Past observation driven changing regime time series models for Forecasting Inflation

Eric Slob (357837)
Supervisor: Prof. dr. R. Paap
Second Reader: Dr. M. van der Wel

August 22, 2016

Abstract

In this paper we propose two time series models for inflation modelling. In both models the previous observation plays an important role for the dynamic structure. In the first model we have time-varying autoregressive parameters, which are dependent on the previous observation. The second model is a mixture model, where the regime probabilities are dependent on the previous observation. We compare the forecasts with a random walk model and a time-invariant autoregressive specification. Both models provide solid density forecasts. We find that combining the two models with an equal-weighing scheme, significantly improves the forecast quality.
Keywords: *inflation, forecasting, Bayesian, time series, mixture modelling, time-varying parameters.*

1 Introduction

Maintaining price stability is considered the best monetary policy a central bank can do to support long-term growth of the economy (Fischer et al., 1996). In order to keep the prices stable, the federal reserve system (FED) sets an explicit target inflation rate for the medium term[1]. In order to influence the inflation rate, the central bank will raise/lower the interest rate. General belief is that a lower interest rate will lead to an acceleration of the economy and hence an increase in inflation (Alvarez et al., 2001). Likewise, a higher interest rate will cool the economy down and lower the inflation. Being able to model and more importantly forecast inflation is of key importance to the central bank, so that they can adapt their interest rate policy. Inflation is also an important variable for other economic agents, such as pension funds and policy makers. A higher inflation leads to a larger cost of borrowing, a falling real income and more uncertainty in business confidence, since there is more uncertainty about prices and costs. Several contracts, such as wages and pension, have agreements on price compensation. A higher inflation leads to a higher compensation and hence it is important for the issuers of these contracts to

[1] See `http://www.federalreserve.gov/faqs/economy_14400.htm`. Accessed 18 July 2016.

have good inflation forecasts.

Since inflation plays an important role for decision making of both central bankers and other agents in the economy, being able to make a good inflation forecast is of great importance. Furthermore, in academic literature, predicting inflation is seen as a manner to get some grip on the characteristics of inflation dynamics.

The most obvious model to forecast inflation is an autoregressive (AR) model as described in Cox et al. (1981). This is a model where the dependent variable depends linearly on its own previous values and some residual term. Speed (1997) has used this model to describe inflation. The biggest problem of this model in practice is that it requires the parameters of the model to be constant over time, both the regression coefficients and the variance of the residual.

However, recent research suggests that the properties of inflation time series have changed over time, both in the mean and persistence of variance. The time-varying mean and persistence of inflation have been shown by Cogley and Sargent (2005), Benati (2004), O'Reilly and Whelan (2005) and Levin and Piger (2004) for respectively the United States, the United Kingdom, the EURO-area and the twelve main OECD economies. Sensier and Van Dijk (2004) found that for over 80% of 214 macroeconomic time series for the U.S. in the time period 1959-1999 most of the observed reduction in volatility is due to breaks in conditional volatility rather than conditional means. Furthermore, Sims and Zha (2006) assert that the time-variation of the dynamics in U.S. macroeconomic time series are entirely due to breaks in variance shocks. Hence, a good extension to improve the basic AR model would be to add structural breaks to the model to incorporate the changing properties of inflation time series. Bai and Perron (1998) presented tests for the presence of multiple structural changes and for the determination of the number of changes present.

For inflation data these structural break models have been estimated by several authors, including Bai and Perron (2003), Levin and Piger (2004) and Culver and Papell (1997). Allowing for a structural break is great for describing the data ex post, yet the main goal is to be able to forecast the inflation. For this purpose these models are not so useful, since one does not know when the next structural break will happen and hence forecasting inflation will fail.

Hamilton (1989) suggested that a possible way to incorporate the switching to another regime is by using a Markov-switching model. This allows one to have different regimes for different behaviour of the inflation rate over time. This has been done by Kim (1993), Simon (1996) and Bidarkota (2001) on inflation data. This way of modelling inflation is good for describing data, yet the big downside is that the switching is an independent random process. In these models the probability of a structural break is constant, whereas research suggests that both the mean and variance are changing over time, see, for example, Cogley and Sargent (2005), Benati (2004), O'Reilly and Whelan (2005) and Levin and Piger (2004). This means that there is little information for predicting in which regime one will be next period and this adds a lot of uncertainty to the model.

To remove this independent random process in the Markov-switching model, Tong and Lim (1980) proposed the threshold autoregressive (TAR) model. This model changes the regimes not randomly, but the regime change is based on the past value of a certain variable. Phiri (2013) has used this model type to describe the inflation in Zambia and Koirala (2012) applied it to the inflation series of Nepal. By adding behaviour to the regime switching, there is more certainty about the point forecasts made. One possible downside is that there exists discontinuity around the threshold what makes forecasts close to the threshold less certain.

Instead of this abrupt change, it is possible to propose a smoother transition function to remove this discontinuity around the threshold. This is the so called smooth transition autoregressive (STAR) model. The two most often used transition functions are the second order logistic function (LSTAR) and exponential function (ESTAR). This type of modelling was first suggested by Chan and Tong (1986). Arango and Gonzalez (2001) have used this model to describe the inflation in Colombia for the past decade. Another way to allow for a more flexible model is by introducing time-varying parameters. This is often done with a random walk for the parameters. Nadal-De Simone (2000) has made use of this to forecast inflation in Chile.

In this paper we propose two non-linear time series models for predicting inflation. In the first model we propose time-varying parameters in a similar fashion to Salimans (2012), yet here we make the parameters dependent on the previous observation of the dependent variable instead of some independent variable. This way of specifying our model allows us to avoid the problem of forecasting in which regime we are and time-varying parameters. The second model is a mixture model, where the mixture components are dependent on the previous observation. This makes it easier to forecast in which regime we are, so we have the advantage of having multiple regimes without the disadvantage of having a lot of uncertainty about the forecasting the regime. Both models make good point and density forecasts for inflation. We find that a linear combination of both models makes excellent point and density forecasts.

The rest of this paper is organized as follows. The proposed models are discussed in Section 2, in Section 3 we will discuss the data, the chosen priors and the posterior results. In Section 4 we use our proposed models to forecast inflation and compare these forecasts with the forecasts of often used models in literature. Finally, Section 5 concludes.

2 Model specification

In Section 2.1, we will start by describing a time series model for inflation where the autoregressive parameters evolve through time based on the level of the previous observation. Next, in Section 2.2 we will discuss an autoregressive mixture model where the observations are assumed to follow a finite mixture of K autoregressive models. The current mixture proportions are based on the level of the previous observation.

2.1 AR model based on previous observation

In economies with an unstable banking policy, we often observe inflation levels that keep on rising. To correct for this most models include some AR terms. Yet, often the behaviour of inflation levels seems to be non-linear. We correct for this non-linear behaviour by proposing an AR model where the parameters are influenced by the previous observation.

To incorporate this into a model we will start first with a standard AR model. This model has the *observation equation*

$$y_t = \alpha + \beta_{1,t} y_{t-1} + \ldots + \beta_{p,t} y_{t-p} + \varepsilon_t \text{ with } \varepsilon_t \sim \mathcal{N}(0, \sigma^2) \ , \tag{1}$$

where y_t is the dependent variable at time t, for $t = 1, \ldots, T$, α is the intercept, $\{\beta_{1,t}, \ldots, \beta_{p,t}\}$ is a collection of p autoregressive coefficients (which can vary over time) and ε_t is the error term.

Let

$$B_t = (\alpha, B_{1,t}, \ldots, B_{p,t}) \ ,$$

$$Y_{t-1} = (y_0, y_1, \ldots, y_{t-1}) \ ,$$

and

$$X_t = (1, y_{t-1}, \ldots, y_{t-p}) \ ,$$

then we can write the model as

$$y_t = X_t B_t + \varepsilon_t \ .$$

We allow the parameter B_t to change based on the level of the previous observation y_{t-1} in a non-linear way. We assume that the coefficients of B_t follow a normal distribution where the mean and variance are both dependent on y_{t-1}, with the assumption that $y_t > 0$ for all t. So we expect that if we had a large value for the observation at the previous moment in time, the parameter value in the next period will be larger and more volatile. With the larger parameter value we hope to capture the observed periods with high inflation. Furthermore, in these high inflation periods we see that the inflation is more volatile than normal, so we hope to capture this by allowing the parameter to be more volatile. This leads to the *transition equation*

$$B_t \sim \mathcal{N}(C y_{t-1} + \beta_0, D y_{t-1}^2) \ , \tag{2}$$

where C and D are parameters that need to be estimated, with $D \geq 0$. We can clearly see that the mean and variance are heavily influenced by the previous observation of y_{t-1}. The initial state B_0 can be given a fixed value or assumed to have a normal distribution with known mean and variance. Another possibility for B_0 is to condition on the first observation and let t run from 2 to T. Since we are dealing with AR models, it is quite likely that this observation would only be used as a independent variable anyway. We call the model in (1) and (2) the PREVOBS-AR model.

Now we have proposed a specification where the behaviour of the inflation is non-linear like we see in real-time data. One nice property of this specification is that we can estimate a non-linear model in a linear fashion. With this model, we hope to be able to describe clusters of inflation. Usually inflation is stable, however there are time periods where the inflation is large for a while. In such a time period, a normal AR model would in general underestimate the inflation in the next period in such a cluster (as it should slowly return to the average inflation), whereas our model should not have this issue since we inflate our parameters. Another often used method in the literature is to propose a random walk for β_t. This is good for describing data, since a part of the variance of σ will be captured through the β_t. Yet when one wants to make forecasts, our model provides more certainty about the value of β_t through the dynamic structure in comparison with this random walk method.

For inference we opt for the Bayesian approach in the next section.

2.1.1 Parameter Estimation

Posterior results will be obtained from the Gibbs sampler of Geman and Geman (1984). For the Gibbs sampler we will first need the complete data likelihood function. First we will derive the likelihood function for the *observation equation* and the *transition equation.* The likelihood for the *observation equation* is

$$f(Y|B_t, X_t\sigma^2) \propto (\frac{1}{\sigma^2})^{T/2}\prod_{t=1}^{T}\exp[-\frac{1}{2\sigma^2}(y_t - X_tB_t)'(y_t - X_tB_t)] \tag{3}$$

and for the *transition equation*

$$f(B_t|C, D, y_t) \propto \prod_{t=1}^{T}(\frac{1}{|Dy_{t-1}^2|})^{1/2}\exp[-\frac{1}{2}(B_t - Cy_{t-1})'D^{-1}y_{t-1}^{-2}(B_t - Cy_{t-1})] \ . \tag{4}$$

Now if we adapt standard priors for the remaining parameters: $1/\sigma^2 \sim \mathrm{Ga}(\alpha_0/2, \delta_0/2)$, $C \sim \mathcal{N}(0, \frac{1}{c_0}I)$ and $D^{-1} \sim \mathrm{W}_p(\nu_0, S_0)$ then the joint posterior distribution is given by

$$\begin{aligned}\pi(B_t, C, D, \sigma^2|y, X) \propto\ & (\frac{1}{\sigma^2})^{T/2}\prod_{t=1}^{T}\exp[-\frac{1}{2\sigma^2}(y_t - X_tB_t)'(y_t - X_tB_t)] \\ & \times \prod_{t=1}^{T}(\frac{1}{|Dy_{t-1}^2|})^{1/2}\exp[-\frac{1}{2}(B_t - Cy_{t-1})'D^{-1}y_{t-1}^{-2}(B_t - Cy_{t-1})] \\ & \times (\frac{1}{\sigma^2})^{\alpha_0/2-1}\exp[-\frac{\delta_0}{2\sigma^2}]\frac{1}{|D|^{(\nu_0-K-1)/2}}\exp[-\frac{1}{2}\mathrm{tr}(S_0^{-1}D^{-1})] \\ & \times \frac{1}{|\frac{1}{c_0}I|}\exp[-\frac{1}{2}C'c_0IC] \ .\end{aligned} \tag{5}$$

This implies the following full conditional posterior distributions for the model parameters, which closely resemble to those used in Chib (2001):

$$\begin{aligned} B_t|y_t, \sigma^2, C, D &\propto \mathcal{N}(\bar{\beta}_t, \tilde{\beta}_t), \\ (1/\sigma^2)|y, B_t &\propto \text{ Ga}(\alpha_1/2, \delta_1/2), \\ C|y, B_t &\propto \mathcal{N}(\bar{C}, \tilde{C}), \\ D^{-1}|y, B_t, C &\propto \text{W}_p(\nu_1, S_1), \end{aligned} \tag{6}$$

where

$$\begin{aligned} \tilde{\beta}_t &= [(1/\sigma^2)X_t' X_t + D^{-1} y_{t-1}^{-2}]^{-1}, \\ \bar{\beta}_t &= \tilde{\beta}[(1/\sigma^2)X_t' y_t + C' y_{t-1}(D^{-1} y_{t-1}^{-2})], \\ \alpha_1 &= \alpha_0 + T, \\ \delta_1 &= \delta_0 + \sum_{t=1}^{T} (y_t - X_t B_t)'(y_t - X_t B_t), \\ \tilde{C} &= (c_0 D I + Y_{t-1}' Y_{t-1})^{-1}, \\ \bar{C} &= \tilde{C} Y_{t-1} B_t, \\ \nu_1 &= \nu_0 + T, \\ S_1 &= [S_0^{-1} + \sum_{t=1}^{T} (B_t - y_{t-1}C)(B_t - y_{t-1}C)']^{-1}. \end{aligned} \tag{7}$$

This sampling scheme is given in Figure 1.

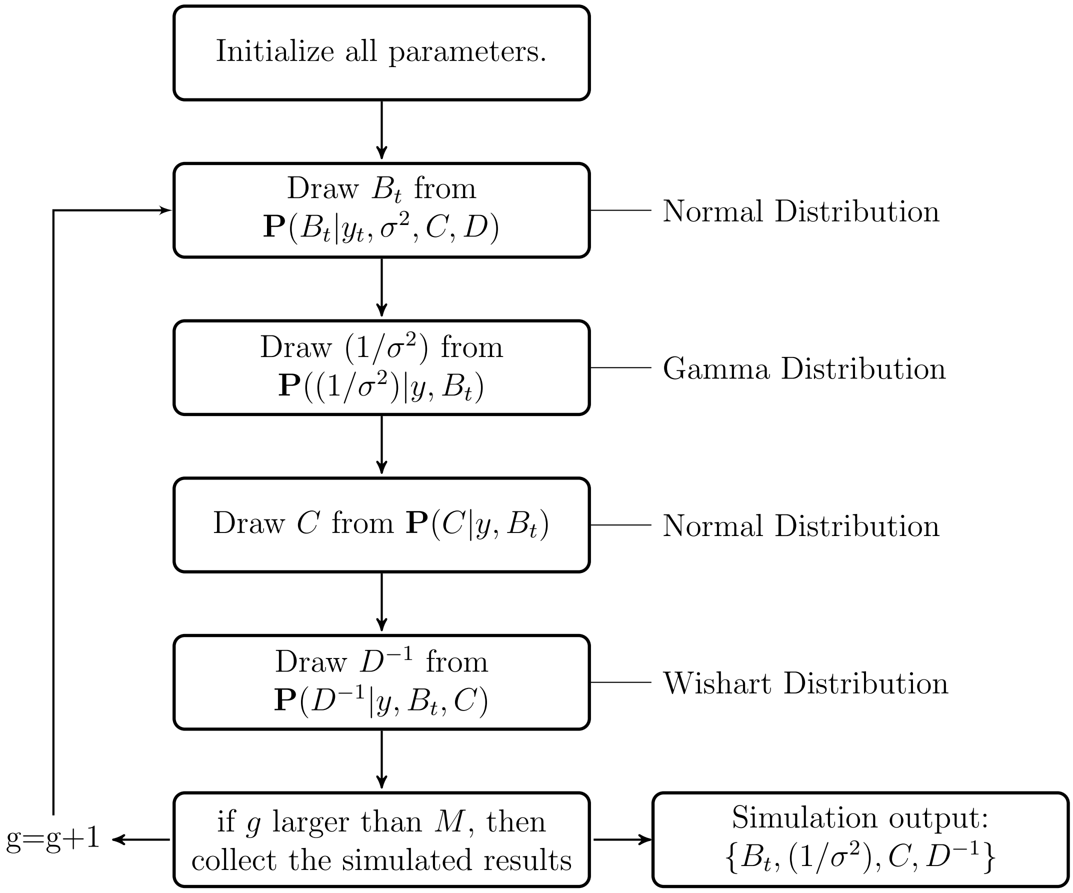


Figure 1: The sampling scheme for PREVOBS-AR. The order of steps is arbitrary, all nodes can be interchanged.

2.2 The mixture of autoregressive models

Gaussian mixture models are often used in statistics to represent the different groups within a population, see for example Kurita et al. (1992) and Yang and Ahuja (1998). Because the general version of this model has a lot of uncertainty in its forecasts due to the uncertainty about in what group we are, this model is not often used for modelling inflation. To remove the regime randomness, we will propose a special behaviour on the mixing probabilities. We believe that the previous observation often contains information on what kind of regime we are, hence we will use that as a proxy to determine in what component we are. This use of a Gaussian mixture gives the possibility to have a smooth transition between the different regions, like a STAR model.

Consider again an univariate, real valued time series y_t, which is observed at equally spaced moments in time $t = 1, 2, \ldots, T$. We are mostly interested in the forecasting of the next observation, so that is $p(y_t|Y_{t-1})$.

We assume that there are $K > 0$ different possible linear models from what y_t is generated, which are characterised by $\theta_k = (\phi_{k0}, \phi_{k1}, \ldots, \phi_{kp_k}, \sigma_k, \mu_k)$ with $k = 1, 2, \ldots, K$. For identifiability we use the restriction $\mu_1 < \ldots < \mu_k$. The selected model is denoted by $m_t = k$. We assume

$$y_t|(\mathcal{F}_{t-1}, m_t = k, \theta_k) = \phi_{k0} + \phi_{k1} y_{t-1} + \ldots + \phi_{kp_k} y_{t-p_k} + \varepsilon_t \text{ with } \varepsilon_t \sim \mathcal{N}(0, \sigma_k^2) \ , \tag{8}$$

where m_t is selected from the different possible models $k = 1, \ldots, K$ and $\mathcal{F}_{t-1}$ is the information set up to time $t-1$. For the chance of model k being selected, we choose a structure that yields odds similar to a probit model, such that

$$p(m_t = k|\theta_1, \ldots, \theta_n, y_{t-1}) \propto \Phi(-|y_{t-1} - \mu_k|) \ , \tag{9}$$

where $\Phi(x)$ is the cumulative Distribution Function of the standard normal distribution. Observe that this specification is symmetrical in μ_k. Now we define

$$\alpha_{kt} = \frac{p(m_t = k|\theta_1, \ldots, \theta_n, y_{t-1})}{\sum_{k=1}^{K} p(m_t = k|\theta_1, \ldots, \theta_n, y_{t-1})} \ , \tag{10}$$

such that the total probability sums to one ($\sum_{k=1}^{K} \alpha_{kt} = 1$). Combining (8), (9), (10) gives the K-component mixture autoregressive model

$$F(y_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \alpha_{kt} \Phi(\frac{y_t - \phi_{k0} - \phi_{k1} y_{t-1} - \ldots - \phi_{kp_k} y_{t-p_k}}{\sigma_k}) \ . \tag{11}$$

Observe that the AR order can be different across the mixture components. Now let $p = \max(p_1, \ldots, p_k)$.

The model has several interesting properties. The model has a conditional distribution that changes over time, since the conditional means of the different components depend on the previous observations. The conditional expectation of y_t is given by

$$E(y_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \alpha_{kt} (\phi_{k0} + \phi_{k1} y_{t-1} + \ldots + \phi_{kp_k} y_{t-p_k}) = \sum_{k=1}^{K} \alpha_{kt} \lambda_{kt} \ . \tag{12}$$

Furthermore, the model is able to account for changing conditional variance, since it depends on the conditional means of the components. The conditional variance of y_t is given by

$$\text{var}(y_t|\mathcal{F}_{t-1}) = \sum_{k=1}^{K} \alpha_{kt} \sigma_k^2 + \sum_{k=1}^{K} \alpha_{kt} \lambda_{kt}^2 - (\sum_{k=1}^{K} \alpha_{kt} \lambda_{kt})^2 \ . \tag{13}$$

Observe that $\sum_{k=1}^{K} \alpha_{kt} \lambda_{kt}^2 - (\sum_{k=1}^{K} \alpha_{kt} \lambda_{kt})^2$ is non-negative and 0 only if $\lambda_{1t} = \lambda_{2t} = \ldots = \lambda_{Kt}$. If the λ_{kt} differ greatly, the variance of y_t is large and the model might be multimodal instead of unimodal.

Now we have a mixture model what is able to describe the different clusters we observe for inflation. In contrast to a normal mixture model, our model is able to make better forecasts on what component we are in and hence reduce the uncertainty. Furthermore, the model should be able to capture the in practice observed changing mean and variance of inflation, since the components allow for this.

For inference we opt for the Bayesian approach. First, in the next section we will discuss the used priors.

2.2.1 Priors

To be able to use the model for forecasting, we need to estimate several parameters. For each component we need to estimate $\phi_{k0}, \phi_{k1}, \ldots, \phi_{kp_k}, \sigma_k$ and μ_k for $k = 1, \ldots, K$.

We take for ϕ_k as a prior the Zellners G-prior distribution from Marin and Robert (2007). This corresponds to $\mathcal{N}(0, c\sigma_k^2(X_p'X_p)^{-1})$, where $c = n$ and

$$X_p = \begin{pmatrix} 1 & y_P & y_{P-1} & \cdots & y_{P-p-1} \\ 1 & y_{P+1} & y_P & \cdots & y_{P-p-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & y_{T-1} & y_{T-2} & \cdots & y_{T-p} \end{pmatrix} .$$

For σ_k^2 we take the inverse gamma distributions as priors, with parameters a and b. This choice of variance is to show that we have very vague knowledge on σ_k^2. Since we have little knowledge about the means of the clusters, the priors for μ_k are $\mathcal{N}(0, \tau)$.

2.2.2 Parameter Estimation

To estimate the model we will use an algorithm similar to the EM algorithm from Dempster et al. (1977). Suppose that the observations $Y = (y_1, \ldots, y_T)$ are generated from (11). Let $Z = (Z_1, \ldots, Z_T)$ be the unobserved random variable, where Z_t is a K-dimensional vector with the kth element equal to 1 if y_t is generated from component k and 0 otherwise.

The (conditional) likelihood is given by

$$\begin{aligned} L &\propto \prod_{t=p+1}^{T} L_t \\ &\propto \prod_{t=p+1}^{T} \sum_{k=1}^{K} z_{kt}\alpha_{kt}\frac{1}{\sqrt{2\pi\sigma_k}}\exp\left[-\frac{1}{2\sigma_k^2}(y_t - \phi_{k0} - \phi_{k1}y_{t-1} - \ldots - \phi_{kp_k}y_{t-p_k})^2\right] . \end{aligned} \tag{14}$$

Observe that the likelihood consists of three parts. The first part z_{kt} is a latent variable for the regime. The second part α_{kt} is the chance that the regime k is selected at time t. The last part is the observation equation.

2.2.3 Sampling

The algorithm produces estimates for the parameters using a sampling scheme what comes quite close to the EM algorithm. The sampling scheme and estimators are similar to those used by Wood et al. (2011). The scheme consists of the following steps.

- Initialize Z, such that $\sum_{k=1}^{K} z_{kt} = 1$.
- Draw the lag p from the multinomial distribution $P(p|y, K, Z)$.
- For $k = 1, \ldots, K$ draw σ_k^2 from the inverse gamma distribution $P(\sigma_k^2|y, z, K, p)$.
- For $k = 1, \ldots, K$ draw ϕ_k from the multivariate normal distribution $P(\phi_k|\sigma_k^2, z, y, p)$.
- For $k = 1, \ldots, K$ sample μ_k using the slice sampling.
- Draw z_{kt} from the multinomial distribution $P(z_{kt}|\phi_k, \sigma_k^2, \mu_k, K, p)$, for $k = 1, \ldots, K$, $t = 1, \ldots, T$,

where

- $p(p|y, K, Z) \propto c^{-K(p+1)/2}|X_p'X_p|^{K/2} \prod_{k=1}^{K} |X_p'Z_kX_p + c^{-1}X_p'X_p|^{-1/2}b_k^{-a_k}$, where $Z_k = \text{diag}(z_t, t = p+1, \ldots, T)$, $a_j = \frac{1}{2}\sum_{t=p+1}^{T} z_t + \alpha$ and $b_j = \frac{1}{2}yM_ky + \beta$, where $M_k = Z_k - Z_kX_p(X_p'Z_kX_p + c^{-1}X_p'X_p)^{-1}X_p'Z_k$,
- $p(\sigma_k^2|y, z, K, p) \propto \text{Ig}(\sum_{k=1}^{K}\sum_{t=1}^{T} z_{tk} + \alpha, \frac{1}{2}y'M_ky + \beta)$,
- $p(\phi_k|\sigma_k^2, z, y, p) \propto \mathcal{N}((X_p'Z_kX_p + c^{-1}X_p'X_p)X_p'Z_ky, \sigma_k^2(X_p'Z_kX_p + c^{-1}X_p'X_p))$,
- Let $p_{kt} = p(y_t|x_{t-1}; \phi_k, \sigma_k^2)$, where $x_{t-1} = (y_{t-1}, \ldots, y_{t-p})'$. Draw the indicators for z_{kt} with $p(z_{kt} = 1|y_t, x_{t-1}; \phi_k, \mu_k, \sigma_k^2) \propto \frac{\alpha_{kt}p_{kt}}{\sum_{k=1}^{K}\alpha_{kt}p_{kt}}$ for $k = 1, \ldots, K$, $t = 1, \ldots, T$.

The full conditional posterior of μ_{kt} is of an unknown form. Finding an appropriate candidate density for a Metropolis Hastings sampler is not straightforward. Therefore we opt for the slice sampler of Neal (2003).

The idea of this technique is to draw uniformly under the curve of the distribution $f(x)$. This is done as follows: first one chooses a starting value x_0 for which $f(x_0) > 0$. Next one should draw a value y_i uniformly on the interval 0 to $f(x_0)$. The next step is to draw a horizontal line across the curve at this y_i value. Across this horizontal line one should sample the next x_i uniformly within the curve. The sample point is now x_i. Now one should use the new x_i as a starting point to generate the next y_i and x_{i+1}. For a more detailed explanation see Neal (2003). The sampling scheme is shown in Figure 2.

Now we have a full sampling scheme, only we need to find a way to choose for how many components we allow in the final model. This will be done by cutting the sample in two parts, a training sample and a forecasting sample. We use the training sample to do parameter estimations for all possible K. Now we look at which K gives the best forecasts in the forecasting sample and chose the optimal K according to the forecasts. Note that one could also use a reversible-jump MCMC to estimate K. Since we believe that we can best use the out-of-sample data to choose K, we opt for the forecasting approach .

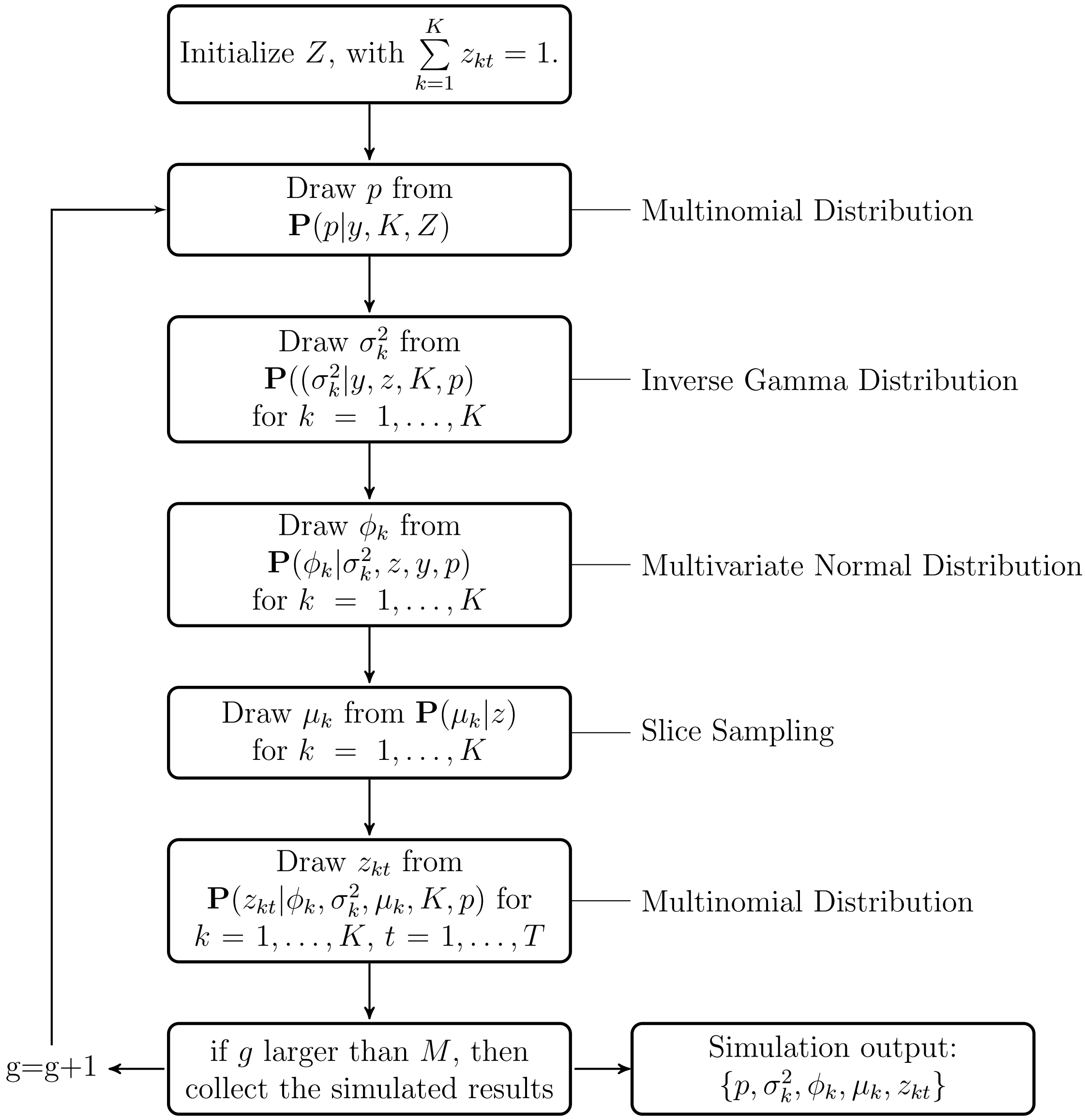


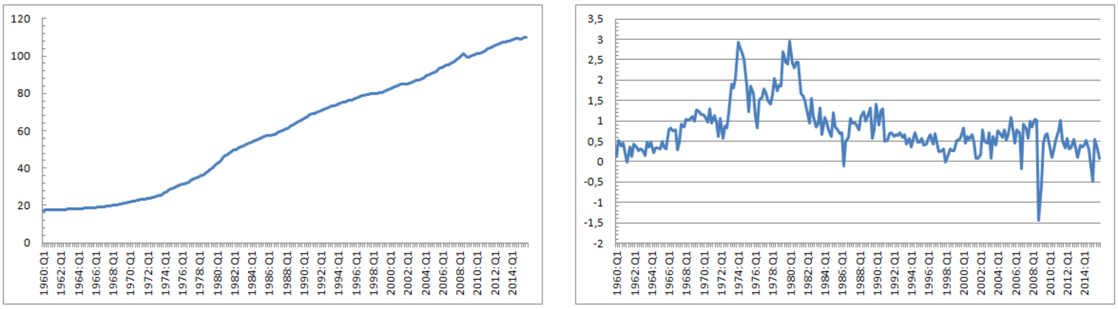
Figure 2: The sampling scheme for the mixture model.

3 Data, Priors, Posterior Results and implied Characteristics

In this section we will discuss how we will operationalize the models on our dataset with an aim to describe the post-WWII behaviour of inflation measures in the US. We discuss the data in Section 3.1, whereas in Section 3.2 we report our prior choices. In Section 3.3 we will review the posterior results. Last in Section 3.4 we will discuss some of the implied characteristics according to the models.

3.1 Data

We will consider a quarterly observed seasonally unadjusted US inflation series for the period 1960Q1-2015Q4. As a proxy for the inflation we use the gross domestic product (GDP) deflator from the Real-Time Data Set for Macroeconomists (RTDSM) of the Federal Reserve Bank of Philadelphia. This is the same dataset as used by Groen et al. (2013), where we use a longer horizon of the dataset. Since for inflation the relative change in the deflator is the most interesting (levels do not really have a clear interpretation), we will model the quarterly log change. Since we are mostly interested in forecasting in *real time*, we will use the first releases to form the time series and hence revisions are ignored.



(a) Level of deflator (b) Ln difference

Figure 3: A graph of the quarterly absolute values and relative change of GDP Deflator from 1960Q1 to 2015Q4.

Figure 3a displays the GDP deflator is shown over the full sample period. We use the first releases of the variable, such that we have data up to 2015Q4. Since we are interested in the relative change of the inflation, we take the difference of logarithm of the series multiplied by 100 to get the percentage change[2]. The resulting series is shown in Figure 3b. In this figure different regimes of the inflation can be noticed. For much of the 1960s the inflation is stable. In 1958, Phillips (1958) came up with the now-infamous 'Philliphs Curve'. This paper linked a high inflation to a low unemployment. The Federal Reserve used this curve to adapt their monetary policy, what caused a period with stable inflation. After this stable period, we see that the inflation rates are fluctuating more or less out of control. The short term relation between the inflation and unemployment is

[2]So the value we use in our models from now on will be $y_t = 100(\ln(\text{inflation}_t) - \ln(\text{inflation}_{t-1}))$.

not observed in the long run, so the economists at the FED did not know how to properly adapt their interest policy. We observe that after 1980 the economy enters a period of low and stable inflation. This sudden decrease in inflation is caused by the raise of interest rates to 20% by the FED chairman. From here up to the financial crisis in 2007, we see a very stable inflation. In 2007 we observe some deflation due to the crisis. Next we see that in the final year of the data, we have an inflation rate close to zero. So our sample can be divided in roughly five periods: first a stable period, next a high inflation period, than another stable period followed by a large drop, than another stable period and finally a near-zero inflation period.

3.2 Priors

To describe the inflation series we perform a Bayesian analysis on the two time series models discussed in Section 2.1 and 2.2. Before conducing Bayesian approach we need to specify our prior settings. In the PREVOBS-AR model we used $\alpha_0 = 2.001$, $\delta_0 = 1$, which is rather uninformative about σ^2. For c_0, we set the prior at 0.5, which is a rather vague prior. Since we also lack clear information about D, we set $v_0 = 2$ and $S_0 = 10$. In the mixture model we also have very vague information about σ^2, such that we choose priors $a = b = 0.05$. For the cluster means μ_k we set $\tau = 2$.

The influence of the chosen priors turns out to be relatively small, since the prior means of most parameters were a priori set at 0 and the prior variance was chosen large. This makes the parameter estimate shrink towards 0 unless the data strongly suggests it to be different from 0. Of course we could have made the variances smaller to get even closer to zero. Unreported results show that this has little effect on the posterior results (we only find little shrinkage effect towards 0). Hence, the information of the data seems to dominate the posterior results.

3.3 Posterior Results

We estimate the PREVOBS-AR model for $p = 1, \ldots, 4$ for the data up to 1999Q4 and with posterior results we construct predictive forecasts for 2000Q1 up to 2009Q4. It turns out that for $p = 2$ both the one quarter and one year ahead forecasts are best using the root mean squared forecast error to evaluate means as point forecasts. We estimate the model with a time varying α_t. However, it turns out that there is no significant evidence that α_t is influenced by y_{t-1}. Furthermore, if we include this into the model the forecast are less accurate. Hence, we did not include this into our model. We also propose to add y_{t-1} to the model in order to capture the non-linear effect. Both β_0 and C for this term are not significant. The inclusion does not improve our forecasts and since we prefer a simple model over a more complicated one, we do not include this term into the model.

Now we re-estimated the models for up to 2009Q4. The posterior means for the parameter estimates are given in Table 1. The small standard deviations exhibited in Table 1 indicate that ergodicity is a reliable paradigm for 5000 iterations (following burn-in of 500 iterations).

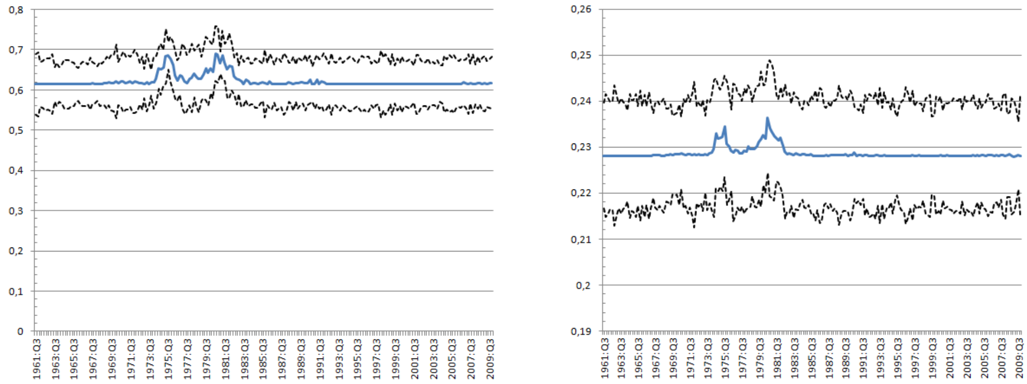
In Table 1 we observe that α and both $\beta_{0,1}$ and $\beta_{0,1}$ are positive. This suggest that a large shock will be followed by above average values in the next periods too. What is interesting to see is that C_1 is positive, meaning that if we get a large previous observation, the parameter in the next period will be larger, since we get a bigger $\beta_{t,1}$. Furthermore observe that D_1 is positive, meaning that a large inflation will also lead to a more volatile $\beta_{t,1}$. For $\beta_{t,2}$, we see that the average value $\beta_{0,2}$ is positive yet smaller than $\beta_{0,1}$ and the coefficient C_2 is also smaller than C_1. This suggests that the past observation contains more information than the observation from two periods ago. We see that for the variance of $\beta_{t,2}$ there is basically no influence by the previous observation. Observe that the sum of $\beta_{t,1}$ and $\beta_{t,2}$ is close to one, what suggests that there is persistence of inflation.

In Figure 4a and 4b, the average posterior values and 95% HPD for $\beta_{1,t}$ and $\beta_{2,t}$ are shown. We observe that the parameter values are relatively constant, yet in the turbulent years between 1973 and 1983, they have much larger values, especially $\beta_{1,t}$.

Table 1: The posterior mean, the upper and lower bound for the 95% HPD of the variables for the PREVOBS-AR model using data up to 2009Q4.

	Mean	HPD_{lower}	HPD_{upper}
α	0.2484***	0.2295	0.2673
$\beta_{0,1}$	0.6082***	0.5313	0.6851
$\beta_{0,2}$	0.2275***	0.2153	0.2379
C_1	0.0100***	0.0061	0.0182
C_2	0.0006**	0.0001	0.0018
D_1	0.0053***	0.0039	0.0083
D_2	0.0008	0	0.000879

Note: *,** and *** respectively mean 0 is not included in the 90%, 95% and 99% HPD region.



(a) $\beta_{1,t}$ (b) $\beta_{2,t}$

Figure 4: The graphs of the posterior means and the 95% HPD of $\beta_{1,t}$ and $\beta_{2,t}$ in the PREVOBS-AR model over the estimation period 1961Q3 up to 2009Q4.

For the mixture model we apply the same procedure with estimating posterior coefficients for data up to 1999Q4 and forecasting for 2000Q1 up to 2009Q4 for $p = 1, \ldots, 4$ and $k = 1, \ldots, 4$. This suggests two clusters and two lags, so $k = 2$ and $p = 2$.

Now we again re-estimated the model for up to 2009Q4. The posterior results are shown in Table 2. The small standard deviations exhibited in Table 2 indicate that ergodicity is a reliable paradigm for 5000 iterations (following burn-in of 500 iterations).

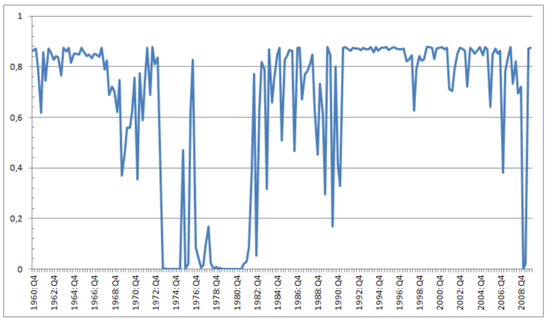
We observe that the first regime with lower values of y_{t-1} has a smaller ϕ_1 and a larger ϕ_2. The σ_k is smaller. These facts together suggest that it is a more stable regime. The second regime has larger ϕ_1, a smaller ϕ_2 and a larger σ_k. This suggests that when we have large inflation in the previous quarter, the next quarter will probably also have a larger inflation. Hence, we have a stable cluster 1, characterized by low inflations, and a more extreme cluster 2, characterized by high inflations.

In Figure 5 the posterior probability of each component is shown. We see that indeed the first mixture component is for the stable periods and the second mixture is for the more volatile periods. This can especially be noticed by inspecting the mixture probabilities between 1973 and 1980. In this period the inflation was high and posterior probability of being in cluster 1 is most of the time very close to 0.

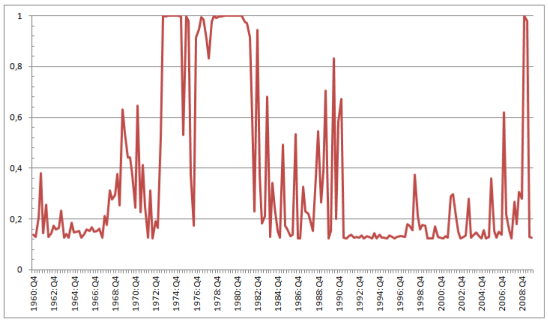
Table 2: The posterior means, the upper and lower bound for the 95% HPD of the parameters for the mixture model. *** means significant at 1 % significance level.

	Mean	HPD_{lower}	HPD_{upper}
μ_1	0.647	0.635	0.659
μ_2	1.313	1.301	1.325
$\phi_{1,0}$	0.1579***	0.1570	0.1588
$\phi_{1,1}$	0.5251***	0.5244	0.5258
$\phi_{1,2}$	0.2440***	0.2433	0.2447
$\phi_{2,0}$	0.3140***	0.3131	0.3149
$\phi_{2,1}$	0.7302***	0.7292	0.7313
$\phi_{2,2}$	0.0615***	0.0605	0.0626
σ_1	0.304	0.299	0.309
σ_2	0.851	0.844	0.858

Note: *** means 0 is not included in the 99% HPD region.



(a) Mixture 1



(b) Mixture 2

Figure 5: The posterior probability for mixture component 1 and 2 for time period 1960Q4 to 2009Q4.

3.4 Implied Characteristics

The PREVOBS-AR model suggests that the parameter values are influenced by the previous observation in a positive direction. With this we mean that on average a large value will lead to larger parameters and hence a larger observation in the next period. This is can be seen in the data in Figure 3b by the clusters of large inflation and small inflations. So this might be a explanation for the inflation clustering we have.

The mixture model suggests that we have two kinds of regimes. First of all we have a stable regime, where inflation over time is quite constant. This can also be observed from Figure 3b. Most of the time the inflation levels are quite stable. The second regime is characterised by very large inflation levels with large variance. According to Figure 3b, this corresponds to the regime of large inflation rates in the 1970s.

4 Real-Time Prediction of U.S. Inflation Rates

The main goal of our paper is to forecast inflation. In this section we will focus on making out-of-sample forecasts. We will make current quarter forecasts and one year ahead forecasts, so that is for $t + 1$ and $t + 5$. First in Section 4.1, we will discuss some simple models which will serve as a benchmark for the performance of our models. The PREVOBS-AR and mixture model seem to perform well in different periods. Hence we will combine the forecasts of the PREVOBS-AR and mixture model. The different ways of combining will be discussed in Section 4.2. There are several ways of producing multiple step ahead forecasts, this will be discussed in Section 4.3. Next in Section 4.4 the methods used to evaluate the methods will be discussed. In Section 4.5 we will compare the forecasting performance of our developed models and the benchmark models.

4.1 Forecast models

As a starting point, we will use the models discussed in Section 2.1 and 2.2. Here we will use the specification of the models which were best according to Section 3.3.

Since we are not only interested in how the models work in comparison with each other, we will also include a random walk (RW) model. Since this model is often hard to beat, this model will serve as a benchmark. We will use the specification from Atkeson and Ohanian (2001). This model assumes that the best forecast in the next period is the average over the past 4 quarters, such that

$$y_{t+1} = \frac{1}{4}\sum_{j=0}^{3} y_{t-j} + \varepsilon_{t+1} \text{ with } \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2) \ . \tag{15}$$

Another model used as benchmark will be the time-invariant autoregressive specification for inflation, where we use the lag orders between 1 and 4:

$$y_{t+1} = \beta_0 + \sum_{j=0}^{p*} \beta_j y_{t-j} + \varepsilon_{t+1} \text{ with } \varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2) \ , \tag{16}$$

where $p*$ is the optimal lag order according to the Bayesian-Schwarz information criterion (BIC) across lag orders up till 4. We will call this model the AR-BIC model. Both benchmark model parameters will be estimated so we can compare them with the other models.

4.2 Combining forecasts

In this section we will discuss the two different methods we use to forecast. We will use the equal-weighted forecast (Equal-Combination) and the time-varying weights approach by Hoogerheide et al. (2010) (TVW-Combination).

The equal-weighted forecast combines the density forecasts of models $1, \ldots, M$ and gives weight $\frac{1}{M}$ to the forecast of each model. This can be written as

$$p_{equal-combination}(y_{t+h}|\mathcal{F}_t) = \sum_{m=1}^{M} \frac{1}{M} p_m(y_{t+h}|\mathcal{F}_t) \ , \tag{17}$$

where $p_m(y_{t+h}|\mathcal{F}_t)$ is the density forecast of model m for y_{t+h} with information up to time t.

The time-varying weights approach combines the density forecasts of models $1, \ldots, M$ and gives weight time-varying weights to each model. These weights w_m are chosen such that they minimize the distance between the vector of observed values $y_{1:T}$ and the space spanned by the constant vector and the vectors of 'predicted' values $\hat{y}_{1:T,m}$ for model m. The weights are assumed to evolve over time in the following fashion:

$$w_t = w_{t-1} + \psi_t \quad \text{ with } \psi_t \sim \mathcal{N}(0, \sigma) \ .$$

This leads to the predictive density equation:

$$p_{TVW-combined}(y_{t+h}|\mathcal{F}_t) = w_{t+h,0} \sum_{m=1}^{M} w_{t+h,m} p_m(y_{t+h}|\mathcal{F}_t) \ . \tag{18}$$

To estimate the weights in (18) the Kalman filter is used, see Hoogerheide et al. (2010) for further details.

4.3 Forecasting Approaches

In this section we will discuss forecasting with the models. The one-step ahead predictive distribution $F(y_{t+1}|\mathcal{F}_t)$ is easy to compute using the observation equations.

However, the m-step ahead predictive distribution is not that easy to calculate. Granger and Terasvirta (1993) gave some fruitful ideas for the m-step forecast. We will discuss three different approaches for the m-step density forecast, the direct, the exact and the Monte Carlo approach.

For the direct density forecast, we pretend that the h step point forecast $\hat{y}_{t+h}$ is the true value of y_{t+h}. So we get

$$\mathcal{F}(y_{t+h}|\mathcal{F}_t) = \mathcal{F}(y_{t+h}|\mathcal{F}_t, y_{t+h-1} = \hat{y}_{t+h-1}, \ldots, y_{t+1} = \hat{y}_{t+1}) \ .$$

This is a very easy to compute forecast, yet some crucial information from the shape of the predictive distribution $\mathcal{F}(y_{t+h-1}|\mathcal{F}_t)$ is not included in the forecast of y_{t+h}. This is mostly a problem when $\mathcal{F}(y_{t+h}|\mathcal{F}_t)$ is multimodal.

To include this missed information, we get to the exact approach. This approach calculates the exact distribution using an integral. The exact predictive distribution is given by

$$\mathcal{F}(y_{t+h}|\mathcal{F}_t) = \int \mathcal{F}(y_{t+h}|\mathcal{F}_t, y_{t+h-1}, \ldots, y_{t+1})\, \mathrm{d}\mathcal{F}(y_{t+h-1}, \ldots, y_{t+1}|\mathcal{F}_t) \ .$$

We might not be able to exactly calculate this integral, yet we can still use numerical methods to evaluate it.

Alternatively, one could take the Monte Carlo approximation. The results from this method are often almost exactly the same as if one would take the exact distribution. The predictive distribution is given by

$$\mathcal{F}(y_{t+h}|\mathcal{F}_t) = \frac{1}{M}\sum_{i=1}^{M} \mathcal{F}(y_{t+h}|\mathcal{F}_t, \{y_{t+h-1}, \ldots, y_{t+1}\})^{(i)}) \ ,$$

where $\{y_{t+h-1}, \ldots, y_{t+1}\})^{(i)}$ are sampled from $\mathcal{F}(y_{t+h-1}, \ldots, y_{t+1}|\mathcal{F}_t)$.

In this paper we opt for the Monte Carlo approach, since the exact method is very hard to compute. Although Marcellino et al. (2006) states that it is still unclear if direct or indirect forecasts give more accurate forecasts. Since we are interested in density forecasts, we believe that the Monte Carlo approach is the best way to evaluate the forecasts from each model.

4.4 Evaluation methods

We will use the models described in the previous subsection to make and evaluate one-quarter and one year ahead forecasts for the GDP deflator in the United states for time period 2010Q1 up to 2015Q4. We will use several measures to evaluate the accuracy of our predictions. First of all we will use the square root of the mean squared forecast error (RMSE) and the mean of the absolute forecast errors (MAE). These can be written down as

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{1}{T - t_0 - h}\sum_{s=s_0-1}^{T-h} \hat{\epsilon}_{s+h}^2} \ , \tag{19}$$

and

$$\text{MAE} = \frac{1}{T - t_0 - h}\sum_{s=s_0-1}^{T-h} |\hat{\epsilon}_{s+h}| \ , \tag{20}$$

where $\hat{\epsilon}_{s+1}$ is the out-of-sample forecast error of a model for y_{t+h}. Gneiting (2011) found that MAE is only a consistent measure when the point forecast is equal to the median of the distribution of forecasts and for the RMSE only when the forecast is equal ot the mean of the distribution of the forecast.

For some distributions, such as the RW and AR-BIC model, the median and mean are equal. Yet for other models, such as those where forecasts are based upon the posterior draws of the Gibbs sampler, this is not the case. There we will base the RMSE and MAE on the mean and median of the distribution of inflation predictions.

One big downside of point forecasts is that they do not incorporate how certain we are about our forecasts. To evaluate how certain we are about our forecasts and how good models are to predict extreme events, we will use density forecast evaluations. There are several possibilities to measure this. The most often used density forecast evaluation method is the log score, since this approximates the likelihood function of a model. One big drawback of this method is that it is sensitive to outliers and does not reward values that are close but not equal to the realization (as shown in Gneiting and Raftery (2007)).

Gneiting and Raftery (2007) and Gneiting and Ranjan (2011) therefore propose the continuous ranked probability score (CRPS), which does not have the drawbacks mentioned for the log score. Therefore we will use this measure to evaluate our density forecasts. The CRPS is defined as:

$$\begin{aligned} CRPS(t+h,l) &= \int_{-\infty}^{\infty} (F(z) - I\{y_{t+h} \leq z\})^2 dz \\ &= E_f|Y_{t+h,l} - y_{t+h}| - \frac{1}{2} E_f |Y_{t+h,l} - Y'_{t+h,l}| \ , \end{aligned} \tag{21}$$

where F is the cumulative density function (CDF) that corresponds to the predictive density f of model l at time t, $I(.)$ takes the value 1 if $y_{t+h} \leqq z$ and 0 otherwise. E_f is the expectation of predictive density f, $Y_{t+h,l}$ and $Y'_{t+h,l}$ are independent random variables with sampling density for both equal to posterior predictive density of model l for y_{t+1} at time t.

As can be seen from (21), the CRPS measures the distance between the CDF implied by the model and the CDF of the realization. A higher CRPS means a worse forecast density and a lower CRPS means a better forecast density. According to the second equation we can see the CRPS as two parts. The first part is the average absolute distance between the empirical CDF of y_{t+h} seen as a step function and the empirical CDF that is associated with the predictive density of model l. The second part is a measure of the variance of the prediction. This can easily be obtained by random resampling the draws from the MCMC sampler or analytically when we use a Gaussian approximation.

(21) concerns only the evaluation of a single forecast. If we want to see how our CRPS is over the whole forecasting horizon we take the average of them all, which is given by

$$avCRPS_l = \frac{1}{T - t_0 - h} \sum_{s=t_0-1}^{T-h} C\hat{R}PS(s+h,l) \ . \tag{22}$$

4.5 Out-of-sample results

With the models discussed in Section 4.1 forecasts are made for the GDP deflator for the current quarter and one year ahead. The forecasts are made for the period 2010Q1 up to

2015Q4. These forecasts are evaluated using the methods discussed in Section 4.4. In this time period we can see two different periods, the years of recovery up to 2014Q4 and the zero inflation period afterwards. Since there is such a difference in inflation behaviour, we will also calculate the evaluation measures for these two time periods. First we will discuss the full sample, than 2010Q1-2013Q4 and last 2014Q1-2015Q4.

In Table 3 the RMSE, MAE and CRPS for the RW,BIC-AR, PREVOBS-AR, mixture, Equal-Combination and TVW-Combination model are shown for time periods 2010Q1-2015Q4, 2010Q1-2013Q4 and 2014Q1-2015Q2. These are for both the one quarter ahead ($h = 1$) and the one year ahead ($h = 5$). First we inspect the full sample one quarter ahead for each model which is not a combination. We observe that the BIC-AR model is the best performing, whereas it it not significantly better than the RW model[3]. The mixture model is performing quite fine, yet the benchmark models are doing a slightly better job. The PREVOBS-AR model is not performing well, since it has higher scores. For the one year ahead the PREVOBS-AR model is still the worst model. Yet we see that the mixture model has the lowest value for the avCRPS and is not significantly worse than the RW model. The performance of the RW model might be worse here since it is puts a lot of emphasis on the last observation for the forecasting and is very subject to outliers. The mixture model is more controlled and thus less subject to this.

If we now consider the time period 2010Q1-2013Q4, we see that the forecasts are better for every model except the PREVOBS-AR model than for the full sample. The better results are expected, since this is a relatively stable period, just like most of the testing sample.

Now if we focus on 2014Q1-2015Q4, we observe an interesting change. The PREVOBS-AR model is doing not significantly worse than the other models, if we consider RMSE and MAE. This is due to the fact that the PREVOBS-AR model is a model that works the best in more extreme areas of the model. What we notice furthermore is that in this period of low inflation, the mixture model is performing not so good if we take the one quarter ahead forecast. In the sampling period, the mixture model did not have many low inflations/deflation periods and hence the forecasts for this period are not accurate. It might be wise to include a third cluster for these low inflations values for the future, in order to be able to better forecast such low inflation periods.

[3]We use a t-test based upon the DM-statistic as used by Groen et al. (2013). This statistic is based upon the Diebold and Mariano (1995) statistic with the correction from Harvey et al. (1997)

Table 3: The RMSE, MAE and CRPS for the RW,BIC-AR, PREVOBS-AR, mixture, Equal-Combination and TVW-Combination model for time periods 2010Q1-2015Q4, 2010Q1-2013Q4 and 2014Q1-2015Q2. These are based upon the forecasts for one period ahead and one year ahead.

	One quarter ahead ($h = 1$)			*One year ahead ($h = 5$)*		
	RMSE	MAE	avCRPS	RMSE	MAE	avCRPS
	Forecast evaluation sample 2010Q1-2015Q4					
RW	0.311	0.242	0.189	0.309	0.246	0.424
BIC-AR	0.304	0.238	0.184	0.356	0.291	0.328
PREVOBS-AR	0.526	0.476	0.574	0.452	0.402	0.567
Mixture	0.360	0.298	0.213	0.331	0.278	0.277
Equal-Combination	0.170	0.142	0.148	0.178	0.145	0.159
TVW-Combination	0.280	0.215	0.192	0.302	0.256	0.283
	Forecast evaluation sample 2010Q1-2013Q4					
RW	0.264	0.210	0.158	0.254	0.209	0.413
BIC-AR	0.244	0.208	0.152	0.292	0.247	0.284
PREVOBS-AR	0.577	0.536	0.593	0.483	0.430	0.575
Mixture	0.289	0.249	0.173	0.264	0.236	0.226
Equal-Combination	0.165	0.203	0.146	0.166	0.132	0.156
TVW-Combination	0.270	0.133	0.191	0.261	0.224	0.236
	Forecast evaluation sample 2014Q1-2015Q4					
RW	0.388	0.307	0.250	0.396	0.319	0.446
BIC-AR	0.397	0.299	0.246	0.457	0.380	0.416
PREVOBS-AR	0.402	0.356	0.561	0.384	0.346	0.553
Mixture	0.470	0.396	0.293	0.435	0.363	0.381
Equal-Combination	0.180	0.159	0.154	0.199	0.173	0.165
TVW-Combination	0.300	0.239	0.193	0.384	0.290	0.377

The results of the PREVOBS-AR model are not as good as expected. This model flourishes in extreme periods, but in stable time periods the results are below standard. This model might be better to use at a more volatile dataset. Our recommendation is to only use this model when one is in a volatile period of a time series. The performance of the mixture model is quite good. The forecasts are of the same accuracy as the random walk and BIC-AR model for a quarter ahead forecast. For larger horizons, that is when this model has really good density forecasts. It seems that this model is more useful for forecasting in the longer run. Furthermore it might be worthwhile to include another mixture component to capture the low inflation periods.

Now if we look at the combination models, we observe that the TVW-Combination model makes forecasts of equal quality as the RW model in the one quarter ahead forecasts. For the one year ahead forecasts, they are of similar quality, yet the density is captured much better. The Equal-Combination model makes really good forecasts, being significant better than all the other models in every measure and at every time horizon. We found after a closer inspection that the TVW-Combination was always too late with adjusting the weights and overcompensated the weights, what resulted in larger errors.

5 Conclusion

In this paper we propose two models for predicting inflation. In both models the previous observation plays an important role for the dynamic pattern. The first model is an autoregressive model with time-varying parameters which are dependent on the previous observation. In the second model is a mixture of autoregressive models, where the regime probabilities are dependent on the previous observation.

The real time inflation forecasting performance of the two models is evaluated using MAE, RMSE and avCRPS. We use a random walk model and a time-invariant autoregressive specification as benchmarks. We find that both models provided accurate density forecasts. We notice that the time-varying AR model is most fruitful in times of extreme inflations. The mixture model has the best performance during stable inflation periods, yet it fails to make good forecasts in the low inflation periods, since there was no low inflation in the training sample. We find that combining the two models with an equal-weighing scheme, drastically improves the forecasts in all the used measures.

A suggestion for further research might be to use the AR model with time varying parameters where we include more predictors for the inflation, such as short-term interest rates and the unemployment ratio. Another idea might be to make a model that switches between the mixture model for the stable periods and the time varying AR model in the more extreme periods.

References

Alvarez, F., Lucas, R. E., and Weber, W. E. (2001). Interest rates and inflation. *The American Economic Review*, 91(2):219–225.

Arango, L. E. and Gonzalez, A. (2001). Some evidence of smooth transition nonlinearity in Colombian inflation. *Applied Economics*, 33(2):155–162.

Atkeson, A. and Ohanian, L. E. (2001). Are Phillips curves useful for forecasting inflation? *Fed. Reserve Bank Minneapolis Quart. Rev.*, 25(1):2–11.

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 25(1):47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1):1–22.

Benati, L. (2004). Evolving post-world war II U.K. economic performance. *Journal of Money, Credit and Banking*, 36(4):691–717.

Bidarkota, P. V. (2001). Alternative regime switching models for forecasting inflation. *Journal of Forecasting*, 20(1):21–35.

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of time series analysis*, 7(3):179–190.

Chib, S. (2001). Markov chain Monte Carlo methods: computation and inference. *Handbook of econometrics*, 5:3569–3649.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic dynamics*, 8(2):262–302.

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., and Lauritzen, S. L. (1981). Statistical analysis of time series: some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, 8(2):93–115.

Culver, S. E. and Papell, D. H. (1997). Is there a unit root in the inflation rate? Evidence from sequential break and panel data models. *Journal of Applied Econometrics*, 12(4):435–444.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 39(1):1–38.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263.

Fischer, S. et al. (1996). Why are central banks pursuing long-run price stability? *Achieving price stability*, 2:7–34.

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Gneiting, T. and Ranjan, R. (2011). Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422.

Granger, C. W. J. and Terasvirta, T. (1993). *Modelling non-linear economic relationships*. OUP Catalogue. Oxford University Press.

Groen, J. J., Paap, R., and Ravazzolo, F. (2013). Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics*, 31(1):29–44.

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

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2):281–291.

Hoogerheide, L., Kleijn, R., Ravazzolo, F., Van DijK, H. K., and Verbeek, M. (2010). Forecast accuracy and economic gains from bayesian model averaging using time-varying weights. *Journal of Forecasting*, 29(1-2):251–269.

Kim, C.-J. (1993). Unobserved-component time series models with markov-switching heteroscedasticity: Changes in regime and the link between inflation rates and inflation uncertainty. *Journal of Business & Economic Statistics*, 11(3):341–349.

Koirala, T. P. (2012). Inflation persistence in Nepal: A TAR representation. Technical report.

Kurita, T., Otsu, N., and Abdelmalek, N. (1992). Maximum likelihood thresholding based on population mixture models. *Pattern Recognition*, 25(10):1231–1240.

Levin, A. T. and Piger, J. (2004). Is inflation persistence intrinsic in industrial economies? Working Paper Series 0334, European Central Bank.

Marcellino, M. G., Stock, J. H., and Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of econometrics*, 135(1):499–526.

Marin, J.-M. and Robert, C. (2007). *Bayesian core: a practical approach to computational Bayesian statistics*. Springer Texts in Statistics.

Nadal-De Simone, M. F. (2000). *Forecasting Inflation in Chile Using State-Space and Regime-Switching Models*. Number 0-162. International Monetary Fund.

Neal, R. M. (2003). Slice sampling. *Annals of statistics*, 31(3):705–767.

O'Reilly, G. and Whelan, K. (2005). Has Euro-area inflation persistence changed over time? *The Review of Economics and Statistics*, 87(4):709–720.

Phillips, A. W. (1958). The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1862-1957. *Economica*, 25(100):283–299.

Phiri, A. (2013). Inflation and economic growth in Zambia: A threshold autoregressive (TAR) econometric approach. Mpra paper, University Library of Munich, Germany.

Salimans, T. (2012). Variable selection and functional form uncertainty in cross-country growth regressions. *Journal of Econometrics*, 171(2):267–280.

Sensier, M. and Van Dijk, D. (2004). Testing for volatility changes in US macroeconomic time series. *Review of Economics and Statistics*, 86(3):833–839.

Simon, J. (1996). A Markov-switching model of inflation in Australia. RBA Research Discussion Papers rdp9611, Reserve Bank of Australia.

Sims, C. A. and Zha, T. (2006). Were there regime switches in US monetary policy? *American Economic Review*, 96(3):54–81.

Speed, C. (1997). Inflation modelling. In *7th International AFIR Colloquium August*.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(3):245–292.

Wood, S., Rosen, O., and Kohn, R. (2011). Bayesian mixtures of autoregressive models. *Journal of Computational and Graphical Statistics*, 20(1):174–195.

Yang, M.-H. and Ahuja, N. (1998). Gaussian mixture model for human skin color and its applications in image and video databases. volume 3656 of *Storage and Retrieval for Image and Video Databeses VII*, pages 458–466.