Quantile regression forecasts of inflation under model uncertainty

Dimitris Korobilis
University of Glasgow*

June 13, 2016

Abstract

This paper examines the performance of Bayesian model averaging (BMA) methods in a quantile regression model for inflation. Different predictors are allowed to affect different quantiles of the dependent variable. Based on real-time quarterly data for the US, we show that quantile regression BMA (QR-BMA) predictive densities are superior and better calibrated compared to those from BMA in the traditional regression model. Additionally, QR-BMA methods compare favorably to popular nonlinear specifications for US inflation.

Keywords: Bayesian model averaging; quantile regression; inflation forecasts; fan charts

JEL Classification: C11, C22, C52

1 Introduction

Quantile regression generalizes traditional least squares regression by fitting distinct regression lines for each quantile of the distribution of the variable of interest. Least squares regression only produces coefficients that allow to fit the mean of the dependent variable conditional on some explanatory/predictor variables. In that respect, quantile regression is more appropriate for making inferences about predictive distributions and assessing forecast uncertainty. At the same time quantile regression estimates are more robust against outliers in the dependent variable. Therefore, quantile regression can be used to discover predictive relationships between the dependent and exogenous variables,

*Adam Smith Business School, University of Glasgow, Gilbert Scott Building, G12 8QQ Glasgow, UK. Tel: +44 141 330 2950. email: Dimitris.Korobilis@glasgow.ac.uk.

when typical regression modelling fails to indicate the existance of predactibility in these exogenous variables; see Koenker (2005).

In this paper we examine the forecasting performance of Bayesian quantile regression. The final aim is to produce quantile forecasts for inflation using several potential explanatory variables and examine the role of model uncertainty in quantile forecasts. Bayesian model averaging (BMA) and selection (BMS) methods have been traditionally used to deal with model uncertainty in forecasting regressions. Following Alhamzawi and Yu (2012) and Yu et al. (2013), it is shown that application of BMA to the quantile regression model, allows for forecasting each quantile of inflation using a different set of predictors whereas estimation is quite straightforward using Bayesian methods. By using model selection and averaging in a quantile regression setting means that we can approximate complex forms of the posterior predictive density of inflation, despite the fact that the quantile regression model specified in this paper is inherently linear[1]. Although a large empirical literature using quantile regression exists, applications of (Bayesian) model averaging are scarce. The only exception is the study of Crespo-Cuaresma, Foster and Stehrer (2011), however, these authors approximate Bayesian inference by using Least Squares and the Bayesian Information Criterion (BIC).

This paper integrates two vastly expanding literatures. On the one hand, there are several studies which develop estimation, inference and forecasting in (Bayesian) quantile regression models, such as Gaglianone and Lima (2012), Geraci and Bottai (2007), Gerlach, Chen and Chan (2011), Lancaster and Jun (2010), Meligkotsidou, Vrontos and Vrontos (2009), Schüler (2014), Tsionas (2003) and Yu and Moyeed (2001). On the other hand, there is a vast literature in macroeconomic and financial forecasting that shows the superiority of Bayesian model averaging and selection methods over other alternatives; see Koop and Korobilis (2012) and Wright (2008), among others.

Empirical evaluation of the quantile regression BMA method is based on real-time forecasting of quarterly US consumer price index inflation, observed for the period 1947Q1-2015Q3, using 16 potential predictors also measured in real-time. We show which predictors are relevant for each quantile of inflation at various forecast horizons, and we compare my results to Bayesian model averaging in the mean regression specification, as well as popular nonlinear regression specifications that have been shown to forecast inflation well. Based on predictive likelihoods (Geweke and Amisano, 2011) the quantile regression BMA provides superior density forecasts compared to regular regression BMA, and naive quantile regression methods without BMA.

In the next Section we present the Bayesian quantile dynamic regression model and the BMA prior, and in Section 3 we present the empirical results. Section 4 concludes the paper and discusses further extensions.

[1]A recent exception is Bernardi, Casarin and Petrella (2016). These authors allow for quantile regressions with time-varying parameters and dynamic assessment of model uncertainty using dynamic BMA.

2 Bayesian quantile regression

Following Yu and Moyeed (2001) the quantile regression model has a convenient mixture representation which, as explained below, is particularly convenient for Bayesian estimation using the Gibbs sampler. In particular, I consider for the inflation process y_t the following linear model

$$y_t = x_t'\beta_p + \varepsilon_t, \tag{1}$$

where x_t is a $n \times 1$ vector of explanatory variables and own lags, and β_p is a vector of coefficients dependent on the p-th quantile of the random error term ε_t which is defined as the value q_p for which $\Pr(\varepsilon_t < q_p) = p$. In typical specifications of quantile regression (Koenker, 2005) the distribution of ε_t is left unspecified (that is, it is a nonparametric distribution F_p), and estimation of β_p is the solution to the following minimization problem

$$\min_{\beta} \sum_{t=1}^{T} \rho_p\left(\varepsilon_t\right), \tag{2}$$

where the loss function is $\rho_p\left(u\right) = u\left(p - I\left(u < 0\right)\right)$ and $I\left(A\right)$ is an indicator function which takes value one if event A is true, and zero otherwise.

The major contribution of Yu and Moyeed (2001) was to show that the minimization problem shown in equation (2) is equivalent to maximizing a likelihood function under the asymmetric Laplace error distribution; see also Tsionas (2003). Reed and Yu (2011) have recently established, both theoretically and empirically, that the assymetric Laplace likelihood accurately approximates the true quantiles of many distributions having different properties. At the same time, as shown in Kotz et al. (1998), the asymmetric Laplace distribution can admit various mixture representations. In Bayesian analysis a popular representation is that of a scale mixture of normals with scale parameter following the exponential distribution. This mixture formulation allows for the likelihood function to be written in conditionally Gaussian form, and inference based on conditional posterior distributions is straightforward. Even when the joint posterior distribution of model parameters is of complex form (as it is the case when the likelihood is asymmetric Laplace - no matter what the prior is), one can rely on the Gibbs sampler (Reed and Yu, 2011) in order to sample from these conditional posteriors. When the conditional lilkelihood admits a normal or a mixture of normals form, these conditional posteriors belong to known distributions and, thus, easy to draw samples from; see the Techincal Appendix for details.

Following Kozumi and Kobayashi (2011) we can represent the error distribution ε_t using the form

$$\varepsilon_t = \theta z_t + \tau\sqrt{z_t}u_t, \tag{3}$$

where $z_t \sim \mathit{Exponential}\left(1\right)$, that is, a variate from an exponential distribution with rate parameter one, and u_t is distributed standard normal. In this formulation it holds that $\theta = \left(1 - 2p\right)/p\left(1 - p\right)$, and $\tau^2 = 2/p\left(1 - p\right)$, for a given quantile $p \in [0, 1]$. Supplanting the formula for ε_t into equation (1) gives the new quantile regression form

$$y_t = x_t' \beta_p + \theta z_t + \tau \sqrt{z_t} u_t, \tag{4}$$

and the conditional density of y_t given the Exponential variates z_t is Normal and is of the form

$$\begin{aligned} f(\mathbf{y}|\beta_p, \mathbf{z}) \quad \propto \quad & \left(\prod_{i=1}^{T} z_t^{-\frac{1}{2}} \right) \times \\ & \exp\left\{ -\frac{1}{2} \sum_{i=1}^{T} \frac{(y_t - x_t'\beta_p - \theta z_t)^2}{\left(\tau\sqrt{z_t}\right)^2} \right\}, \end{aligned}$$

where $\mathbf{y} = (y_1, ..., y_T)'$ and $\mathbf{z} = (z_1, ..., z_T)'$.

Given this likelihood formulation we can now define prior distributions. Bayes theorem says that the posterior distribution is simply the product of the (conditionally) Normal likelihood and the prior. In particular, Yu and Moyeed (2001) prove that all the posterior moments of β_p exist when the prior for β_p is Normal. In this paper we consider the conditionally Normal prior

$$(\beta_{i,p}|\gamma_{i,p}, \delta_{i,p}) \sim (1 - \gamma_{i,p}) N\left(0, \underline{c} \times \delta_{i,p}^2\right) + \gamma_{i,p} N\left(0, \delta_{i,p}^2\right), \tag{5}$$

$$\left(\delta_{i,p}^{-2}\right) \sim Gamma\,(a_1, a_2), \tag{6}$$

$$(\gamma_{i,p}|\pi_0) \sim Bernoulli\,(\pi_0), \tag{7}$$

$$(\pi_0) \sim Beta\,(b_1, b_2). \tag{8}$$

where $\underline{c} \to 0$ is a fixed hyperparameter. This multi-level prior specification for $\beta_{i,p}$, $i = 1, ..., n$, is a mixture of normals prior. Whenever the indicator variable $\gamma_{i,p} = 1$ then $\beta_{i,p}$ has a Normal prior with variance $\delta_{i,p}^2$. When $\gamma_{i,p} = 0$ then $\beta_{i,p}$ has a Normal prior with mean zero and variance $\underline{c} \times \delta_{i,p}^2$ which will be very close to zero as long as $\underline{c}$ is selected to be small enough (in the empirical application of this paper $\underline{c} = 0.00001$). Such an extremely informative prior means that predictor $x_{i,t}$ is not relevant for the p-th quantile. The indicators $\gamma_{i,p}$ are estimated from the data, thus they have their own Bernoulli prior with probability π_0. Additionally, in order to avoid subjectively selecting the hyperparameters π_0 and $\delta_{i,p}^2$, we introduce hyper-prior distributions on them so that they are estimated from the likelihood.

Posterior computation is relatively straightforward, as all conditional posterior distributions belong to known families and can easily be sampled from using Markov Chain Monte Carlo methods. In particular, we sequentially sample from the posteriors of each unkown parameter conditional on all other parameters using a standard Gibbs sampler algorithm, which is provided in the Technical Appendix. For results in the next Section referring to the full sample of the data, 50000 Monte Carlo iterations have been used, 10000 iterations are discarded for convergence, and from the remaining 40000 draws only every 40th draw is retained; see also Appendix B for a justification of this approach and the selection of draws. For the more demanding recursive forecasting exercise, also outlined in the next Section, 22000 iterations are used of which 2000 are discarded, and only every 20th draw is stored (for a total of 1000 draws from the posterior densities used for inference). In both cases, convergence of the Gibbs sampler is quite satisfactory and the number of iterations can be considered sufficient for the nature of the application and the length of the sample. Convergence diagnostics are also provided in the Appendix.

3 Empirics

3.1 Data and models

In this section we examine whether QR-BMA can provide superior point and density forecasts compared to popular linear and nonlinear specifications that have been considered successful at forecasting inflation. We consider real-time data for CPI for the period 1947Q1-2015Q3 as the dependent variable, and two own lags of inflation as well as 16 variables measured in real-time as the potential predictors. In particular, the dataset contains various measures of economic activity (e.g. unemployment, investment), money supply (e.g. M1) and expectations (e.g. default yield spread). All predictors are either measured in real time, or their final vintage is used if they are not subject to revisions (e.g. interest rates). Further important variables can also be used as predictors (e.g. surveys), however, they either are not available in real-time, or their sample is considerably smaller which would make any forecast comparison less reliable (due to small estimation and evaluation samples). The data, which are downloaded from the Real Time Data Research Center of the Philadelphia Fed, and the St Louis Federal Reserve Economic Database (FRED), are explained in detail in the Data Appendix.

For the purpose of forecasting, the model in equation (1) is re-written as

$$y_{t+h,p} = x_t' \beta_p + \varepsilon_{t+h}, \tag{9}$$

for $t = 1, ..., T - h$, and a similar expression holds for the transformed model in (4). This is a typical specification of a generalized backwards-looking Phillips curve model for forecasting the p-th quantile of inflation; see Stock and Watson (2007). When computing quantile forecats, we follow Gaglianone et al (2012) and collect the quantities $y^p_{T+h|T}$ using a fairly large grid for p [2] and construct the full predictive density using kernel smoothing based on a Gaussian kernel.

For comparison we also estimate and forecast with several alternative specifications which have been shown to forecast inflation well:

[2]For each draw from the Gibbs sampler we generate forecasts of quantiles $p \in [0.05, 0.10, ..., 0.90, 0.95]$, i.e. we obtain 19 quantiles. We do not consider the 5% probability from each tail of the predictive distribution for reasons explained in Gaglianone et al. (2012).

1. BMA regression (Wright, 2008): The standard "mean"regression model is of the form

$$y_{t+h} = x_t'\beta + \varepsilon_{t+h},$$

where now there is a single β such that $E(\mathbf{y}) = E(\mathbf{x})\beta$. For the sake of simplicity and comparability, BMA is implemented using the prior described in equations (B.2)-(B.3), which is now applied on β instead of β_p (and posterior expressions are a special case of the ones derived for the BMA quantile regression). The same 16 predictors defined for the benchmark QR-BMA are also used in the estimation of the BMA regression.

2. UCSV regression (Stock and Watson, 2007): The Unobserved Components Stochastic Volatility model is already a benchmark model for forecasting inflation. It takes the form

$$\begin{aligned} y_{t+h} &= c_t + \varepsilon_{t+h}, \\ c_t &= c_{t-1} + u_t, \end{aligned}$$

where c_t is trend inflation and the disturbance terms ε_{t+h} and u_{t+h} have stochastic volatilities (such that log-variances follow random walks).

CPS regression (Cogley, Primiceri and Sargent, 2010): This is the UC-SV model of Stock and Watson where now there is one autoregressive term for inflation, i.e. y_t, which also has a time-varying coefficient.

3. TVP-DMA regression with Dynamic Model Averaging (Koop and Korobilis, 2015): This model generalizes the UC-SV and CPS models by allowing inflation to depend on further predictors. Similar to the benchmark QR-BMA and BMA regression models, the same 16 real-time predictors are used for estimation of this model. In order to deal with overparametrization concerns (especially compared to the parsimonious UC-SV model) Koop and Korobilis (2012) suggest to perform Bayesian model averaging at each point in time, leading to a Dynamic Model Averaging (DMA) scheme.

All the models above rely on various tuning hyperparameters and prior distributions, given that estimation in the original respective papers is Bayesian. In that respect, and given that these are highly nonlinear models, we try to follow settings which are fairly uninformative or that broadly follow the recommendations of the original authors. For example, for the time-varying coefficients in the UCSV, CPS and TVP-DMA models we use the same initial condition, $N(0, 10)$, while the prior on state covariances is a "business as usual prior" in the sense of Cogley and Sargent (2005). Therefore, in the USSV and the CPS models the prior scale is $0.0001 \times I$, while in the TVP-DMA the relevant forgetting factor is set to $\lambda = 0.99$; the reader should consult the original papers for more details.

3.2 How does QR-BMA work?

This subsection clarifies certain features of the QR-BMA algorithm, and explains why this can be potentially useful for forecasting inflation. The first interesting exercise is to pin down the relevant predictor variables selected by BMA for each quantile of inflation. Table 1 presents these variables for the QR-BMA model estimated at three representative forecast horizons, $h = 0$ (short horizon), $h = 4$ (medium), and $h = 8$ (long). Additionally results for five representative quantiles are presented, $p = 0.05, 0.25, 0.5, 0.75, 0.95$. These results refer to the full sample of the data, 1947Q1 - 2015Q3; a different sample will imply different relevant predictors. Koop and Korobilis (2012) have shown in particular, that predictors of inflation are extremely unstable over time, so relevant predictors are expected to be quite different for different samples.

Table 1: Selected predictors of CPI inflation per quantile; horizons $h = 0$, $h = 4$, and $h = 8$; full sample 1947q1-2015q3

Predictors at horizon $h = 0$ (nowcasting)

Predictor	1^{st} lag	IPM	HSTARTS	CUM	RINVBF	ROUTPUT	RUC	ULC	WSD	DYS	NAPMNOI
p=.5		•		•	•	•		•	•		•
p=.25	•									•	
p=.5	•									•	
p=.75	•									•	
p=.95		•					•				•

Predictors at horizon $h = 4$

Predictor	1^{st} lag	IPM	HSTARTS	CUM	RINVBF	ROUTPUT	RUC	ULC	WSD	DYS	NAPMNOI
p=.5	•	•		•	•	•		•	•		•
p=.25	•									•	
p=.5	•									•	
p=.75	•										
p=.95	•	•					•				•

Predictors at horizon $h = 8$

Predictor	1^{st} lag	IPM	HSTARTS	CUM	RINVBF	ROUTPUT	RUC	ULC	WSD	DYS	NAPMNOI
p=.5	•	•		•		•		•	•		
p=.25	•		•			•				•	
p=.5	•		•							•	
p=.75	•		•							•	
p=.95	•	•					•				

Note: Predictors with probability of inclusion < 0.5 for any of the models, are not included in the table.

The Table shows which predictors have probability of inclusion in the final regression which is higher than 0.5, i.e. what Barbieri and Berger (2004) call the "median probability model". Consistent with the Bayesian variable selection literature (e.g. Chipman et al., 2001), these probabilities are calculated as the mean of the posterior of $\gamma_{i,p}$: the posterior distribution of $\gamma_{i,p}$ is a sequence of zeros and ones, so that the posterior mean denotes a probability for each variable i in each quantile p.

The main message of this Table is two-fold: i) more predictors are relevant when forecasting inflation at longer horizons, and ii) more predictors are relevant when forecasting extreme quantiles. The first message is already well established in the relevant literature. Papers such as Koop and Korobilis (2012) find that only two variables are relevant for forecasting CPI one-step ahead (inflation expectations and M1 in their model), while many more variables are relevant as the forecast horizon increases. The second message, though, is a novel one in this literature and an encouraging one. It says that in times of extreme changes in inflation rates, that is changes that are beyond the "median expectation" of consumers and/or the central bank, there is potential predictive ability conveyed in several variables.

In order to understand why this can be the case, it would be interesting to examine predictive distributions from the QR-BMA for two recent extreme events. One is the big drop in inflation in 2009Q1 following the collapse of the global banking system and the turbulence in commdoity markets. Note that in monthly data the largest drop was actually in November 2008, but since this paper relies on quarterly data which are averages of monthly ones the decrease is dated as 2009Q1. The second important extreme event for US inflation was the deflation in 2015Q1, when annualized quarter-on-quarter inflation rates hit negative territory. Both these events are very important for policy-makers, so that accurate prediction of their occurrence is of paramount interest.

The top and bottom panels of Figure 1 show one-step ahead predictive distributions of inflation estimated in 2008Q4 and 2014Q4, respectively. The graphs show the realized value one-step ahead (that is, 2009Q1 and 2015Q1), and the predictive distribution of the QR-BMA, the regular (mean) regression BMA, and the Survey of Professional Forecasters (SPF) [3]. The quantile regression model predictive distribution can be multi-modal and asymmetric, while the regression model distribution is always (conditionally) Normal and, thus, symmetric. The SPF distribution in 2008Q4 is close to Normal and matches very closely that of the mean regression BMA. In this specific case the QR-BMA gives quite a lot of weight on highly negative outcomes, and considerably lower weight on positive outcomes for inflation. By allowing this kind of multimodality the QR-BMA distribution gives twice as much mass closer to the realization for inflation (which was -9.2% at an annual rate). In contrast, the symmetric regression and SPF distributions are restricted into giving equal weight to highly negative and highly positive outcomes.

In 2014Q4 a slightly different story is observed. The QR-BMA distribution is again multimodal, although this time looks symmetric. However, it is striking that, while the SPF distribution is a mixture with possibly three modes, it is quite concentrated

[3] Detailed data on the SPF are also available from the Philadelphia Fed website.

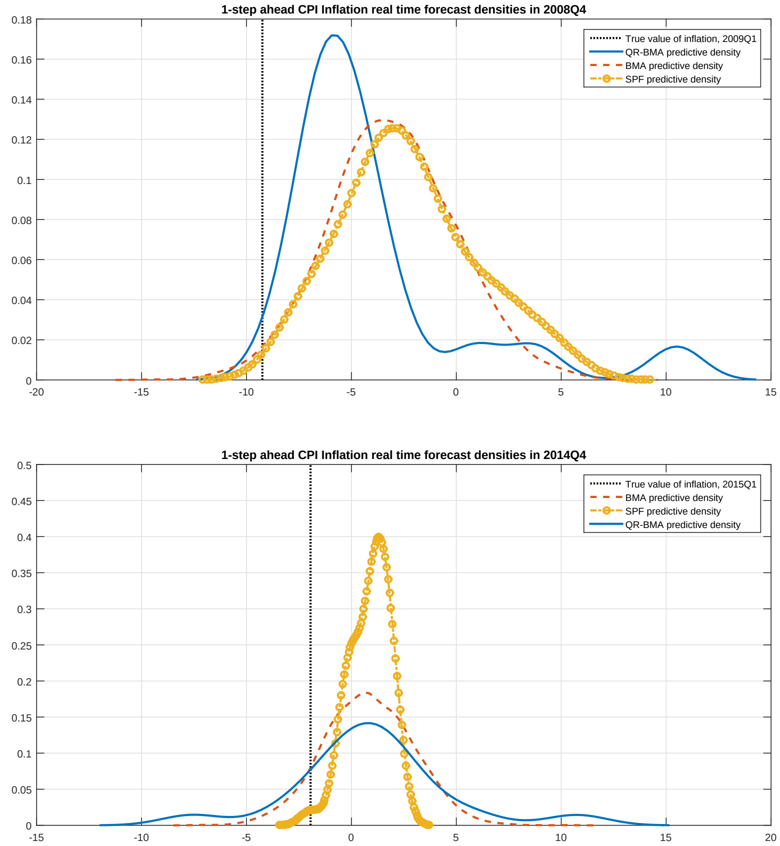

Figure 1: Predictive densities for abruptly occurred deflation events in U.S. data

around a certain value, that is 1%. Forecast disagreement is quite low, however, the vast majority of forecasters failed to account for the possibility of disinflation. In contrast, there is a mode with very low mass at the realized value of inflation, signifying that a small proportion of forecasters had correct expectations about inflation. Nevertheless, the model-based distributions (regular BMA and QR-BMA) perform much better in this particular case, since they both assign more probability on the true realization of inflation than the SPF distribution does.

3.3 Forecasting results

In order to evaluate the forecast performance of each model we consider a recursive pseudo-out-of-sample (poos) procedure: we start with estimation of model parameters for the 1975Q1 vintage (sample is 1947Q1-1974Q4), nowcast the 1975Q1 observation and forecast out-of-sample for each horizons $h = 1, 2, 3, 4, 5, 6, 7, 8$. Then take the 1975Q2 vintage (sample is 1947Q1-1975Q1) in order to estimate model parameters, nowcast and forecast, and repeat this procedure until the sample is exhausted (i.e. until the vintage

2015Q3-h). All forecasts are evaluated relative to their true value, which we consider to be the last available vintage in the dataset, that is, 2015Q4 (for the sample spanning 1947Q1-2015Q3).

A first assessment of forecasting performance can be implemented by looking at point forecasts. While the quantile regression model is naturally designed to allow for more complex predictive densities, reporting point forecasts is still the most popular way for policy-makers to communicate to the public their expectations about inflation. In that respect, the top panel of Table 2 presents Mean Squared Forecast Errors (MSFEs). In particular, the first row presents the performance on an AR(2) model for inflation, and the subsequent rows provide the MSFEs relative to the MSFE of the AR(2). Values lower than one show that the respective model generates point forecasts that are superior to the AR(2). Asterisks next the relative MSFEs show that the respective model has significantly better forecasting performance compared to the AR(2) at the 1% (***), 5% (***), and 10% (***) level, based on the Diebold-Mariano statistic; see Diebold and Mariano (1995).

While one should be careful when comparing point forecasts [4], the results suggest that QR-BMA improves over traditional BMA to the extend that it is comparable to the nonlinear UCSV specification of Stock and Watson (2007). Regarding the three time-varying parameter specifications, results are slightly different to previous results found in Koop and Korobilis (2012) using a shorter sample (which didn't include the 2009-2015 period) and slightly different set of predictors (they also included measures such as inflation expectations). In particular, the UCSV specification seems to be the best performing model for longer term forecasts ($h > 6$), even though this specification has no explanatory variables. Additionally, the nonlinear CPS and TVP-DMA specifications are not performing that well compared to the benchmark AR(2), even though previous evidence (Koop and Korobilis, 2012) suggests that for short-term forecasts of inflation it is nonlinearity that matters (while there is scant evidence that inflation is affected by predictors).

A natural second step in the analysis is to evaluate density forecasts. The bottom panel of Table 2 shows the mean log predictive scores of all forecasting models. The first row shows the log predictive scores of the AR(2) model, and the subsequent rows the log score differentials for each model from the AR(2). As it is the case with the MSFEs, asterisks show significant differentials at the 1% (***), 5% (***), and 10% (***) level, based on the Diebold-Mariano statistic. These results clearly indicate that QR-BMA offers substantial improvements when considering the whole distribution of forecasts. The improvements are particularly evident in short horizons (nowcasting and $h = 1$), where surprise movements of inflation can result in failure of forecasting models (see also discussion in previous subsection and Figure 1).

[4] All methods but QR-BMA model the conditional mean of y_t, while in QR-BMA results are based on the $p = 0.5$ quantile (median) of y_t.

Table 2: Forecasting results, 1975q1-2015q3

	Mean Squared Forecast Error								
	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
AR(2)	5.93	5.42	4.84	4.80	4.83	4.93	5.17	5.43	5.59
BMA	0.86**	0.86**	0.85**	0.87**	0.91**	0.91*	0.94*	1.01	1.00
QR-BMA	0.81***	0.75***	0.76***	0.82***	0.84**	0.87**	0.93*	0.98	0.99
UCSV	0.82***	0.76***	0.73***	0.75***	0.80**	0.85**	0.88**	0.90*	0.91*
CPS	1.15	1.04	0.84**	0.88**	0.86**	0.86**	0.95	1.06	1.25
TVP-DMA	1.06	0.83***	0.68***	0.76***	0.81***	0.84***	0.88**	0.92*	0.96
	Mean Log Predictive Scores								
	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
AR(2)	-3.73	-3.42	-3.16	-3.15	-3.15	-3.14	-3.17	-3.19	-3.20
BMA	0.65**	0.94***	0.84***	0.71***	0.69***	0.57***	0.56**	0.54**	0.54**
QR-BMA	1.42***	1.11***	0.87***	0.82***	0.80***	0.78***	0.78***	0.78**	0.77**
UCSV	-0.22	0.60**	0.83***	0.88**	0.88***	0.84***	0.63**	0.61**	0.65**
CPS	-0.49	0.72***	0.88***	0.99***	0.94***	0.98***	0.84***	0.87***	0.64**
TVP-DMA	0.08	0.61***	0.18*	1.07***	0.96***	0.85***	0.48**	0.41**	0.55**

Notes: Entries in the top panel of the Table are mean squared forecast errors (MSFE). The first row shows the MSFE of the AR(2) model for inflation, and the remaining rows the MSFE of competing models relative to the AR(2). Values lower (higher) than one mean that the respective model is doing better (worse) than the AR(2). MSFEs for the QR-BMA are based on the regression line estimated for the median ($p = 0.5$), while all other methods model the mean regression for inflation.

The bottom panel shows mean log predictive scores (MLPS). The first row shows the MLPS of the AR(2) model, and the remaining rows the MLPS differential for model M_i relative to the $AR(2)$.

We test equal predictive accuracy both in terms of the MSFE and the MLPS by means of the Diebold-Mariano statistic using the finite sample adjustment of Harvey et al. (1997) and Newey-West standard errors with one lag. Stars next to the relative MSFEs and MLPSs differentials denote that the (point and density, respectively) forecast performance of the respective model model is significantly better than the AR(2) at the 1% (***), 5% (***), and 10% (***) level.

Additional insight can be provided by the Probability Integral Transform (PIT), which is used here in order to evaluate the correct specification of predictive densities. For a given probability density function $p(y_{t+h}|y_t)$, the PIT is the corresponding cumulative density function (CDF) evaluated at the realization y_{t+h}:

$$z_{t+h} = \int_{-\infty}^{y_{t+h}} p(u|y_t)du \equiv P(y_{t+h}|y_t). \tag{10}$$

If the estimated predictive density is consistent with the "true" predictive density, then the sequence of all z_{t+h} in the out-of-sample evaluation period (i.e. 1975Q1-2015Q3-h) is is an independent and identically distributed (iid) Uniform (0,1) and its cumulative distribution function is the 45° line; Diebold, Gunther and Tay (1998).

A visual assessment by means of plotting histograms of the PITs is a first way of evaluating predictive densities: the closer the PITs look to a continuous Uniform distribution, the better calibrated they are. Nevertheless, more formal metrics exist, which allow to formally test uniformity of the PITs. Following Rossi and Sekhposyan (2014), we can test how close the CDF of the PITs is to that of the uniform distribution using the Kolmogorov-Smirnov (KS) test and its Anderson-Darling (AD) modification. Additionally, we can use the result of Berkowitz (2001) that if $z_{t+h} \overset{iid}{\sim} U(0,1)$ then $\zeta_{t+h} \equiv \Phi^{-1}(z_{t+h}) \overset{iid}{\sim} N(0,1)$ where $\Phi^{-1}(\cdot)$ is the inverse of the Normal CDF denoted by $\Phi(\cdot)$. The Doornik-Hansen (DH) test is used to assess normality of the transformed variable ζ_{t+h}. Finally, the Ljung-Box (LB) test is used to test for independence in the first and second central moments of the PITs; see Rossi and Sekhposyan (2014) and references therein. We denote by LB1 the LB statitic for testing independence of the mean, and LB2 the statistic for testing independence of the variance of the PITs.

Table 3 provides diagnostics related to calibration of predictive densities, and is similar to Table 1 of Metaxoglou, Pettenuzzo and Smith (2016). Entries are p-values, and values lower than 0.05 signify rejection of the null hypothesis of the test at the 5% level. In the case of BMA for both horizons the KS test rejects the null of uniformity of the PITs. For the other models there is evidence that the PITs are uniformly distributed, even though this evidence is weaker (i.e. p-values marginally higher than 0.05) in the case of the UCSV and TVP-DMA models for $h = 1$. The DH test of the inverse Normal of the PIT also shows that BMA is the only misspecified model at both forecast horizons, while UCSV and TVP-DMA are misspecified at horizon $h = 1$. The LB statistics for serial correlation in the mean and variance of the PITs, LB1 and LB2 respectively, suggest that PITs for all models are serially correlated - the only exception being the variance of the PIT for the UCSV at $h = 0$. These statistics are not illuminating in comparing the various models, but the KS and DH statistics suggest that the quantile regression model generates predictive densities which are as well-calibrated as the predictive densities of popular nonlinear specifications for inflation.

Table 3: PIT statistics

	Horizon $h = 0$				Horizon $h = 1$			
	KS	DH	LB1	LB2	KS	DH	LB1	LB2
BMA	0.042	0.001	0.002	0.006	0.022	< 0.001	< 0.001	< 0.001
QRBMA	0.284	0.384	0.000	0.013	0.142	0.458	0.001	< 0.001
UCSV	0.464	0.068	0.002	0.369	0.088	0.036	< 0.001	< 0.001
CPS	0.301	0.164	0.000	0.003	0.165	0.137	< 0.001	0.001
TVP-DMA	0.294	0.416	0.009	0.171	0.066	0.017	< 0.001	0.001

Notes: Entries are p-values of the respective statistics for the Probability Integral Transforms (PITs) of the predictive densities: Kolmogorov-Smirnov (KS), Doornik-Hansen (DH), Ljung-Box test of the mean (LB1), and Ljung-Box test of the variance (LB2); see Rossi and Sekhposyan (2014) for more information.

Finally, if the policy-maker is interested in a certain quantile of inflation, as it is typically the case with Value-at-Risk (VaR) forecasting in finance, one can use several measures such as the DQ test of Engle and Manganelli (2004). In a recent paper, Gerlach, Chen and Lin (2016) show how to implement Bayesian variants of popular tests proposed in the quantile regression literature. Since VaR and similar measures are not of interest for assessing forecasts of inflation, such tests are not presented in this paper.

4 Conclusions

This paper proposes a new empirical procedure for implementing Bayesian model averaging, which allows different predictor variables to affect different quantiles of the dependent variable. The benefits of this flexible approach are evaluated using real-time data for CPI inflation for the US and a number of predictor variables. Results indicate that the quantile regression BMA approach indeed finds that different predictors are relevant for each quantile of inflation, and that by taking this feature into account, predictive distributions are superior.

References

[1] Alhamzawi, R. and Yu, K. (2012). Variable selection in quantile regression via Gibbs sampling. *Journal of Applied Statistics* 39(4), 799-813.

[2] Barbieri, M. M. and Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics* 32, 870-897.

[3] Berkowitz, J. (2001). Testing Density Forecasts, with Applications to Risk Management. *Journal of Business and Economic Statistics* 19(4), 465-474.

[4] Bernardi, M., Casarin, R. and Petrella, L. (2014), Dynamic model averaging for quantile regression. arXiv:1602.00856.

[5] Chipman, H., George, E. I. and McCulloch, R. E. (2001). The practical implementation of Bayesian model selection. In: P. Lahiri (Ed.), *Model selection* (pp. 67–116). Lecture Notes –Monograph Series, Vol. 38. Beachwood, OH: Institute of Mathematical Statistics.

[6] Cogley, T. and Sargent, T. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics* 8(2), 262-302.

[7] Crespo-Cuaresma, J., Foster, N. and Stehrer, R. (2011). Determinants of regional economic growth by quantile. *Regional Studies* 45(6), 809-826.

[8] Diebold, F.X., Gunther, T.A. and Tay, A. S. (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review* 39(4), 863-883.

[9] Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy *Journal of Business and Economic Statistics* 13, 253-263.

[10] Engle R. F. and Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* 22, 367-381.

[11] Gaglianone, W. P. and Lima, L. R. (2012). Constructing density forecasts from quantile regressions. *Journal of Money, Credit and Banking* 44, 1589-1607.

[12] Geraci, M. and Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics* 8, 140-154.

[13] Gerlach R, Chen C. W. S. and Chan, N. Y. C. (2011). Bayesian time-varying quantile forecasting for value-at-risk in financial markets. *Journal of Business and Economic Statistics* 29(4), 481-492.

[14] Gerlach R, Chen C. W. S. and Lin, E. M. H. (2016). Bayesian assessment of dynamic quantile forecasts. *Journal of Forecasting* (early online publication), doi: 10.1002/for.2408.

[15] Geweke, J. and Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* 26(2), 216-230.

[16] Geyer, C. (2011). Introduction to Markov Chain Monte Carlo. in Brooks, S., Gelman, A. and Meng, X-L (Eds.) Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC: Boca Raton.

[17] Harvey, D., Leybourne, S. and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13, 281-291.

[18] Koenker, R. (2005).*Quantile Regression*. New York: Cambridge University Press.

[19] Koop, G and Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review* 53(3), 867-886.

[20] Kozumi, H. and Kobayashi, G. (2011). Gibbs sampling methods for Bayesian quantile regression. *Journal of Statistical Computation and Simulation* 81(11), 1565-1578.

[21] Lancaster, T. and Jun, S. J. (2010). Bayesian quantile regression methods. *Journal of Applied Econometrics* 25, 287-307.

[22] Meligkotsidou, L., Vrontos, I. D. and Vrontos, S. D. (2009). Quantile regression analysis of hedge fund strategies. *Journal of Empirical Finance* 16, 264-279.

[23] Metaxoglou, K., Pettenuzzo, D. and Smith, A. (2016). Option-implied equity premium predictions via entropic tilting. Working Papers 99, Brandeis University, Department of Economics and International Businesss School.

[24] Pettenuzzo, D. and Timmermann, A. (2015). An empirical investigation of the nature of the break point process for U.S. GDP and inflation. *Journal of Business and Economic Statistics*, forthcoming.

[25] Reed, C. and Yu, K. M. (2011). Efficient Gibbs sampling for Bayesian quantile regression. manuscript.

[26] Schüler, Y. S. (2014). Asymmetric effects of uncertainty over the business cycle: A quantile structural vector autoregressive approach. Working Paper Series University of Konstanz 2014-02.

[27] Stock, J. H. and Watson, M. W. (2007). Why has U.S. inflation become harder to forecast?. *Journal of Money, Credit and Banking* 39, 3-33.

[28] Tsionas, E. (2003). Bayesian quantile inference. *Journal of Statistical Computation and Simulation* 73, 659-674.

[29] Wright, J. E. (2008). Forecasting US inflation by Bayesian model averaging, *Journal of Forecasting* 28, 131-144.

[30] Yu, K. M., Chen, C. W.S., Reed, C. and Dunson, D. B. (2013). Bayesian variable selection in quantile regression. *Statistics And Its Interface* 6, 261–274.

[31] Yu, K. M. and Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters* 54, 437-447.

A Data Appendix

Real time data are from Philadelphia Fed (https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data), and variables which are not revised (final vintages) are from FRED (http://research.stlouisfed.org/fred2/). The dependent variable is CPI (Consumer Price Index, Quarterly Vintages). All variables are transformed to be approximately stationary, implying that instead of forecasting second differences of CPI, inflation rates are forecast (which are close to being random walk). In particular, if $z_{i,t}$ is the original untransformed series, the transformation codes are (column Tcode below): 1 - no transformation (levels), $x_{i,t} = z_{i,t}$; 4 - logarithm, $x_{i,t} = ln\,(z_{i,t})$; 5 - first difference of logarithm, annualized, $x_{i,t} = 400 \times ln\,(z_{i,t}/z_{i,t-1})$.

No	Mnemonic	Description	Tcode	Source
1	CPI	Consumer Price Index, Quarterly Vintages	5	Philly
2	IPM	Industrial Production Index, Manufacturing	5	Philly
3	HSTARTS	Housing Starts	4	Philly
4	CUM	Capacity Utilization Rate, Manufacturing	5	Philly
5	M1	M1 Money Stock	5	Philly
6	RCOND	Real Personal Cons. Expenditures, Durables	5	Philly
7	RCONS	Real Personal Cons. Expenditures, Services	5	Philly
8	RG	Real Government Cons. & Gross Inv., Total	5	Philly
9	RINVBF	Real Gross Private Domestic Inv., Nonresidential	5	Philly
10	ROUTPUT	Real GNP/GDP	5	Philly
11	RUC	Unemployment Rate	2	Philly
12	ULC	Unit Labor Costs	5	Philly
13	WSD	Wage and Salary Disbursements	5	Philly
14	DYS	Default yield spread (Moody's BAA - AAA)	1	St Louis
15	NAPM	Purchasing Manager's Index	1	St Louis
16	NAPMII	Inventories Index	1	St Louis
17	NAPMNOI	New Orders Index	1	St Louis

B Technical Appendix

B.1 Posterior inference in the Bayesian quantile regression with model averaging prior

The transformed quantile regression model is given in equation (4) which we rewrite here for convenience

$$y_t = x_t'\beta_p + \theta z_t + \tau\sqrt{z_t}u_t, \tag{B.1}$$

with x_t' being the (fixed) exogenous variables, and $z_t \sim Exponential\,(1)$ and $u_t \sim N\,(0,1)$ are new variables introduced when transforming the likelihood (see main text for more details). The prior we use is of the form

$$\begin{aligned}
(\beta_{i,p}|\gamma_{i,p}, \delta_{i,p}) &\sim (1-\gamma_{i,p})\,N\left(0, \underline{c}\times\delta_{i,p}^2\right) + \gamma_{i,p}N\left(0,\delta_{i,p}^2\right), && \text{(B.2)}\\
\left(\delta_{i,p}^{-2}\right) &\sim Gamma\left(\underline{a}_1,\underline{a}_2\right), && \text{(B.3)}\\
(\gamma_{i,p}|\pi_0) &\sim Bernoulli\,(\pi_0), && \text{(B.4)}\\
(\pi_0) &\sim Beta\left(\underline{b}_1,\underline{b}_2\right). && \text{(B.5)}
\end{aligned}$$

where $(\underline{a}_1,\underline{a}_2,\underline{b}_1,\underline{b}_2)$ are prior hyperparameters chosen by the researcher, and $\underline{c}$ is a fixed parameter set very close to zero. In order to obtain draws from the posteriors of all the unknown parameters, we sample sequentially from the following conditional distirbutions

1. Sample $\beta\,(p)$ conditionally on knowing all other parameters (incl z_t) and, of course, the data x_t, y_t, from:
$$\beta_p|\gamma_p,\tau^2,\mathbf{z},\mathbf{x},\mathbf{y} \sim N\left(\overline{\beta},\overline{V}_\beta\right),$$
where $\overline{V}_\beta = \left(\sum_{t=1}^T \frac{\widetilde{x}_t'\widetilde{x}_t}{\tau^2 z_t} + \Delta_p^{-1}\right)^{-1}$ and $\overline{\beta} = \overline{V}_\beta\left[\sum_{t=1}^T \frac{\widetilde{x}_t(y_t-\theta z_t)}{\tau^2 z_t}\right]$, and Δ is a diagonal prior variance matrix with diagonal element $\delta_{i,p}^2$ if $\gamma_{i,p}=1$ or $\underline{c}\delta_{i,p}^2$ if $\gamma_{i,p}=0$

2. Sample $\delta_{i,p}^{-2}$ conditional on other parameters and data from:
$$\delta_{i,p}^{-2}|\beta_{i,p},\mathbf{x},\mathbf{y} \sim Gamma\left(\overline{a}_1,\overline{a}_2\right),$$
where $\overline{a}_1 = \underline{a}_1 + \frac{1}{2}$, $\overline{a}_2 = \frac{(\beta_{i,p})^2}{2} + \underline{a}_2$.

3. Sample $\gamma_{i,p}$ conditional on other parameters and data from:
$$\gamma_{i,p}|\gamma_{-/i,p},\beta_{i,p},\mathbf{z},\mathbf{x},\mathbf{y} \sim Bernoulli\left(\overline{\pi}\right),$$
where $\overline{\pi} = \frac{\pi_0 f\left(\gamma_{i,p}=1|\gamma_{-/i,p};\mathbf{x},\widetilde{\mathbf{y}}\right)}{\pi_0 f\left(\gamma_{i,p}=1|\gamma_{-/i,p};\mathbf{x},\widetilde{\mathbf{y}}\right)+(1-\pi_0)f\left(\gamma_{i,p}=0|\gamma_{-/i,p};\mathbf{x},\widetilde{\mathbf{y}}\right)}$, $\widetilde{\mathbf{y}} = \mathbf{y}-\theta\mathbf{z}$, and $\gamma_{-/i,p}$ denotes the vector γ_p with its i-th element removed (i.e. condition $\gamma_{i,p}$ on all remaining

$n-1$ elements in γ_p). The function $f\left(\gamma_{i,p}=1|\gamma_{-/i,p};\mathbf{x},\widetilde{\mathbf{y}}\right)$ is the likelihood of the model

$$\widetilde{y}_t = y_t - \theta z_t = x_t'\beta_p + \tau\sqrt{z_t}u_t,$$

evaluated assuming $\gamma_{i,p}=1$, and similarly for the function $f\left(\gamma_{i,p}=0|\gamma_{-/i,p};\mathbf{x},\widetilde{\mathbf{y}}\right)$.

4. Sample π_0 conditional on other parameters and data from:

$$\pi_0|\gamma_p,\beta_p,\mathbf{z},\mathbf{x},\mathbf{y}\sim Beta\left(\overline{b}_1,\overline{b}_2\right),$$

where $\overline{b}_1 = n_\gamma + \underline{b}_1$ and $\overline{b}_2 = n - n_\gamma + \underline{b}_2$, and n_γ denotes the number of elements in γ_p which are one, i.e. $n_\gamma = \sum_i \gamma_{i,p} = 1$.

5. Sample z_t conditional on other parameters and data from:

$$\mathbf{z}|\beta_p,\gamma_p,\mathbf{x},\mathbf{y}\sim GIG\left(\frac{1}{2},\overline{\kappa}_1,\overline{\kappa}_2\right),$$

where $\overline{\kappa}_1 = \left[\sum_{t=1}^{T}\left(y_t - x_t\beta_p\right)/\tau\right]$ and $\overline{\kappa}_2 = \sqrt{2+\theta^2}/\tau$. The p.d.f of the Generalized Inverse Gaussian density is of the form

$$f\left(x|v,a,b\right) = \frac{\left(b/a\right)^v}{2K\left(ab\right)}x^{v-1}\exp\left\{-\frac{1}{2}\left(a^2x^{-1}+b^2x\right)\right\},$$

with $x>0$, $-\infty<v<\infty$, $a,b\geq 0$.

B.2 Convergence diagnostics

This subsection assesses convergence of the Markov chain Monte Carlo algorithm in the baseline application to the U.S. data. In general, for the simple univariate regression model of this application convergence to the posterior is not sensitive to different starting points of the chain (selected randomly), or the size of the burn-in period (which, to some extend, is in line with the point of Geyer, 2011, that a burn-in period is not necessary for finding a good starting point in MCMC samplers).

Convergence of the posterior sampling algorithm is relatively straightforward for the case of the univariate regression models we examine in this paper. However, the parameter draws for different quantiles are quite autocorrelated (as well as cross-correlated, i.e. across quantiles). A simple fix for such an efficiency issue is to do thinning, that is retain only n-th draw from the MCMC chain. The top and middle panels of Figure B.1 show the first order autocorrelation of the draws of β_p for all $p\in[0.05,0.10,...,0.90,0.95]$. This figure reveals that if we save a single draw every 40 iterations of the Gibbs sampler, then this is enough to ensure that autocorrelation among draws is reasonably low.

The bottom panel of Figure B.1 presents the inefficiency factors (IFs) of the posterior estimates of β_p. The IF is the inverse of the relative numerical efficiency measure of

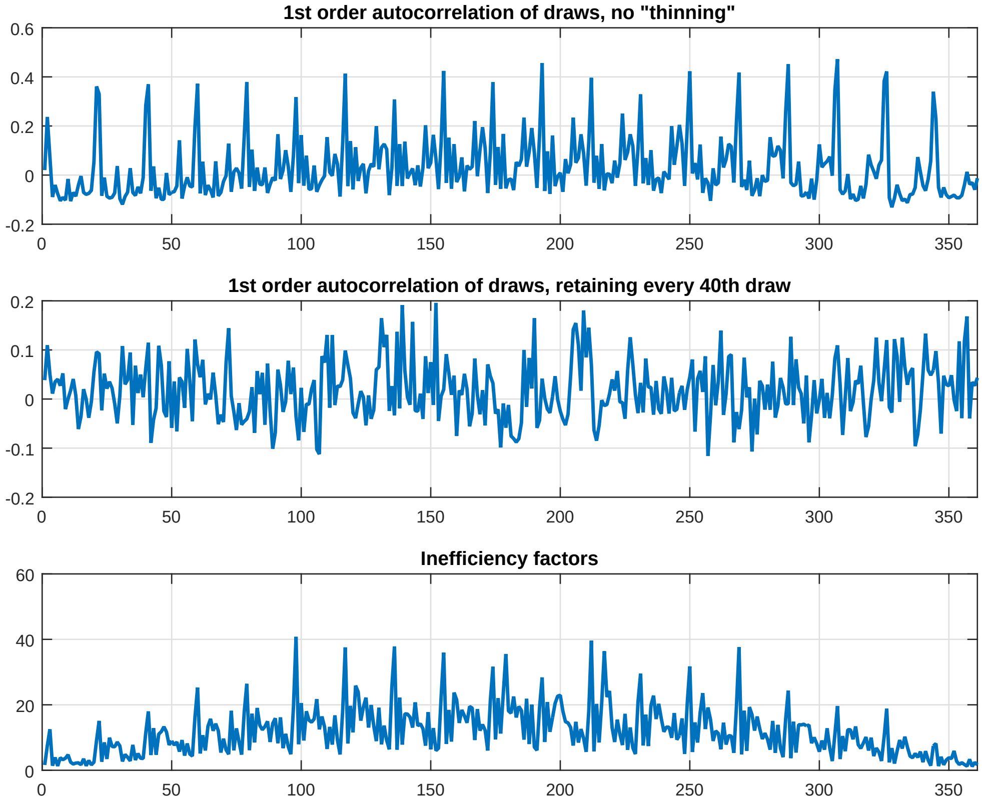

Figure B.1: MCMC diagnostics, β_p coefficients only

Geweke (1992), that is, the IF is an estimate of $(1 + 2\sum_{k=1}^{\infty} \rho_k)$, where ρ_k is the k-th order autocorrelation of the chain which is estimated using a 4% tapered window for the estimation of the spectral density at frequency zero. As a rule of thumb, IFs equal or lower than 20 are considered satisfactory. One can further increase the length of the MCMC chain in order to achieve accuracy, however, in a recursive forecasting exercise this is costly computationally and a balance between precision and computation has to be achieved.