. We have found that our proposed model outperforms both the Markov switching model with constant probabilities and with transition probabilities depending on a lagged dependent variable. In particular, the patterns filtered by our model can be easily interpreted, with higher (lower) persistence for high (low) volatility regimes in the beginning of the sample compared to the later part of the sample. Higher volatilities appear to re-occur again at the very end of the sample, during the financial and sovereign debt crises. We conclude that the model can provide a useful benchmark in settings where transition probabilities in a regime switching model may vary over time.

References

- Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* 60(2), 255–265.
- Blasques, F., S. J. Koopman, and A. Lucas (2012). Stationarity and ergodicity of univariate generalized autoregressive score processes. *Discussion Paper Tinbergen Institute TI 12-059/4*.
- Blasques, F., S. J. Koopman, and A. Lucas (2014). Maximum likelihood estimation for generalized autoregressive score models. *Discussion Paper Tinbergen Institute TI 14-029/III*.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 31(3), 307–327.
- Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization* 31(4), 942–959.
- Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8, 93–115.
- Creal, D., S. J. Koopman, and A. Lucas (2008). A general framework for observation driven time-varying parameter models. *Discussion Paper Tinbergen Institute TI 08-108/4*.
- Creal, D., S. J. Koopman, and A. Lucas (2011). A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations. *Journal of Business & Economic Statistics* 29(4), 552–563.
- Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics* 28(5), 777–795.
- Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation driven mixed-measurement dynamic factor models. *Review of Economics and Statistics*, forthcoming.
- De Lira Salvatierra, I. and A. J. Patton (2013). Dynamic copula models and high frequency data. *Duke University Discussion Paper*.
- Delle Monache, D. and I. Petrella (2014). A score driven approach for gaussian state-space models with time-varying parameter. *Working Paper, Imperial College London*.
- Diebold, F., J. Lee, and G. Weinbach (1994). Regime Switching with Time-Varying Transition Probabilities. In C. Hargreaves (Ed.), *Nonstationary Time Series Analysis and Cointegration*, pp. 283–302. Oxford University Press.

- Doornik, J. (2013). A Markov-switching model with component structure for US GNP. *Economics Letters* 118(2), 265–268.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.
- Engle, R. F. and J. R. Russell (1998). Autoregressive Conditional Duration: A New Model for irregularly Spaced Transaction Data. *Econometrica* 66(5), 1127–1162.
- Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics* 12(3), 299–308.
- Francq, C. and M. Roussignol (1998). Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum-likelihood estimator. *Statistics: A Journal of Theoretical and Applied Statistics* 32(2), 151–173.
- Francq, C. and J.-M. Zakoïan (2001). Stationarity of multivariate markov-switching ARMA models. *Journal of Econometrics* 102(2), 339–364.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.
- Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987). Generalised residuals. *Journal of Econometrics* 34(1), 5–32.
- Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42(1), 27–62.
- Hamilton, J. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57(2), 357–384.
- Hamilton, J. D. and B. Raj (2002). New directions in business cycle research and financial analysis. *Empirical Economics* 27(2), 149–162.
- Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Econometric Series Monographs. Cambridge University Press.
- Harvey, A. C. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, forthcoming.
- Hurvich, C. M. and C.-L. Tsai (1991). Bias of the corrected AIC criterion for underfitted regression and time series models. *Biometrika* 78(3), 499–509.
- Kim, C. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics* 60(1), 1–22.

- Kim, C.-J., J. C. Morley, and C. R. Nelson (2004). Is there a positive relationship between stock market volatility and the equity premium? *Journal of Money, Credit and Banking* 36, 339–360.
- Koopman, S. J., A. Lucas, and M. Scharth (2012). Predicting time-varying parameters with parameter-driven and observation-driven models. *Tinbergen Institute Discussion Papers* 12-020/4.
- Krengel, U. (1985). *Ergodic theorems*. Berlin: De Gruyter studies in Mathematics.
- Lucas, A., B. Schwaab, and X. Zhang (2014). Measuring credit risk in a large banking system: econometric modeling and empirics. *Journal of Business and Economic Statistics*, forthcoming.
- Maheu, J. M. and T. H. McCurdy (2000). Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics* 18(1), 100–112.
- Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59(2), 347–370.
- Oh, D. H. and A. J. Patton (2013). Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. *Duke University Discussion Paper*.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *The Annals of Mathematical Statistics* 23(3), 470–472.
- Smith, D. R. (2008). Evaluating Specification Tests for Markov-Switching Time-Series Models. *Journal of Time Series Analysis* 29(4), 629–652.
- Straumann, D. and T. Mikosch (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics* 34(5), 2449–2495.
- Turner, C. M., R. Startz, and C. R. Nelson (1989). A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics* 25(1), 3–22.