Stochastic levels and duration dependence in US unemployment

Bert de Bruijn[a,b,*]
Philip Hans Franses[a,b,†]

[a] Econometric Institute, Erasmus School of Economics,
Erasmus University Rotterdam, the Netherlands

[b] Tinbergen Institute, the Netherlands

September 23, 2015

Econometric Institute Report Series, EI2015-20

Abstract

We introduce a new time series model that can capture the properties of data as is typically exemplified by monthly US unemployment data. These data show the familiar nonlinear features, with steeper increases in unemployment during economic downswings than the decreases during economic prosperity. At the same time, the levels of unemployment in each of the two states do not seem fixed, nor are the transition periods abrupt. Finally, our model should generate out-of-sample forecasts that mimic the in-sample properties. We demonstrate that our new and flexible model covers all those features, and our illustration to monthly US unemployment data shows its merits, both in and out of sample.

Keywords: Markov switching, Duration Dependence, Gibbs sampling, Unemployment, Stochastic levels.

JEL classifications: C11, C22, C24, C53, E24.

*Econometric Institute, Erasmus School of Economics, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands; phone: +3110 4088902; email: debruijn@ese.eur.nl

†Econometric Institute, Erasmus School of Economics, PO Box 1738, NL-3000 DR Rotterdam, The Netherlands; phone: +3110 4081273; fax: +3110 4089162; email: franses@ese.eur.nl (corresponding author)

1 Introduction

In this paper we introduce a new time series model that can capture the properties of data as exemplified by monthly US unemployment data as depicted in Figure 1. Clearly the data show nonlinear features, as the increases in unemployment during economic downswings are much steeper than the decreases during economic prosperity. At the same time, the levels of unemployment in each of the two states do not seem fixed, nor are the transition periods abrupt. Finally, one would want a time series model that can generate out-of-sample forecasts that mimic the in-sample properties. Our new and flexible model will be shown to cover just those aspects, and our illustration to the very same US unemployment data shows its merits.

To analyze time series data with regime switching features, a natural starting point is the familiar Markov Switching model. Markov Switching (MS) models (Hamilton, 1989) are suitable for data fluctuating around two levels, where these levels associate with each of the two states. In the initially proposed MS models, the occurrence of a state at time t is only dependent on the state at time $t-1$ and it is governed by transition probabilities p_{ij}, the probability of switching from state i to state j. In these initial models, the probabilities are fixed across the sample. Figure 2a shows a simulated example of a MS model with 2 states. Recent applications of such basic MS models include Kim (2009), Bauwens et al. (2010), Nalewaik (2011), Cunningham and Kolet (2011), Guérin and Marcellino (2013) and Chen and Schorfheide (2013).

One of the features of such a basic MS model is that it is not capable of dealing with cyclicality, which entails for example that the forecasts produced by a basic two-state MS model are monotonically convergent. One modification to account for cyclical behavior is to introduce duration dependence in the transition probabilities (Diebold and Rudebusch, 1990; Durland and McCurdy, 1994). For example, one could let the probabilities be $p_{ij} = F(\beta_0 + \beta_1 d_t)$ with F any CDF and d_t the duration of the current spell at time t, which is the period since the last state switch. This means that the probability of switching now has become dependent on the duration of the spell. Figure 2b shows an example of a duration dependent MS model with

2 states, in which the duration dependence is positive, that is, the probability of switching out of a state increases the longer the time series has been within that state. Such MS models are implemented in, among others, Sichel (1991), Lunde and Timmermann (2004), Lam (2004), Layton and Smith (2007), Castro (2010) and Cunningham and Kolet (2011).

A common feature of the two MS models so far is that the mean in each of the states is fixed. For example, if one were to use a two-state MS model to model unemployment from 1980 to now, one assumes that the states of high unemployment and low unemployment imply the same mean for both the eighties and the current decade. This assumption might not be considered as realistic, which is also clear from Figure 1. Hence, one may wish to allow the means to be stochastic. In our model we alleviate the restriction by allowing the means to alternate in such a way that the difference is different each time. An example of the kind of data that can be generated by such a model can be found in Figure 2c.

Finally, as already indicated, and is visible from Figure 1, the transitions from one state to the other may not be immediate, as there might be a gradual transition from the previous state mean to the new state mean. At the same time, the time it takes to switch from one regime to the other may also not be the same across the entire sample, and hence we wish to allow the transition process to be a stochastic process too. Figure 2d shows a simulated time series with these properties, and it is clear that the pattern starts to come close to the unemployment data graphed in Figure 1.

To wrap up, in the present paper we propose a Markov Switching model with duration dependence, and with stochastic processes for the levels in each of the states and for the transitions from one regime to the other. We will illustrate our new model for monthly US unemployment from 1948 to 2012 (see Figure 1). As this new model is computationally demanding and also requires the data to be informative, we run various simulations to see how well parameters can be estimated.

The outline of the remainder of this paper is as follows. In Section 2 we will formally introduce our new MS model. Using simulations, we will highlight some of the data characteristics that align with this model. Section 3 discusses estimation

of the parameters and inference of the latent variables. We will also show how one can produce forecasts using this model, and we will demonstrate that these forecasts continue the in-sample data features into the future. In Section 4 we illustrate our model and the associated estimation procedure for US unemployment. We also outline how our new model can be used for real-time monitoring of the data. In Section 5 we simulate from the DGP using the estimates from Section 4 to investigate how accurate the parameters can be estimated. Finally, we conclude in Section 6 with some final remarks and thoughts for further research.

2 Modeling and simulations

We first reintroduce the Markov Switching model. We choose for a notation that will make it easier to describe extensions. Denote the time series of interest as y_t with $t = 1, \ldots, T$. We relate y_t to the two state means μ_0 and μ_1, and the differences between the data and the state means are contained in the error term $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. In the basic MS model the probability of being in one of the states $s_t \in 0, 1$ depends only on the state in the previous time period, and in the basic MS model these probabilities are assumed as fixed. The so far discussed properties of the model can be captured by the following expressions:

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{1}$$

$$\mu_t = \mu_{s_t} \tag{2}$$

$$P(s_t = j | s_{t-1} = i) = p_{ij}, \quad i, j \in 0, 1 \tag{3}$$

The time period of the κ^{th} switch is described by the variable τ_κ, so $s_{\tau_\kappa} \neq s_{\tau_\kappa - 1}\ \forall \kappa$ and $s_t = s_{t-1}$ for all other t.

A first extension of this model that is often considered in practice concerns allowing for transition probabilities that are duration dependent. One way to do this is to use a link function to transform a linear function of the state duration to a variable between 0 and 1. One possible link function that is commonly used in various applications is the standard normal CDF $\Phi(.)$. If the duration of the relevant state at time t is captured by the variable d_t, then a first extension amounts to replacing (3)

by $P(s_t = j|s_{t-1} = i) = \Phi(\beta_0 + \beta_1 d_t)$. This makes the transition probability dependent on the duration d_t, but note that it also assumes that switches from one state to the other occur in a similar way. One way to incorporate a possible difference in such switching behavior is to consider

$$\begin{array}{rcl} P(s_t = j|s_{t-1} = i) & = & \Phi(\beta_0 + \beta_1 I[s_{t-1} = 1] + \beta_2 d_t + \beta_3 d_t I[s_{t-1} = 1]) \\ & = & \Phi(\beta' D_t) \end{array} \tag{4}$$

with $I[.]$ the indicator function and $D_t' = [1 \quad I[s_{t-1} = 1] \quad d_t \quad d_t I[s_{t-1} = 1]]$. If both β_1 and β_3 are zero, then there is no difference in the switching behavior across the two states. If both β_2 and β_3 are zero, there is no duration dependence, and the model reduces to the basic MS model as in equations (1)-(3).

In practice we need to estimate the value of d_t when $t = 1$ as we do not know whether a switch has occurred just before the start of the sample or whether it has occurred a long time before that. For this, we introduce the variable d_1^*, and we will set d_1 equal to that, and calculate the other d_t's by either adding 1 to the previous value, or by resetting it to 1.

Next, we propose a second extension by allowing for stochastic state means, instead of fixing these to two values μ_0 and μ_1. To allow for this in the notation, we introduce the difference between the type of state s_t, which is either H (high) or L (low), and the sequential number of the state at time t, for which we extend the variable κ to have an index $\kappa_t = 0, 1, \ldots$. The state type will switch around each time the data enter a new state. We assume the following relation between two subsequent states means, that is

$$\mu_{\kappa_t} \sim N(\mu_{\kappa_t - 1} + \Delta\mu^* \times (-1)^{I[s_t = L]}, \sigma^2_{\Delta\mu}) \tag{5}$$

This relation assumes that the new state mean on average differs $\Delta\mu^*$ from the previous state mean. This difference is however not fixed, so it is not exactly the same each time. Also, whether the change in the state mean is upwards or downwards depends on what type of state s_t will be associated with the new state mean μ_{κ_t}. We do not want new state means to be on the wrong side of the previous state mean (for example, having a state mean of type $s_t = H$ being lower than the directly preceding state mean of the low type). Therefore, we adjust the preceding relation

to a Truncated Normal distribution with parameters for the bounds denoted by lb_{κ_t} and ub_{κ_t} as

$$\mu_{\kappa_t} \sim TN(\mu_{\kappa_t - 1} + \Delta\mu^* \times (-1)^{I[s_t = L]}, \sigma^2_{\Delta\mu}, lb_{\kappa_t}, ub_{\kappa_t}) \tag{6}$$

$$lb_{\kappa_t} = \begin{cases} -\infty & \text{if } s_t = L \\ \mu_{\kappa_t - 1} & \text{if } s_t = H \end{cases} \qquad ub_{\kappa_t} = \begin{cases} \mu_{\kappa_t - 1} & \text{if } s_t = L \\ \infty & \text{if } s_t = H \end{cases} \tag{7}$$

The observation mean μ_t can now be generalized from (2) to

$$\mu_t = \mu_{\kappa_t} \tag{8}$$

Our third and final extension concerns the stochastic linear transitions. We introduce the notation λ_{κ_t} to denote the time taken by the transition of state mean $\mu_{\kappa_t - 1}$ to μ_{κ_t}. The linear transition property indicates that between the start of the transition τ_{κ_t} and the end of the transition at $\tau_{\kappa_t} + \lambda_{\kappa_t}$, the state mean is not equal to either $\mu_{\kappa_t - 1}$ or μ_{κ_t}, but a weighted sum of these, with weights dependent on the length of the transition period. To calculate weights, we can use the duration variable d_t. This results in partly replacing (8) by

$$\mu_t = \frac{d_t}{\lambda_{\kappa_t}} \mu_{\kappa_t} + (1 - \frac{d_t}{\lambda_{\kappa_t}}) \mu_{\kappa_t - 1} \tag{9}$$

We write "partly" because this replacement is only relevant for the cases in which $t \in [\tau_{\kappa_t}; \tau_{\kappa_t} + \lambda_{\kappa_t}]$. For the other cases, (8) remains valid, that is, $\mu_t = \mu_{\kappa_t}$ if $t \in [\tau_{\kappa_t} + \lambda_{\kappa_t}; \tau_{\kappa_t + 1}]$. We impose a distribution on λ_{κ_t} that needs to be positive, which is why we use the lognormal distribution. We allow the precise distribution to be dependent on whether the switch is upwards or downwards, as there might be a difference in transition speed. For the upwards switch, we propose $\lambda^u_{\kappa_t} \sim LN(\lambda^*_u, \sigma^2_{\lambda,u})$. Similarly for the downwards switch, we assume $\lambda^d_{\kappa_t} \sim LN(\lambda^*_d, \sigma^2_{\lambda,d})$. Finally, we impose that the transition periods have come to an end before the next one starts. The latter amounts to the restriction

$$\tau_{\kappa_t + 1} - \tau_{\kappa_t} \geq \lambda_{\kappa_t} \tag{10}$$

for all κ_t.

To wrap up, our new model reads as

$$y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{11}$$

$$\mu_t = \begin{cases} \mu_{s_t} & \text{if } t \in [\tau_{\kappa_t} + \lambda_{\kappa_t}; \tau_{\kappa_t+1}] \\ \frac{d_t}{\lambda_{\kappa_t}}\mu_{\kappa_t} + (1 - \frac{d_t}{\lambda_{\kappa_t}})\mu_{\kappa_t-1} & \text{if } t \in [\tau_{\kappa_t}; \tau_{\kappa_t} + \lambda_{\kappa_t}] \end{cases} \tag{12}$$

$$P(s_t = j | s_{t-1} = i) = \Phi(\beta D_t) \tag{13}$$

$$D_t' = [1 \quad I[s_{t-1} = L] \quad d_t \quad d_t I[s_{t-1} = L]] \tag{14}$$

$$\mu_{\kappa_t} \sim TN(\mu_{\kappa_t-1} + \Delta\mu^* \times (-1)^{I[s_t=L]}, \sigma_{\Delta\mu}^2, lb_{\kappa_t}, ub_{\kappa_t}) \tag{15}$$

$$lb_{\kappa_t} = \begin{cases} -\infty & \text{if } s_t = L \\ \mu_{\kappa_t-1} & \text{if } s_t = H \end{cases} \qquad ub_{\kappa_t} = \begin{cases} \mu_{\kappa_t-1} & \text{if } s_t = L \\ \infty & \text{if } s_t = H \end{cases} \tag{16}$$

$$\lambda_{\kappa_t}^u \sim LN(\lambda_u^*, \sigma_{\lambda,u}^2) \qquad \lambda_{\kappa_t}^d \sim LN(\lambda_d^*, \sigma_{\lambda,d}^2) \tag{17}$$

$$\tau_{\kappa_t+1} - \tau_{\kappa_t} \geq \lambda_{\kappa_t} \tag{18}$$

Our new model includes 12 parameters that need to be estimated from the data, and these are σ_ε, $\Delta\mu^*$, $\sigma_{\Delta\mu}$, λ_u^*, $\sigma_{\lambda,u}$, λ_d^*, $\sigma_{\lambda,d}$, β_0, β_1, β_2, β_3, d_1^*. Additionally, we need to estimate $3\kappa_T + 1$ latent variables, that is $\mu_{\kappa_t}, \tau_{\kappa_t}, \lambda_{\kappa_t}$ $\forall \kappa_t$, plus the start state mean μ_0. The number of latent variables depends on the size of the sample and on the frequency of state switches. The other variables such as the observation mean μ_t or the state duration d_t can be directly calculated from the estimates of the latent variables.

Hypothetical data

To examine how time series data can look like if they are generated from the new model, we run a few simulations. We generate data from four data generating processes (DGPs). The reference DGP, for which the associated hypothetical data are plotted in the top left part of each upcoming graph, is based on the following parameter configuration, that is, $\sigma_\varepsilon = 1$, $\mu_0 = 0$, $\Delta\mu^* = 6$, $\sigma_{\Delta\mu} = 2$, $\lambda_u^* = \lambda_d^* = 2.5$, $\sigma_{\lambda,u} = \sigma_{\lambda,d} = 0.5$, $\beta = [-4 \; 0 \; 0.1 \; 0]'$ and $d_1^* = 0$, where we set the sample size at $T = 500$. This corresponds with a duration-dependent model in which the transition behavior is the same for upward and downward switches. The other DGPs differ from the benchmark DGP each time for only a few parameters, that is, we consider (i) $\sigma_{\Delta\mu} = 1$, which amounts to a process with more similar-sized jumps between the state means, and thus this process is closer to a model without stochastic means.

Next, we consider (ii) $\beta = [-3\ -3\ 0.12; 0]'$, for which the most important difference is $\beta_1 = -3$ instead of $\beta_1 = 0$. The intercept for downward switches is $\beta_0 + \beta_1$, thus a negative value for β_1 makes downward transitions take more time to initialize than upward transitions, which have just β_0 as intercept. Finally, we consider (iii) $\lambda_d^* = 4$, which means that the downward transitions take more time to complete than the upward transitions.

Figure 3 shows several simulated series using each of the four configurations of the parameters. The top-left graph shows three series generated using the reference DGP, with two series having a different μ_0 which ensures that these lines do not overlap. The state durations are relatively stable, which illustrates the duration dependence. Also, the stochastic means are evident from the fact that the level is not the same each time the data switch between states. Especially the top line shows a quickly changing mean.

The other graphs in Figure 3 show comparable but slightly different behavior. The top-right panel shows alternative (i), which incorporates a lower $\sigma_{\Delta\mu}$. This is visible as this series shows less drifting behavior, and stays closer to its starting value. Alternative (ii) in the lower-left graph depicts the case in which the duration dependence is different per state. The graph clearly shows that more time is spent in the high states than in the low states. The lower-right graph represents alternative (iii), in which the transition time is different across the types of switches. The series in this graph clearly have a longer upwards transition time than downwards.

Figure 4 shows the first half of one of the series for each type, and then uses 10000 simulations to construct a prediction interval for the remainder of the series. The top-left graph shows how our new general models short-term forecasts can capture the cyclical behavior rather well. Of course, for the longer term one eventually becomes less certain about whether there will be an upward or a downward state. Also, the confidence intervals increase as the aggregated effect of the unknown future information increases. The smaller $\sigma_{\Delta\mu}$ in alternative (i) is clearly visible in the top-right graph of Figure 4, as the intervals are smaller, especially for the longer term where the aggregated effect of $\sigma_{\Delta\mu}$ could have much of an effect. The intervals for both alternatives (ii) and (iii) seem to be comparable in size for all horizons relative

to the reference DGP. This shows that the estimation of the stochastic mean shall be the most important part of the estimation process as this mean dominates the uncertainty in long-term forecasts.

Finally, Figure 5 shows a simulated histogram of the length of a full cycle (switching up and down) for each parameter configuration, based on 50000 replications. While the first three histograms (with symmetric switching behavior between both types of states) have an approximately symmetrically distributed cycle length, the last histogram shows a much more asymmetric distribution. This shows that an asymmetric switching occurrence (alternative (ii)) and an asymmetric transition length (alternative (iii)) can have different effects on the cycle length.

3 Estimation and inference

In this section we present the estimation routine for the estimation of the parameters and latent variables in our new general model. We also show how these estimates can be used for forecasting purposes.

Parameter estimation

We start with assuming that the number of state switches in the sample is known and is equal to K. To estimate the parameters we will make use of Gibbs Sampling with Data Augmentation. This method uses conditional distributions of parameters and latent variables given other parameters and latent variables to draw parameter values in an iterative manner. If chosen starting values of the parameters and latent variables are reasonably close to their posterior distribution, then after convergence the draws will be draws from the posterior distribution of the parameters. From these, one can take for example the mean to obtain a point estimate. We denote the draw in iteration m with the superscript (m). For example, $\mu_0^{(251)}$ denotes the value of the latent variable μ_0 in iteration round 251.

The conditional distributions need to be constructed for different sets of variables. Per set, one needs to be able to draw all parameters and latent variables within that set simultaneously (given the other parameters and latent values),

so we need to group them accordingly. The sets that we create are as follows: $[\sigma_\varepsilon,\ \sigma_{\Delta\mu},\ \sigma_{\lambda,u},\ \sigma_{\lambda,d},\ d_1^*]$, $[\Delta\mu^*,\ \lambda_u^*,\ \lambda_d^*,\ \beta]$, $[\lambda_1,\ \lambda_2,\ ...,\ \lambda_K]$, $[\mu_0]$, $[\mu_1]$, ..., $[\mu_K]$, $[\tau_1]$, $[\tau_2]$, ..., $[\tau_K]$. This amounts to $2K+4$ sets. We denote the sets using the notation B_1, B_2, ..., B_{2K+4}. To denote all sets except B_i, we use the notation B_{-i}. To denote the sets with a lower of higher index, we use $B_{<i}$ and $B_{>i}$.

The conditional distributions

We now present and discuss the conditional distributions of each individual set.

First we discuss the conditional distribution $B_1^{(m)}|B_{-1}^{(m-1)}$. This set consists of all the σ type parameters. The conditional distribution can be derived for each parameter separately, as these parameters do not directly affect each others contribution to the likelihood function. To draw $\sigma_\varepsilon^{(m)}$, we calculate the residuals of (11) and denote these residuals as $\hat{\varepsilon}_t^{(m)}$. Then we have

$$\sigma_\varepsilon^{2(m)}|B_{-1}^{(m-1)} \sim IG(\sum_{t=1}^{T}\hat{\varepsilon}_t^2, T) \tag{19}$$

with IG denoting the Inverted Gamma distribution. Similarly, the other σ type variables can be drawn by rewriting their defining equations in a residual form, that is,

$$\sigma_{\Delta\mu}^{2\ (m)}|B_{-1}^{(m-1)} \sim IG(\sum_{i=1}^{K}(|\mu_i-\mu_{i-1}|-\Delta\mu^*)^2, K-1) \tag{20}$$

$$\sigma_{\lambda,u}^{2\ (m)}|B_{-1}^{(m-1)} \sim IG(\sum_{s_\kappa=H}(\lambda_\kappa^u-\lambda_u^*)^2, \sum_{\kappa=1}^{K} I[s_\kappa=H]) \tag{21}$$

$$\sigma_{\lambda,d}^{2\ (m)}|B_{-1}^{(m-1)} \sim IG(\sum_{s_\kappa=L}(\lambda_\kappa^d-\lambda_d^*)^2, \sum_{\kappa=1}^{K} I[s_\kappa=L]) \tag{22}$$

Finally, for the draw of d_1^* we only need to observe the moment of the first switch $\tau_1^{(m-1)}$ and the duration dependence parameters $\beta^{(m-1)}$. The contribution to the likelihood of d_1^* is then the probability of switching at $t=\tau_1$ times the probability of not switching earlier, like

$$L(d_1^*) \propto \Phi(\beta^{(m-1)}D_{d_1^*+\tau_1-1}^{(m-1)})\prod_{t=1}^{d_1^*+\tau_1-1}(1-\Phi(\beta^{(m-1)}D_t^{(m-1)})) \tag{23}$$

We draw the new value for $d_1^{*(m)}$ from 0, 1, ... using the probabilities $p(j) = \frac{L(j)}{\sum_{i=1}^{\infty} L(i)}$.

Next, for the draws of $B_2^{(m)}|B_1^{(m)}, B_{>2}^{(m-1)}$, we again can split the set into parts that have no influence on each others' likelihood contribution. For $\Delta\mu^*|B_1^{(m)}, B_{>2}^{(m-1)}$ we can rewrite (15) to a normal distribution with mean equal to the average difference between subsequent state means and the variance equal to the sample mean variance, that is,

$$\Delta\mu^*|B_1^{(m)}, B_{>2}^{(m-1)} = N(\frac{1}{K}\sum_{\kappa=1}^{K}|\mu_\kappa - \mu_{\kappa-1}|, \frac{\sigma^2_{\Delta\mu}}{K}) \tag{24}$$

After applying a logarithmic transformation to (17), we can apply the same method to find the conditionals of λ_u^* and λ_d^*:

$$\lambda_u^*|B_1^{(m)}, B_{>2}^{(m-1)} = N(\frac{1}{K}\sum_{s_\kappa=H}\lambda_\kappa^u, \frac{\sigma^2_{\lambda,u}}{\sum_{\kappa=1}^{K} I[s_\kappa = H]}) \tag{25}$$

$$\lambda_d^*|B_1^{(m)}, B_{>2}^{(m-1)} = N(\frac{1}{K}\sum_{s_\kappa=L}\lambda_\kappa^d, \frac{\sigma^2_{\lambda,d}}{\sum_{\kappa=1}^{K} I[s_\kappa = L]}) \tag{26}$$

For the simulation of β, we rewrite (13) by introducing the latent variable z_t:

$$z_t = \beta D_t + \eta_t, \qquad \eta_t \sim N(0,1) \tag{27}$$

$$\begin{aligned} s_t &\neq s_{t-1} \quad \text{if } z_t \geq 0 \\ s_t &= s_{t-1} \quad \text{if } z_t < 0 \end{aligned} \tag{28}$$

Then, we simulate z_t from a truncated normal using the observation that there is a switch or not at time t, that is,

$$z_t^{(m)} \sim \begin{cases} TN(\beta^{(m-1)}D_t^{(m-1)}, 1, 0, \infty) & \text{if } s_t^{(m-1)} \neq s_{t-1}^{(m-1)} \\ TN(\beta^{(m-1)}D_t^{(m-1)}, 1, -\infty, 0) & \text{if } s_t^{(m-1)} = s_{t-1}^{(m-1)} \end{cases} \tag{29}$$

After that, we can simulate $\beta^{(m)}$ using a normal distribution based on the OLS regression of $z_t^{(m)}$ on $D_t^{(m-1)}$, like

$$\beta^{(m)} \sim N(\hat{\beta}_{OLS}^{(m)}, ({D_t}^{(m-1)\prime}{D_t}^{(m-1)})^{-1}) \tag{30}$$

For $B_3^{(m)}|B_{<3}^{(m)}, B_{>3}^{(m-1)}$, we make use of a Metropolis-Hastings sampler (MH; see Chib and Greenberg, 1995) for each individual λ_κ. For the MH sampler we need a candidate-generating function and a likelihood function for evaluation. For the candidate-generating function, we make use of (17) and (18) to draw from $g(\lambda_\kappa)$, which is a truncated log-normal distribution with parameters λ_u^* and $\sigma_{\lambda,u}$ for an

upwards transition (λ_d^* and $\sigma_{\lambda,d}$ for downwards) and an upperbound equal to $\tau_{\kappa+1} - \tau_\kappa$. For the likelihood function we use the contribution of λ_κ to the likelihood, $f(\lambda_\kappa)$, which is based on its effect on μ_t via (12) and on its own likelihood via (17). The probability of accepting the candidate is

$$\alpha = \min\left(\frac{f(\lambda_\kappa^*)g(\lambda_\kappa^{(m-1)})}{f(\lambda_\kappa^{(m-1)})g(\lambda_\kappa^*)}, 1\right), \tag{31}$$

otherwise, $\lambda_\kappa^{(m)} = \lambda_\kappa^{(m-1)}$. As the definition $g(\lambda_\kappa)$ is part of the definition $f(\lambda_\kappa)$, this drops out of the fraction and thus we can also define $h(\lambda_\kappa) = \frac{f(\lambda_\kappa)}{g(\lambda_\kappa)}$, which only looks at the contribution to the likelihood based on its effect on μ_t via (12), and then use

$$\alpha = \min\left(\frac{h(\lambda_\kappa^*)}{h(\lambda_\kappa^{(m-1)})}, 1\right) \tag{32}$$

We use this approach for each individual λ_κ. The drawn value of one λ_κ will not affect the distribution of the other λ_κ's, which is the reason we can include them all in one set B_3.

For the sets $B_4^{(m)}|B_{<4}^{(m)}, B_{>4}^{(m-1)}$; ... ; $B_{K+4}^{(m)}|B_{<K+4}^{(m)}, B_{>K+4}^{(m-1)}$ we can use the same approach. For each B_i with $i = 4, \ldots, K+4$, we only draw μ_κ with $\kappa = i-4$. For this, we at first will neglect the restriction that subsequent state means must alternately be higher and lower. We can then rewrite (11), (12) and (15) to a regression of y_t and $\Delta\mu^*$ on transformations of d_t and λ_{κ_t} and on 1 and -1, that is,

$$y_t = \begin{cases} \mu_{\kappa_t} + \varepsilon_t & \text{if } t \in [\tau_{\kappa_t} + \lambda_{\kappa_t}; \tau_{\kappa_t+1}] \\ \mu_{\kappa_t}\frac{d_t}{\lambda_{\kappa_t}} + \mu_{\kappa_t-1}(1 - \frac{d_t}{\lambda_{\kappa_t}}) + \varepsilon_t & \text{if } t \in (\tau_{\kappa_t}; \tau_{\kappa_t} + \lambda_{\kappa_t}) \end{cases} \tag{33}$$

$$\Delta\mu^* = \begin{cases} \mu_\kappa - \mu_{\kappa-1} + \zeta_\kappa & \text{if } s_\kappa = H \\ \mu_{\kappa-1} - \mu_\kappa + \zeta_\kappa & \text{if } s_\kappa = L \end{cases} \tag{34}$$

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2) \qquad \zeta_\kappa \sim N(0, \sigma_{\Delta\mu}^2) \tag{35}$$

We standardize all equations by dividing each term by the associated standard deviation and collect the variables on the right hand side (except for μ_κ itself) in the matrix X. We collect the μ_κ variables in the vector μ. Without the alternating increase or decrease in state mean, we could now sample all μ_κ using $\mu \sim N(\hat{\mu}_{OLS}, (X'X)^{-1})$. As we do not want to interfere in the alternating state mean restriction, we can sample the μ_κ one by one, conditional on all the others, using the standard formula

for conditional normal distributions[1]. We restrict these to be either higher or lower than both the newly drawn previous state mean $\mu_{\kappa-1}^{(m)}$ and the next state mean of the previous iteration $\mu_{\kappa+1}^{(m-1)}$.

For the sets $B_{K+5}^{(m)}|B_{<K+5}^{(m)}, B_{>K+5}^{(m-1)}$; ... ; $B_{2K+4}^{(m)}|B_{<2K+4}^{(m)}, B_{>2K+4}^{(m-1)}$ we can use the same approach for each individual set. Each one of these sets consists of only one latent variable, that is, τ_{κ_t}. For this, we again make use of a MH-sampling method, for which we again need a likelihood-evaluating function f and a candidate generating function g. As candidate-generating function, we consider a discrete uniform distribution between the end of the previous transition $(\tau_{\kappa-1}^{(m)} + \lambda_{\kappa-1}^{(m)})$ and the end of the current transition $(\tau_{\kappa}^{(m-1)} + \lambda_{\kappa}^{(m)})$. As we want to let τ_κ only influence the start of the transition and not also the end, we adjust the transition length: $\lambda_\kappa^* = \lambda_\kappa^{(m-1)} + (\tau_\kappa^{(m-1)} - \tau_\kappa^*)$. For the evaluation of the likelihood, we need to observe that τ_κ influences the likelihood in two ways, that is, (i) changing the observation mean μ_t via (12), and (ii) the time periods during which the probit in (13) equals 1 (and thus also when it is 0). The adjusted λ_κ also affects (12), and along with that it contributes to the likelihood via (17). As our candidate-generating function is a uniform distribution, its pdf has the same value for each input in its support and it disappears from the equation to calculate the acceptance probability. This means that in this case the probability of accepting the candidate $\tau_\kappa^{(m)} = \tau_\kappa^*$ equals

$$\alpha = \min\left(\frac{f(\tau_\kappa^*, \lambda_\kappa^*)}{f(\tau_\kappa^{(m-1)}, \lambda_\kappa^{(m-1)})}, 1\right), \tag{36}$$

otherwise, $\tau_\kappa^{(m)} = \tau_\kappa^{(m-1)}$, and similarly for $\lambda_\kappa^{(m)}$.

Finally, we relax the assumption that the number of switches is known to be K. Instead, assume that the number of state switches in the sample is an element of {K, K+1, K+2}. That is, we allow for two states for which it is open whether they are part of the sample or not. For this, we will allow for two additional sets of state parameters, μ_{K+1}, μ_{K+2}, λ_{K+1}, λ_{K+2}, τ_{K+1} and τ_{K+2}. The state means and transition speeds might be partly simulated using the time series, if the associated

[1] If X_1 and X_2 are both multivariate normally distributed vectors with means μ_1 and μ_2, covariance matrices Σ_1 and Σ_2 and cross-covariance matrix Σ_{12}, then $X_1|X_2 \sim N(\mu_1 + \Sigma_{12}'\Sigma_2^{-1}(X_2 - \mu_2), \Sigma_1 - \Sigma_{12}'\Sigma_2^{-1}\Sigma_{12})$

switches occur before $t = T$. Otherwise, the simulated values will be entirely due to the distribution assumptions (13)-(18). Similarly, the simulation of the parameters β, $\Delta\mu^*$, $\sigma_{\Delta\mu}$, λ_u^*, λ_d^*, $\sigma_{\lambda,u}$ and $\sigma_{\lambda,d}$ now incorporate the additional two states and its latent values.

In practice, we would advise to set K such that K+1 equals the suspected number of switches in the sample. This way one can account for the situation that the expected last switch might not have happened, or otherwise, that an unexpected switch did occur. To evaluate the number of switches outside this interval, one could compare the estimated average likelihoods for different choices of K, possibly including a penalizing term for higher values of K.

Forecasting and real-time monitoring

We can of course also construct forecasts of y_t, $t > T$, and also for the associated latent values. For this one can fix the parameters to the mean or median of the draws obtained using the Gibbs sampling method. To account for parameter uncertainty, it is however better to draw the parameters used in forecasting from the entire posterior distribution. This can be easily done in practice by constructing a forecast in each iteration of the estimation process using the values of the parameters in that iteration. More elaborate sampling methods need to be used if one wants to forecast from a different starting point than at the end of the estimation sample. In that case, we would also need to re-estimate all the latent variables using only information up until that starting point.

We can forecast the observation mean μ_t by simulating from (13)-(18). To account for the restriction in (18), we first simulate the transition length λ, and then simulate the next state switch moment τ so that the λ is smaller than the difference between the two subsequent state switches. To obtain a full forecasting distribution, we also simulate the observation error ε_t using (11). If necessary, point and interval forecasts can be obtained using expected quantiles of this distribution.

Related to forecasting is the concept of real-time monitoring, in which the estimates of now relevant latent variables and short-horizon forecasts are updated each time a new data point becomes available. The approach for this is comparable to

the one we use for forecasting, with the main difference that for real-time monitoring one needs to reapply the sampling procedure each time period. Due to the newly available sample point, the latent variables associated with the final states in the sample will change, and the best way to account for this is by re-estimating the previous latent variables and the parameters. This estimation process may take some time (depending on processor speed, programming efficiency and software), which can be in contrast to the goals of real-time monitoring, for which in fact one needs the updated estimates as quickly as possible after obtaining the new data point. For a quicker updating of the latent variables of the final states, one could fix the parameters and the latent variables of the previous states. That way, re-estimating will be done using less sets of parameters in the Gibbs sampling, which leads to less autocorrelation in the draws and thus to a smaller simulation sample that is necessary to obtain an accurate distribution.

4 Application to US unemployment

In this section we illustrate our model and estimation process on monthly unemployment in the United States for the period 1948 to 2012. Our estimation sample runs until 1992 (covering 540 months), which leaves 240 months for the forecast evaluation. For the estimation sample we restrict our K to be an element of $\{15, 16, 17\}$, based on visual inspection of the data.

Estimation results

Our estimation results are shown in Table 1 for which we have used in total 110000 iterations in the sampling process. After accounting for the burn-in period of 10000 iterations and a thinning factor of 45, this results in 2000 as-good-as-independent draws from the posterior distribution of the parameters and the latent variables.

The estimates of β_1 and β_3 show that there is asymmetry concerning the switching behavior, although it is not statistically significant. The estimate of β_2 shows that the upwards switch is not duration dependent, as zero is approximately in the middle of the HPD interval. In contrast, for the downwards switch the results show that

there might be duration dependence as zero is on the border of the interval for $\beta_2 + \beta_3$. In fact, there are many ways in which we could have applied our thinning differently and that could have resulted in an interval that would not include zero.

Next, the estimate of $\Delta\mu^*$ shows that on average the high states and low states in unemployment differ about 2.8 percentage points. The ratio of $\frac{\Delta\mu^*}{\sigma_{\Delta\mu}}$ suggests that switches in the wrong direction are not likely even if we would remove the truncation, as the average of this ratio is 2.810, of which the negative (-2.810) is the 0.25-th percentile of the normal distribution. In fact, if we would calculate for each iteration the probability that one individual state mean change is in the wrong direction ($\Phi(\frac{0-\Delta\mu^*}{\sigma_{\Delta\mu}})$), then the HPD-region of these probabilities only runs to 0.033 and it has a mean of 0.008. The median is even lower, that is, 0.003. This shows that the truncation restriction is not of much influence on our estimation results.

Finally, we discuss the estimates of the transition speed parameters. The estimates of λ_u^*, λ_d^*, $\sigma_{\lambda,u}$ and $\sigma_{\lambda,d}$ seem to be quite accurate, considering their relatively small HPD intervals. The average transition length however is an exponential function of both of them, which can result in blowing up small differences to large effects. The average upwards transition length, $e^{\lambda_u^* + \frac{1}{2}\sigma_{\lambda,u}^2}$, is 14.6, which means that on average it takes almost five quarters for unemployment to transit from a local minimum to a local maximum, what would be called a recession. The opposite movement takes much more time, as the average for $e^{\lambda_d^* + \frac{1}{2}\sigma_{\lambda,d}^2}$ amounts to more than three years. This shows the familiar property for unemployment that an increase in unemployment is much quicker than a decrease.

Table 2 shows the average results for the latent variables. The first two columns present results on the timing of the start of the switch (τ_κ). Next, the results on the length of the transitions (λ_κ) follow, and the final two columns contain results on the state means (μ_κ).

The first fifteen states all have a state switch that falls entirely inside the sample that runs to $t = 540$. These state switches can thus be estimated quite accurately in most cases. Only the 10th state switch, which occurred around December 1971 ($t = 287$), has a standard deviation that is relatively high (4.150). For this state, the observations gradually start to decline and so there is no clear visual starting point

of the transition. The model arrives at the same conclusion. The last two states do not fall entirely inside the sample. For $\kappa = 17$, the state switch is definitely after $t = 540$, but for $\kappa = 16$, this is not so clear. On average, the switch occurs after $t = 540$, but in fact in 50.5% of the iterations the state switch τ_{16} occurs on or before $t = 540$. This shows why it is important to account for multiple possible numbers of switches as it might be unclear whether a switch has occurred or not, and then one can account for both situations. As can be expected, both τ's that fall (partially) outside the sample are estimated less accurately than those inside the sample.

For the values of λ_κ in Table 2, we can observe an alternating pattern of high and low values. This is due to the different transition behavior for upwards and downwards transitions, which was also evident from the final two rows of Table 1. Again, of all in-sample states, state 10 has the most uncertain estimate of the latent variable. This makes sense, because if the start of the transition is unclear, then the length of the transition is also most likely unclear, as that depends on the start. Also, the two latent variables that fall outside the sample ($\lambda_1 6$ and $\lambda_1 7$) are both less accurate than all in-sample estimates, which is similar to the situation for the corresponding τ variables.

For the in-sample state means (μ_κ in Table 2), we see again an alternating pattern of high and low values, which follows directly from the relation between two subsequent state means. The two state means that are most close to each other on average are $\mu_9 \approx 5.013$ and $\mu_{10} \approx 6.031$, which still differ more than 1 percentage point. Also, there is no single value that is clearly less accurate than the others. The highest in-sample posterior standard deviation is 0.185 for μ_1, while the lowest is 0.060 for μ_8. The numbers in between seem to be about evenly spread out. Even the two state means around the tenth switch, μ_9 and μ_{10}, are both estimated quite accurately, in contrast to the situation for τ_{10} and λ_{10}. Again, both state means that are (partially) outside of the sample are estimated less accurately.

Based on our estimates of the parameters and the latent variables, some interesting graphs can be constructed. Figure 6 shows the original data along with the estimated mean and HPD intervals for the state means μ_t until the end of the estimation sample $t = 540$. We also present 7.5 years of forecasts, all constructed

using information until $t = 540$. It can be seen that the in-sample intervals are much smaller than the out-of-sample intervals, as could well be expected. Quite noticeable is that the forecasts exhibit the same cyclical property of the model, and this is a feature that does not follow from a standard MS model with two states.

Figure 7 shows the probability of switching out of the state for both the upward and the downward switch, with on the horizontal axis the time already spent in that state. These probabilities are calculated for the entire posterior distribution of β and they incorporate the parameter uncertainty. For both state types, the probability of switching away increases the longer the duration of the state. For the downward switch this increase is obtained earlier than for the upward switch, as its line is mostly above the line of the upward switch.

Figure 8 shows the histogram of the durations of both states, accounting for both the uncertainty in β and for the completed transition restriction in (18). The histograms clearly show that the downward switch happens faster than the upward switch. In fact, the averages amount to 2.3 years and 5.6 years, so upward switches occur after more than twice the time of downward switches, which means that high states last twice as short as the lower states. For both states an immediate switch is not impossible, although it is unlikely. As a comparison, for a standard MS model this histogram would be monotonically declining, and an immediate switch is actually the most likely.

Finally, Figure 9 shows the histogram of the length of a total cycle, thus incorporating both an upward and a downward switch using the values of β in the same iteration of the Gibbs sampler for both cycles. The pattern is comparable to the patterns in Figure 8, which was to be expected as it is the sum of both histograms in that figure. The average cycle length amounts to 7.8 years, which corresponds well with the common socio-economic cycle periods mentioned in de Groot and Franses (2012).

Real-time monitoring

An interesting application of our model is real-time monitoring, in which one investigates how much the estimates of latent variables get updated when new data

points become available. We now illustrate how real-time monitoring can be applied using our model. For this we first re-estimate all latent variables using the data only until $t = 1$, only until $t = 2$, up to until $t = 540$, while fixing the parameters to the posterior distribution that has been estimated previously. For this part of the estimation process, we can use less draws as we can randomly draw the parameters from the posterior distribution, thereby decreasing the autocorrelation of the draws of the latent variables. That is why we decided to use 10000 draws with a burn-in sample of 1000 for each individual monitoring process.

This results in 540 estimated distributions for each individual latent variable. Using this we can calculate all sorts of statistics to see how the estimates evolve over time. For example, we can calculate the estimated mean of state 5, μ_5, using any possible set of information, to see from which point on the estimate of μ_5 does not vary any more. Or, we can find the width of the 95% HPD interval of μ_5, to see when the estimate of μ_5 first meets a certain accuracy requirement.

Figure 10 shows the average width (calculated over the entire sample period) of the 95% HPD interval for μ_t. This width has been calculated using information starting from $t-100$ up until $t+100$, thus using a total time period of over 16 years around each observation. This shows how the accuracy in estimating the observations' mean evolves when more information about that mean becomes available. It can be seen that the width decreases the fastest just before and just after time t. There is a slower decrease for the time periods that are well before t, while after a few months after t there is no information gain left anymore.

Next, Figure 11 shows the in-sample Root Mean Squared Error for each observation y_t, calculated using the in-sample forecast $\hat{y}_{t|t+h}$. As expected, for low h this value approaches σ_ε, while for high values of h the forecast error is larger. Even though this graph shows a different characteristic than Figure 10, they both show a similar pattern.

Forecasting

We have constructed forecasts for the last 240 hold-out observations, allowing for varying forecast horizons. We produce these forecasts starting at different starting

points, for which we each time need to re-estimate the latent variables to account for the new information. We will not update the parameter estimates, however, as these are based on the first 540 observations. For simulating the latent variables we have used the same approach as discussed for real-time monitoring, with again 10000 iterations per information set.

We will compare the forecasting performance of our model with several others. We use two simple forecasting models as a baseline comparison. The first of these is just taking the average of y_t for $t = 1, \ldots, 540$, so this means that we forecast the future using the sample average of the estimation sample. The second simple model is the Random Walk model, in which each forecast is just the most recently observed value at that time. In other words, $\hat{y}_{t+h|t} = y_t$. We also compare our model with two other Markov Switching models, namely the two-state and the three-state models. These models do not incorporate the duration dependence or stochastic means of our model, but instead they use fixed transition probabilities and fixed state means. To use an approach that comes close to our approach, we have estimated these models using Gibbs sampling on the same estimation sample and we forecast using draws from the entire posterior distribution of the parameters. For these simpler models, we have used 50000 iterations to estimate the parameters (after 100 burn-in iterations), and 5000 iterations for updating the estimates as the forecasting windows moves.

Table 3 shows the results of this forecast comparison. For each competing model, the RMSE (Root Mean Squared Error) has been calculated and then divided by the RMSE of our model. Values above 1 indicate that the alternative model performs worse, and values below 1 indicate the opposite. As can be seen, our model is the best model for the short-to-mid-term: for forecasting 6 months to 2 years ahead, our model beats the Random Walk model and a simple first order autoregression, and it is much better than both other MS models. For the other forecast horizons, our model is beaten by other models. On the very short horizon of 1 month, this defeat is no surprise, as our model takes no short-term information into account and the two models that beat our model here do. On the other hand, we easily outperform both other MS models again for 1 month ahead. On the longer term (more than 3

years), our model loses from all alternatives. For 3 to 5 years ahead the forecasting performance is not too bad as our model still beats the Random Walk, and no other model outperforms by more than 8%. For the two longest horizons however, our model performs very poorly. This might be because in the long-run our simulated state cycles are often out-of-sync with reality. This latter feature is studied next, using simulations.

5 Simulations

In this section we investigate the accuracy of our estimation method in practically realistic situations. We do this by simulating multiple time series from a Data Generating Process (DGP) and by applying our estimation method to these time series.

As DGP we use the model and its parameters as presented in Section 4. We set the sample size in our simulations at 540, which is the length of the time series used in Section 4, and at 2160. The number of simulated time series in both cases is 400, while we use 10000 iterations in the estimation process after a burn-in of 1000 iterations.

Various summary statistics of the simulation results are shown in Table 4. As these results are calculated across 400 time series with 10000 iterations each in the estimation process, we report summary scores of various statistics, like for example the standard deviation of the mean. In that case, the standard deviation is calculated based on 400 values of the mean, of which each individual value is based on 10000 iterations.

Table 4 also shows the values of the parameters in the DGP. For most parameters, the mean of the mean and median of the median are quite close to the true DGP values. This holds true for $T = 540$ and even more so if $T = 2160$. Moreover, in all cases the spread in the point estimate (*StDev of mean*) is smaller when using more observations, which was to be expected. For some variables the *StDev of mean* for $T = 2160$ is about 10% of the same statistic for $T = 540$, while for most it is about a half. The least improvement is made for d_1^* (only a 16% drop to 84% of

previous value), which might be explained by the fact that this value is for a large part affected by the initial observations only, and these cannot be influenced by the choice of T.

The β parameters are the only ones that are not always accurately estimated, especially if $T = 540$. The reason for the large differences in the mean of the posterior distribution (column *StDev of mean*) is the apparently small number of state switches in a sample of this length. On average the number of state switches is about equal to 15, as in the estimation sample of Section 4, which is already quite small, and for some of the simulated series this number dropped to as low as 10. Naturally, estimating the parameters in a probit model with four explanatory variables using only 10 observations with a '1' results in substantial uncertainty around the estimates. Having a longer time series obviously will make this situation less likely, and this is evident from the much lower values of *StDev of mean* for β when $T = 2160$. Also, for larger T the point estimates are on average much closer to the true values, even though there is still room for improvement.

The final two columns for both values of T provide an indication of how the DGP configurations are located as compared to the posterior. The *mean of StDev* shows how narrow (or wide) the estimated posterior is. Most posteriors are much more narrow for a higher T, with d_1^* as only exception as that parameter is not really affected by the value of T. The *mean of quantile* shows the quantiles where the true parameters are located. Both sample sizes show a similar pattern, so the sample size does not seem to matter much.

6 Conclusion

In this paper we have introduced a new model that can deal with changing levels and cyclicality in time series. We have proposed a Markov switching model with two states that each have a stochastic mean, where the transition behavior of these states is governed by duration dependence and stochastic linear transition periods. We have shown with artificial data that data from this model have characteristics comparable with actual data. We have presented an estimation method that uses

Gibbs Sampling with Data Augmentation, which also generates a density forecast. We have applied this estimation method to postwar monthly US unemployment and we have found that for two to three years ahead forecasts, our model has superior forecasting performance compared to a set of benchmark models. We have also shown, using a set of simulations, that the parameters of our model can be estimated quite accurately, granted that there is a sufficient number of state switches.

We envisage various potential extensions to our model and analysis. The major drawback of our model, as we have seen in the simulation exercise, is the potential difficulty in estimating parameters that fully depend on state switches. For many currently available samples of macroeconomic data, one typically encounters a limited number of state switches. One way to alleviate this is to jointly model several time series for which a common parameter can be assumed. Alternatively, the parameters of the different time series can be linked using an underlying joint distribution.

Applications to other than macroeconomic series can be particularly interesting. We then think of high-frequency financial time series data or data in marketing contexts, where different regimes may occur much more frequently. Other extensions could include implementing an autoregressive model to the stochastic-mean part of the model (15) or to the shocks in (11). An alternative distribution like a log-normal distribution instead of the truncated normal in (15) could also be considered.

References

Bauwens, L., Preminger, A., Rombouts, J., 2010. Theory and inference for a Markov switching GARCH model. The Econometrics Journal 13, 218–244.

Castro, V., March 2010. The duration of economic expansions and recessions: More than duration dependence. Journal of Macroeconomics 32 (1), 347–365.

Chen, F., F. D., Schorfheide, F., December 2013. A Markov-switching multifractal inter-trade duration model, with application to US equities. Journal of Econometrics 177 (2), 320–342.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings Algorithm. The American Statistician 49 (4), 327–335.

Cunningham, R., Kolet, I., 2011. Housing market cycles and duration dependence in the United States and Canada. Applied Economics 43 (5), 569–586.

de Groot, B., Franses, P. H., January 2012. Common socio-economic cycle periods. Technological Forecasting and Social Change 79 (1), 59–68.

Diebold, F., Rudebusch, G., June 1990. A nonparametric investigation of duration dependence in the American business cycle. Journal of Political Economy 98 (3), 596–616.

Durland, J., McCurdy, T., 1994. Duration-dependent transitions in a Markov model of U.S. GNP growth. Journal of Business & Economic Statistics 12 (3), 279–288.

Guérin, P., Marcellino, M., 2013. Markov-Switching MIDAS models. Journal of Business & Economic Statistics 31 (1), 45–56.

Hamilton, J., March 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57 (2), 357–384.

Kim, C., January 2009. Markov-switching models with endogenous explanatory variables II: A two-step MLE procedure. Journal of Econometrics 148 (1), 46–55.

Lam, P., February 2004. A Markov-Switching model of GNP growth with duration dependence. International Economic Review 45 (1), 175–204.

Layton, A., Smith, D., December 2007. Business cycle dynamics with duration dependence and leading indicators. Journal of Macroeconomics 29 (4), 855–875.

Lunde, A., Timmermann, A., 2004. Duration dependence in stock prices. Journal of Business & Economic Statistics 22 (3), 253–273.

Nalewaik, J., April-June 2011. Incorporating vintage differences and forecasts into Markov switching models. International Journal of Forecasting 27 (2), 281–307.

Sichel, D., May 1991. Business cycle duration dependence: A parametric approach. The Review of Economics and Statistics 73 (2), 254–260.

Tables and Figures

Table 1: Results on the posterior density of the main parameters of applying the SMDDMS-SLT model to US unemployment.

Parameter	Average	Standard Error	95% HPD interval	
σ_ε	0,313	0,010	0,292	0,332
β_0	-1,967	0,342	-2,688	-1,347
β_1	-0,594	0,476	-1,471	0,415
β_2	0,021	0,019	-0,016	0,055
β_3	-0,011	0,019	-0,051	0,023
$\Delta\mu^*$	2,806	0,280	2,244	3,333
$\sigma_{\Delta\mu}$	1,040	0,222	0,673	1,469
λ_u^*	2,375	0,265	1,857	2,909
λ_d^*	3,184	0,326	2,554	3,792
$\sigma_{\lambda,u}$	0,713	0,125	0,519	0,953
$\sigma_{\lambda,d}$	0,816	0,164	0,566	1,144
d_1^*	46,455	29,592	0	98
$\beta_0+\beta_1$	-2,561	0,329	-3,218	-1,940
$\beta_2+\beta_3$	0,011	0,006	-0,001	0,022
$\frac{\Delta\mu^*}{\sigma_{\Delta\mu}}$	2,810	0,605	1,663	3,982
$\Phi(\frac{0-\Delta\mu^*}{\sigma_{\Delta\mu}})$	0,008	0,015	0,000	0,033
$e^{\lambda_u^*+\frac{1}{2}\sigma_{\lambda,u}^2}$	14,558	4,612	7,857	23,568
$e^{\lambda_d^*+\frac{1}{2}\sigma_{\lambda,d}^2}$	36,669	17,654	15,177	63,317

Table 2: Average results on the latent variables. For example, the sixth state switch is estimated to start around $t = 127$, the transition length for this switch is about 9 and the level of the sixth state is about 5.7.

	τ_κ		λ_κ		μ_κ	
κ	Mean	StDev	Mean	StDev	Mean	StDev
0					7,273	0,097
1	11,386	0,734	8,634	1,358	4,019	0,185
2	22,685	0,979	17,582	1,402	8,024	0,059
3	69,541	0,593	5,732	1,104	5,093	0,155
4	79,874	1,515	10,316	2,370	6,885	0,062
5	116,981	0,655	7,565	1,110	3,555	0,176
6	127,021	0,999	9,207	1,523	5,746	0,089
7	150,353	1,573	6,686	2,133	4,342	0,090
8	159,719	1,667	69,800	3,184	7,373	0,060
9	263,398	0,985	12,534	1,750	5,013	0,107
10	287,225	4,150	17,099	5,828	6,031	0,090
11	318,829	0,854	8,314	1,293	2,337	0,120
12	329,791	1,674	40,408	3,041	5,156	0,113
13	380,478	1,521	36,920	1,817	1,394	0,074
14	418,860	0,816	65,760	1,992	5,672	0,069
15	507,878	1,605	21,436	3,643	3,474	0,133
16	541,674	6,440	18,445	10,247	6,070	1,105
17	578,684	17,344	9,725	6,210	3,264	1,461

Table 3: RMSE relative to our model. A value larger than 1 indicates that the corresponding model performs less than our model for the corresponding horizon, while a value smaller than 1 indicates that the model outperforms our model.

Forecast horizon	Fixed Mean	Random Walk	AR1	MS2	MS3
1 month	9,599	0,762	0,766	5,290	4,615
6 months	3,314	1,095	1,083	2,256	2,016
1 year	1,838	1,096	1,055	1,496	1,426
1.5 years	1,366	1,106	1,038	1,228	1,233
2 years	1,142	1,131	1,033	1,083	1,163
3 years	0,978	1,184	1,035	0,972	1,036
4 years	0,951	1,218	1,031	0,955	0,973
5 years	0,947	1,175	0,968	0,941	0,921
7.5 years	0,769	0,786	0,680	0,751	0,726
10 years	0,791	1,062	0,783	0,770	0,787

Table 4: Summary results of multiple parameter distributions estimated using simulated data from the same Data Generating Process for time series length $T = 540$ and $T = 2160$, with 400 time series for each case. Parameters have been estimated using Gibbs sampling with 10000 iterations (after a burn-in of 100 iterations). The table presents the mean and standard deviation of the mean of the posterior distribution, the median of the median of the posterior distribution, the mean of the standard deviation of the posterior distribution, and the mean of the quantile of the DGP parameter in the posterior distribution.

		$T = 540$					$T = 2160$				
	Data Generating Process	Mean of mean	StDev of mean	Median of median	Mean of StDev	Mean of Quantile	Mean of mean	StDev of mean	Median of median	Mean of StDev	Mean of Quantile
σ_ε	0,313	0,315	0,011	0,315	0,012	0,544	0,315	0,007	0,314	0,005	0,565
β_0	-1,967	-3,813	3,225	-2,919	1,105	0,072	-2,597	0,304	-2,568	0,255	0,030
β_1	-0,594	0,116	3,404	-0,354	1,462	0,599	-0,319	0,421	-0,352	0,338	0,693
β_2	0,021	0,094	0,162	0,053	0,047	0,835	0,037	0,016	0,036	0,010	0,821
β_3	-0,011	-0,068	0,163	-0,030	0,050	0,315	-0,025	0,017	-0,023	0,010	0,251
$\Delta\mu^*$	2,806	2,742	0,309	2,743	0,344	0,445	2,775	0,169	2,781	0,163	0,447
$\sigma_{\Delta\mu}$	1,040	1,080	0,247	1,023	0,279	0,482	1,045	0,111	1,036	0,119	0,482
λ_u^*	2,375	1,914	5,935	2,350	1,356	0,495	2,344	0,163	2,342	0,156	0,442
λ_d^*	3,184	3,151	0,364	3,155	0,358	0,469	3,161	0,187	3,176	0,165	0,467
$\sigma_{\lambda,u}$	0,713	0,726	0,130	0,675	0,202	0,416	0,683	0,055	0,672	0,063	0,321
$\sigma_{\lambda,d}$	0,816	0,764	0,139	0,709	0,200	0,302	0,738	0,062	0,728	0,070	0,185
d_1^*	46,455	49,858	18,771	49	24,947	0,497	51,329	15,241	50	28,022	0,516

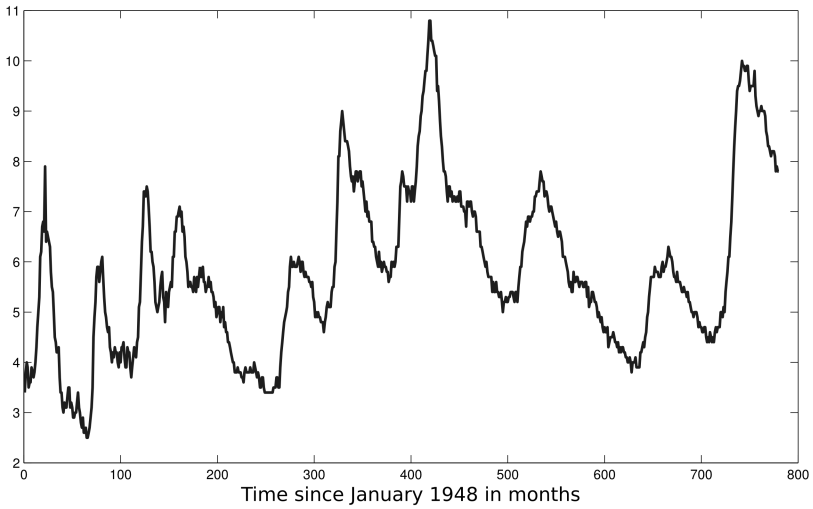

Figure 1: Monthly unemployment in the United States in the period 1948 to 2012.

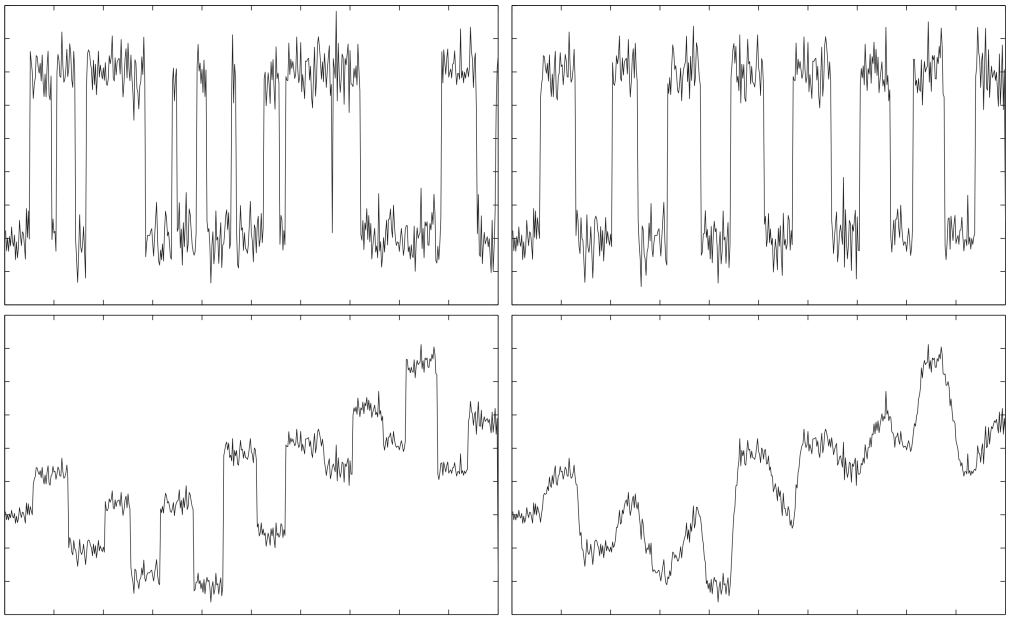

Figure 2: Four stylized series to characterize the gradual steps from a standard MS model to our model.

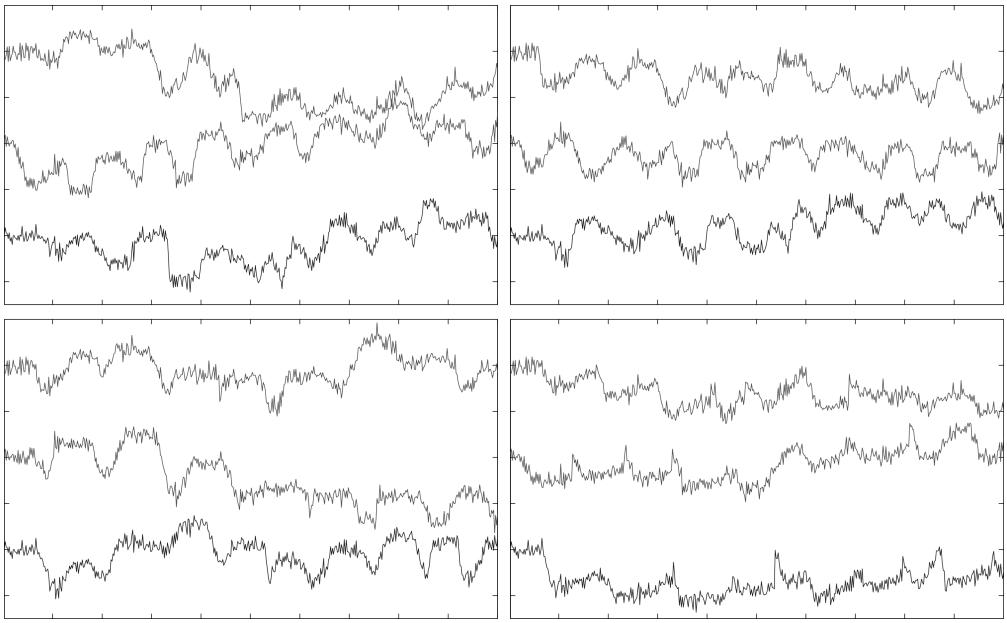

Figure 3: Three sample series of each of a reference DGP and three alternatives.

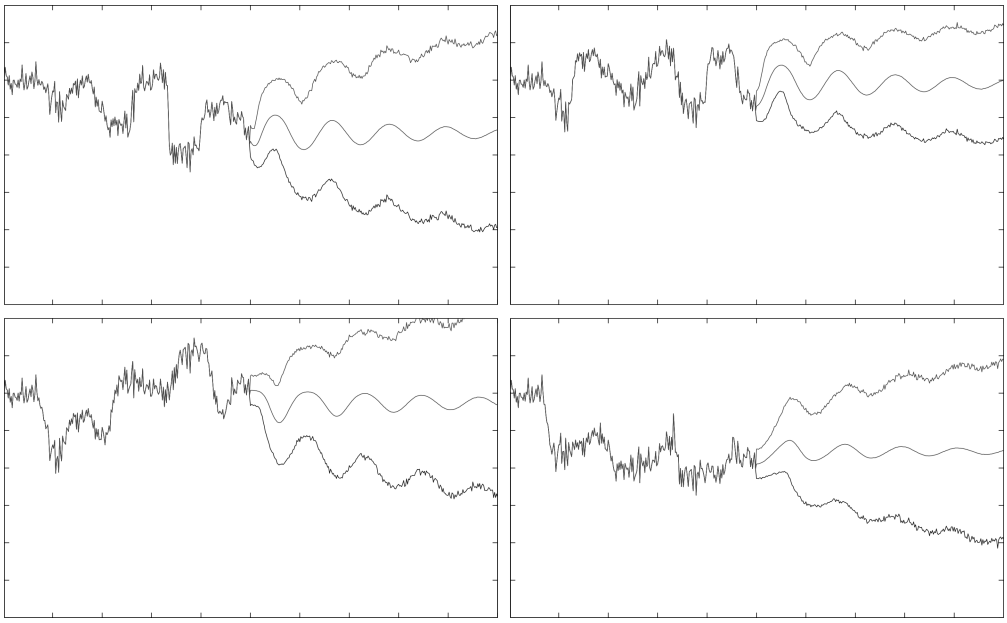

Figure 4: Simulated confidence intervals for the second half of one series for each of a reference DGP and three alternatives.

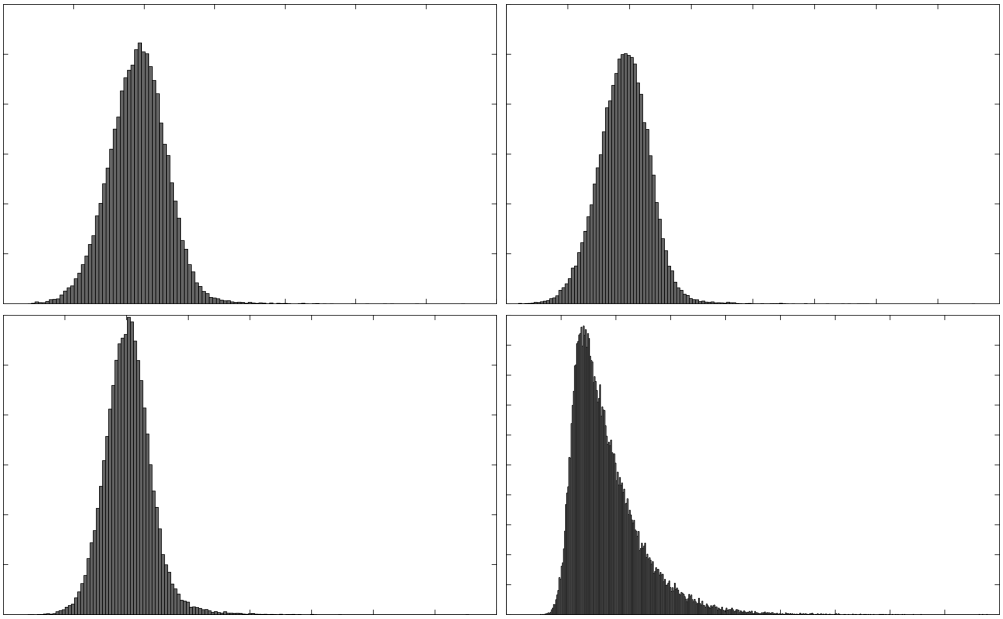

Figure 5: Simulated histograms for the length of a full cycle (consisting of an upward and a downward switch).

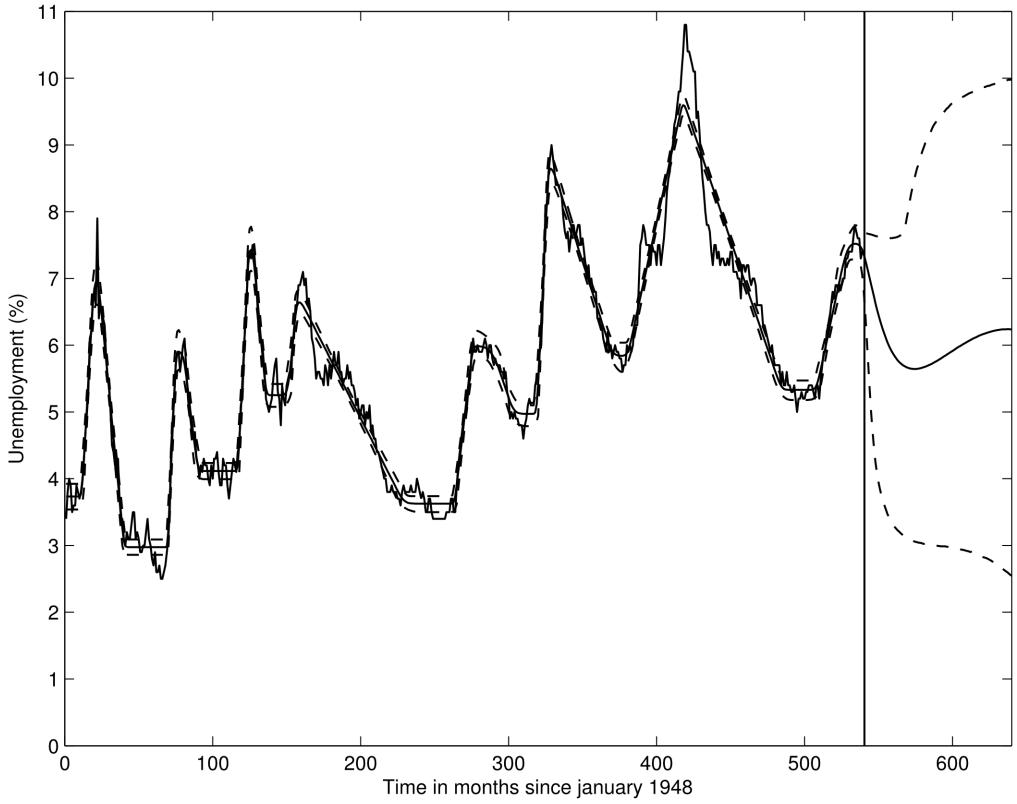

Figure 6: The mean and 95 % bounds for the estimation and forecasted observation means. Up until the vertical line at $t = 540$ the mean and bounds are in-sample and they are shown together with the original unemployment series for the period 1948 to 1992. For $t > 540$, mean and bounds of out-of-sample forecasts are shown, all constructed with the information set at $t = 540$.

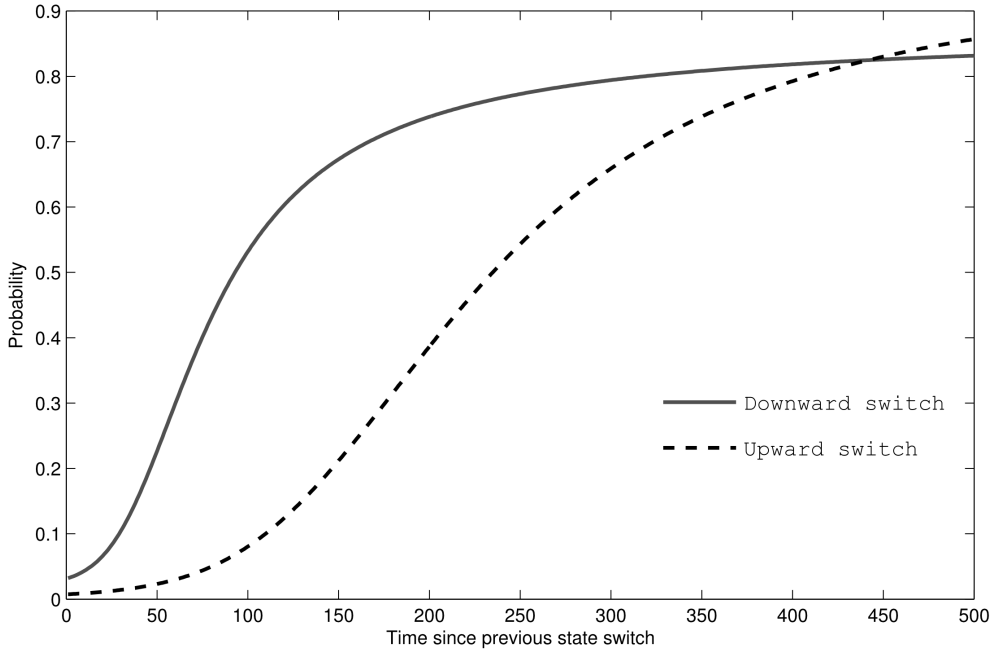

Figure 7: Estimated posterior probability of switching out of a state, accounting for the uncertainty in β by using the entire posterior distribution.

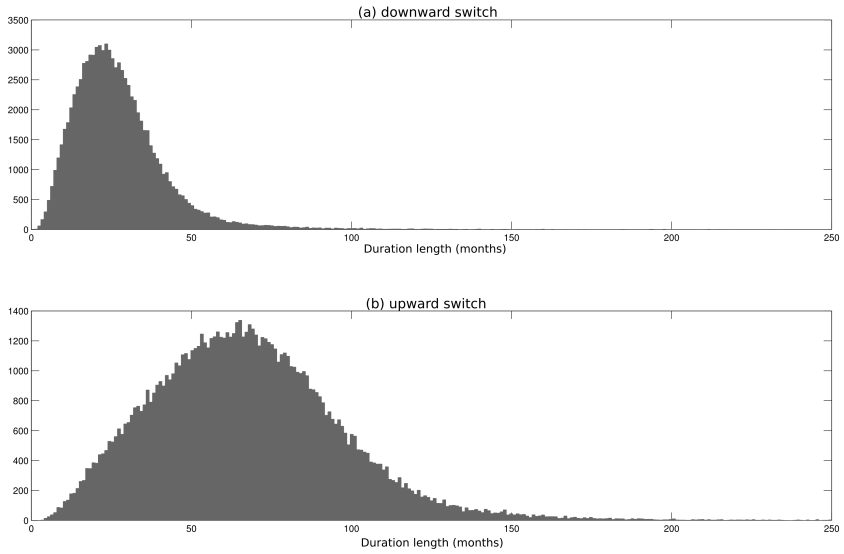

Figure 8: Estimated histograms for the duration of one state, accounting for the uncertainty in β by using the entire posterior distribution.

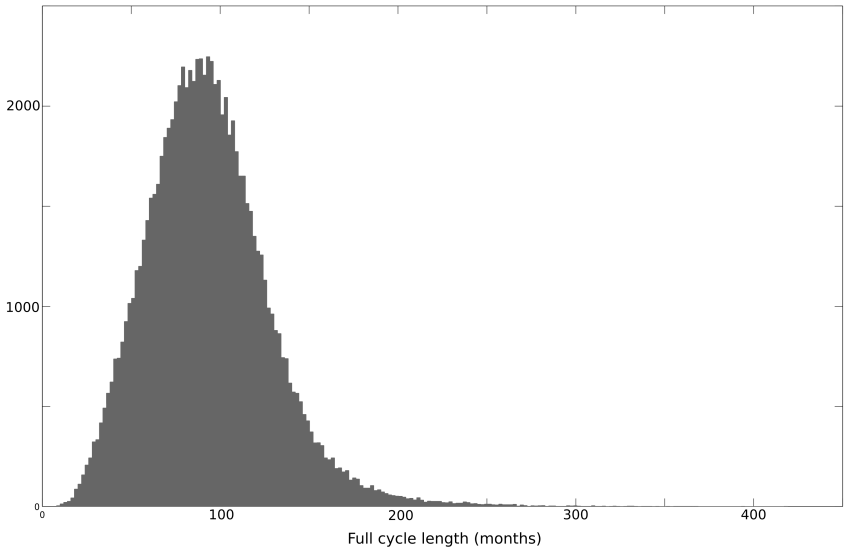

Figure 9: Estimated histograms for the length of one full cycle, which incorporates both one upwards switch and one downwards switch together with their transition periods.

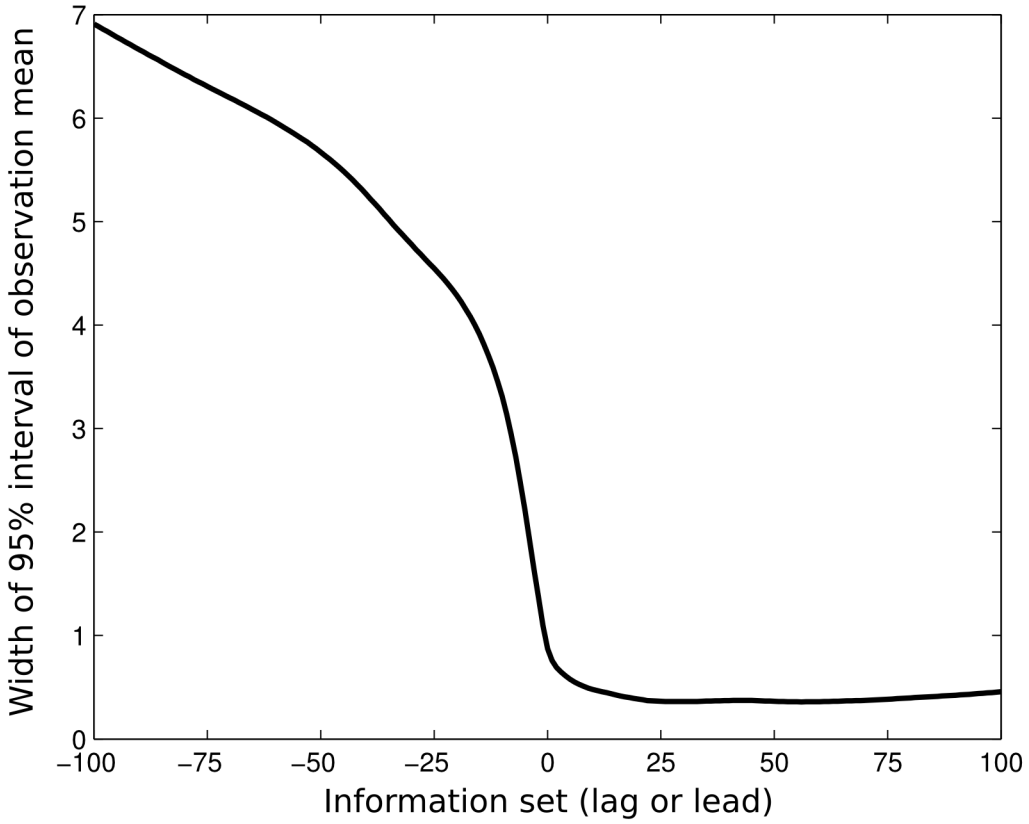

Figure 10: The average width of the 95 % interval of the observation mean for information sets that lead or lag with horizons up to $h = 100$.

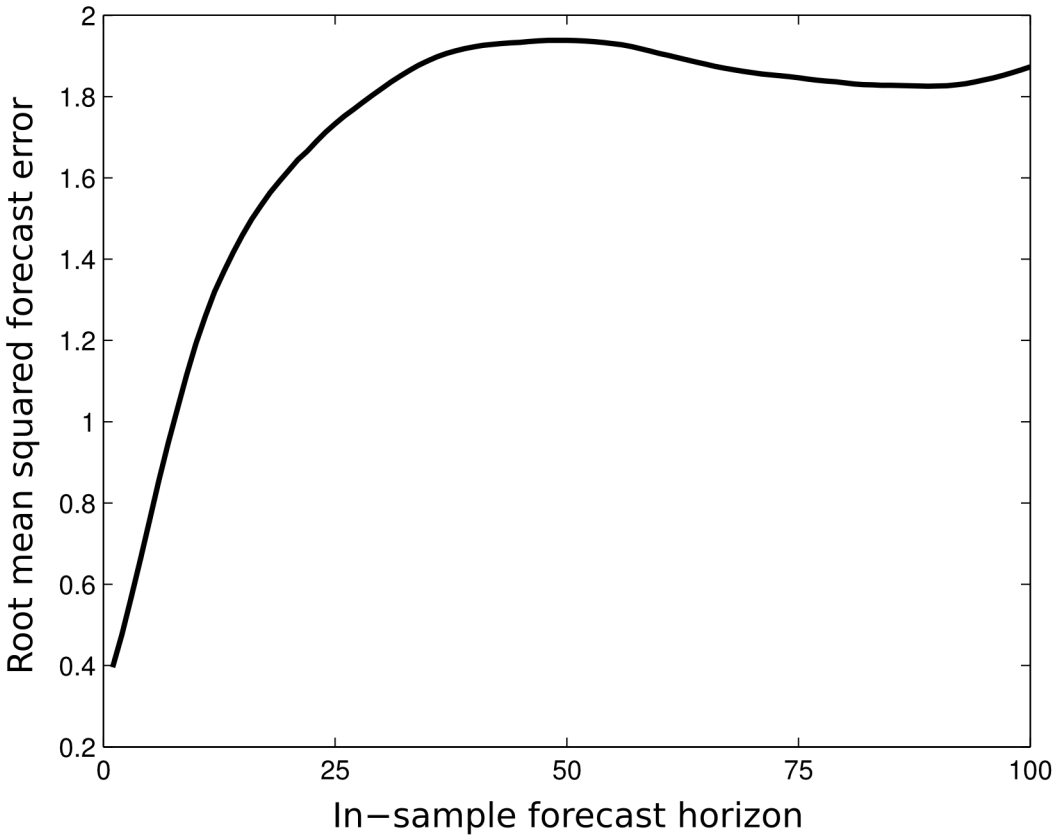

Figure 11: The in-sample root mean squared forecast error for horizons up to $h = 100$.