Density forecasts of inflation through Quantile Regression Bayesian Model Average

Pablo BARRIO

September 20, 2024

Contents

1 Introduction

In the field of macroeconomics, the easiest and most used type of predictions are point forecasts. That is, when we predict the value of an indicator several periods ahead (or none for nowcasting). There is a large literature on models for point foreacasts which are usualy simple to interpret. However, point forecasts usually only perform well in normal times. When forecasting risky events, these types of models pale in comparison to what are called density forecasts. Predicting probability densities may be of greater use in macroeconomics, specially during uncertain times. Instead of predicting one value with some degree of certainty given by statistical tests, forecasting densities allows to gain insight of the probabilities associated to different events. For instance, these models may give the probability that inflation may increase or decrease considerably, which is valuable information for central banks for monetary policy making. In fact, international institutions such as central banks are already using predictive densities because of their obvious advantages compared to simple point predictions.

Bayesian Models already allow for such predictions by using the posteriors of coefficients. In "Quantile regression forecasts under model uncertainty", Dimitris Korobilis adds to the bayesian framework by introducing a new method that uses the results from Bayesian Quantile regressions to build density forecasts. In the paper, this method is proven to be superior to the traditional bayesian model average (BMA) and to other less traditional models.

We start by doing a brief literature review in section 2. In section 3, we replicate Korobilis' results and further study the statistical properties of the Quantile Regression Bayesian Model Average (QR-BMA) by working on the same data as used in the paper. Finally, in section 4 we proceed to study the model on simulated data. This allows us to study how the model performs when confronting different datasets with different statistical properties.

2 Literature Review

In "Quantile regression forecasts under model uncertainty", Dimitris Korobilis [3] does real time density forecasts of inflation. This paper stands on two expanding bodies of literature. The first being the research on bayesian estimation, inference and forecast of quantile regressions, and the second being the research on Bayesian Model Averaging (BMA) and selection methods.

A well known reference for Bayesian Model Averaging is "Forecasting U.S. Inflation by Bayesian Model Averaging" (6), a paper in which Wrigth presents it and uses it to forecast U.S. inflation. The main idea behind this bayesian technique is to take the forecasts estimated with different models and to compute a weighted average of these forecasts. The weights correspond to the probability of each model, hence the term "bayesian". This technique can be applied to many different cases. One of the most popular models used for BMA is a mean regression bayesian model. By using this method we shrink down the number of models considered for forecasting to a single form. For instance, one form may be a linear regression with a set of variables multiplied by their respective coefficients. The set of models considered for averaging are the different sets of values for the coefficients of the predictors. The mean regression bayesian model is particularly useful as it allows to estimate these values, as well as their associated probabilities through an MCMC algorithm, which makes bayesian model averaging straightforward.

As previously mentionned, Korobilis' method, Quantile Regression Bayesian Model Average, uses the results from a quantile bayesian regression to obtain predicted quantiles of inflation and build a density from these point predictions through the use of a gaussian kernel. The priors used in the model, which we will present later, present the major advantage of having in-model variable selection. Indeed, the model penalizes variables through a smooth spike and slab, which allows quantile regressions evaluated at different quantile levels to select a different number of variables. The main results from the paper are that when predicting extreme quantiles, the model takes into account more variables. This may explain why the QR-BMA performs better than a benchmark Bayesian Model Averaging with a mean regression bayesian model (3), as the predicted density is built on quantiles, each of them predicted through a different BMA that uses a different set of coefficients for the variables. Said otherwise, each probability level prediction is based on BMA on a different set of variables that best predict the associated quantile, as opposed to the BMA which uses one set of variables

for the whole density prediction.

The density building step from quantiles in Korobilis' paper is based on a non-parametric method introduced by Gagliannone and Lima (1). However, there are other ways to perform this step, such as introduced by Adriano, Boyarchenko and Giannone (ABG, 5). In the former, the density is estimated through the use of a kernel smoother that allows to estimate the density by using quantiles (or quantiles estimates). In the latter, the authors present a method that matches the empirical quantiles to a skewed t-distribution, which allows for an asymmetric distribution. In their paper (2), Mitchell and al. find that the non-parametric method is slightly better than ABG's method by using both on the forecast of U.S. inflation.

3 Quantile Regression Bayesian Model Averaging under uncertainty

3.1 Data

The data we use corresponds to the Real Time Data Research Center of the Philadelphia Fed and the St Louis Federal Reserve Economic Database (FRED) dataset which contains vintages from years 1974q1 to 2015q1. Having vintages allows to simulate real time forecasts at different periods.

The dataset is made of 16 macroeconomic U.S. variables including CPI, GDP, the Industrial Production Index, the Unemployment rate and M1 among others. Every variable is transformed to be stationary. A full list of the variables, as well as their respective transfromations is available in the Appendix (5).

3.2 Models

We are interested in the performance of two models, the QR-BMA from [3], and the BMA, also used in [3] and introduced in [6].

Quantile Regression We start by introducing the frequentist quantile regression which is important to understand how the QR-BMA works. The quantile regression takes the following shape:

$$y_{t+h} = x_t^t \beta_p + \epsilon_t \tag{1}$$

, where:

- h is the forecasting horizon
- β_p depends on the p_{th} quantile of ϵ_t

In frequentist statistics, quantile regressions are estimated by minimizing this error:

$$\hat{\beta}_p = \arg\min_{\beta} \sum_{t=1}^{T} \rho_p(\epsilon_t) \tag{2}$$

$$\rho_p(u) = u \cdot (p - \mathbf{1}\{u < 0\})$$

QR-BMA The Quantile Regression Bayesian Model Average consists in a quantile bayesian regression and kernel smoother that uses the results from the quantile regression to estimate a probability density function (pdf).

As explained in the paper, it has been established in the literature that the minimization problem (2) associated to a frequentist quantile regression is equivalent to the maximization of a asymmetric Laplace likelihood. Furthermore, in [4], it is proven that an asymmetric Laplace distribution can also be represented through a mixture model. In particular, one popular representation is a scale mixture of normals, with the scale parameter following an exponential distribution. The error term in the model is represented as such:

$$\epsilon_t = \theta z_t + \tau\sqrt{z_t}u_t$$

, where:

- $z_t \sim \mathcal{E}xp(1)$
- $u_t \sim \mathcal{N}(0,1)$
- $\theta = \frac{1-2p}{p(1-p)}$ and $\tau = \frac{2}{p(1-p)}$

The quantile bayesian regression is therefore: $y_{t+h} = x_t^t\beta_p + \theta z_t + \tau\sqrt{z_t}u_t$.

From this representation, Korobilis is able to build a bayesian model, which is characterized by the following likelihood;

$$\mathcal{L}(y|x,\beta_p,z) = \left(\prod_{t=1}^{T} z_t^{-\frac{1}{2}}\right) \times \left(-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_{t+h} - x_t^t\beta_p - \theta z_t)^2}{(\tau\sqrt{z_t})^2}\right)$$

and the following priors:

- $\beta_{i,p}|\gamma_{i,p},\delta_{i,p} \sim (1-\gamma_{i,p})\mathcal{N}(0, 0.00001 \times \delta_{i,p}^2) + \gamma_{i,p}\mathcal{N}(0,\sigma_{i,p}^2)$
- $\delta_{i,p}^{-2} \sim \mathcal{G}amma(a_1, a_2)$
- $\gamma_{i,p}|\pi_0 \sim \mathcal{B}ernouilli(\pi_0)$
- $\pi_0 \sim \mathcal{B}eta(b_1, b_2)$

The β_p follow a smooth spike and slab prior. This ensures that, as we run the estimation of the model, the associated $\beta_{i,p}$ of some variables will have a posterior very close to zero and some others not. This prior works as a penalization parameter (equivalent to a Ridge penalization parameter) in such a way that it selects the variables that best forecast a certain forecast of the target. This is one of the advantages of this model highlighted in the paper. The model is able to select the best predictors and estimate the forecasts.

Moreover, δ^{-2} follows an inverted gamma distribution. We can choose to have a very informative prior or a non-informative prior. The author does not provide information on how he chooses his hyperparameters, however, we have noticed that the model best predicts and selects variables when the prior is rather informative and we allow for δ to have high values. This will be further discussed in the next section.

As for the hyper-parameters of π_0, this choice is rather similar to choosing the parameter of a ridge or lasso regression. By allowing π_0 to have a low value (under 0.5 for instance), we increase the probability that the coefficient of any variable will follow a normal distribution centered around zero and with a very weak variance, meaning we increase the probability that the coefficient will be very close to 0, and therefore we decrease the probability of the variable being selected by the model. We find that an informative prior with low values for π_0 best predicts inflation. Again, this will be further discussed in the next section(4).

To estimate the density of the forecast, we use the Bayesian Model Average of the different quantile forecasts. For example, for the 0.1 level quantile forecast, we simply take the average of the sampled $\beta_{0-.1}$ and multiply it by the regressors to obtain the quantile forecast. Once, we have obtained this quantile forecasts for a range of different

probabilities (0.05, 0.1, ..., 0.9, 0.95), we use a gaussian kernel smoother to estimate the pdf. However, this only allows us to get an estimate between 0.05 and 0.95 probability levels. To obtain en estimate of the tails we simply match the extreme quantile levels of a gaussian distribution to the ones we have found.

BMA The main idea behind Bayesian Model Average is to take the average of forecasts done with different models, weighted by the likelihood/probability of each model, hence the term "bayesian". This technique can be applied to many different settings. For the sake of simplicity, we use BMA on a bayesian mean regression that follows the same priors as the QR-BMA. Moreover, as it is a mean regression, we use a normal likelihood centered around zero with variance $sigma^2$ that we fix to one. We could have added a prior for the $sigma^2$, but since we want to compare the QR-BMA which does not present a variable for the variance of the error term, we fix the variance to the empirical variance of the available sample of the target, so that it is adjusted to the scale of the model.

As for the density forecast, we add a step in the Gibbs sampler that draws the point forecast from the likelihood by taking into account the last drawn values of β at each iteration. This allows to estimate the conditional distribution of inflation as such:

$$(y_{t+1}|x_t) \sim \pi\left(y_{t+1}|x_t\right) = \int \pi(y_{t+1}, \beta|x_t)d\beta = \int \pi(y_{t+1}|x_t, \beta) \times \pi(\beta|x_t)d\beta \qquad (3)$$

Once we have obtained these estimates of inflation, we simply take the empirical quantiles of the sample and use a gaussian kernel smoother to estimate the density.

3.3 Results

In this section, we compare our results using the QR-BMA and the BMA to forecast inflation at different horizons. we focus on one-step ahead forecasts as well as six-steps ahead forecasts so that we may compare the performance of both models at different horizons. As in the original paper, we would like to show that the QR-BMA performs better when forecasting densities, but that it is less the case when forecasting points. For this purpose, we evaluate both models performances when doing both types of forecasting. Because of the time it requires running a Gibbs sampler for a range of quantiles for all vintages (1975Q1-2015Q3), we restrict the window for evaluation to

2008Q1-2015Q3.

Point Forecasts

To evaluate the performance of the models when doing point forecasts we use the Mean Squared Forecast Error metric. We compare point forecasts at the 0.5 quantile level for the QR-BMA, and forecasts from the BMA.

Model	**MSFE**	
	h=1	*h=6*
QR-BMA	7.05	6.5
BMA	11.23	8.04

Table 1: Point forecasts Results

It seems that the QR-BMA and the the BMA perform in a similar manner when doing point forescasts. This is not surprising, as the model used for BMA corresponds to a simple special case of the QR-BMA for the 0.5 quantile level. As a matter of fact, the main difference between both models, apart from the simplification of the priors on the mean regression model, is that the QR-BMA uses an asymmetric Laplace distribution likelihood, whereas the BMA uses a normal distribution likelihood. As explained previously, we should be able to find a more prominent difference between both models when comparing their density forecasts.

In Korobilis' paper, both models are compared to a benchmark AR(2), for which the Mean Squared Forecast error between 1975 and 2015 is 5.42 for a one step ahead forecast, and 5.17 for a six steps ahead forecast. Both models perform better than the AR(2), having lower MSFEs. Given our results, it seems it also is the case for our models. However, we only tested the models on the period between 2008 and 2015, and therefore we cannot give a proper conclusion on this matter.

Density Forecasts

As for evaluating the performance of QR-BMA for density forecasting, we compare it to the predictive density of BMA when forecasting inflation one quarter ahead and six quarters ahead between 2008 and 2015. Density forecasts from the QR-BMA and the BMA regression are made using the same method introduced in (1). We use the mean

log predictive scores (MLPS) to evaluate both models.

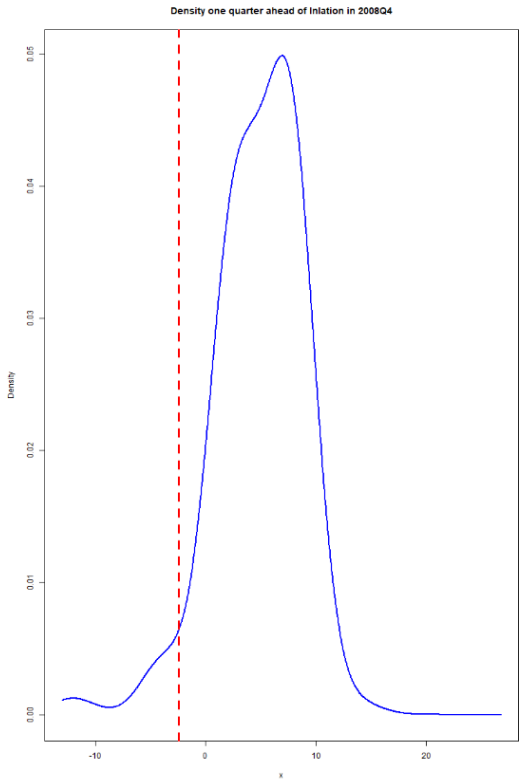
Model	**MLPS**	
	h=1	*h=6*
QR-BMA	-3.3	-2.99
BMA	-4.4	-4.1

Table 2: Density forecasts Results

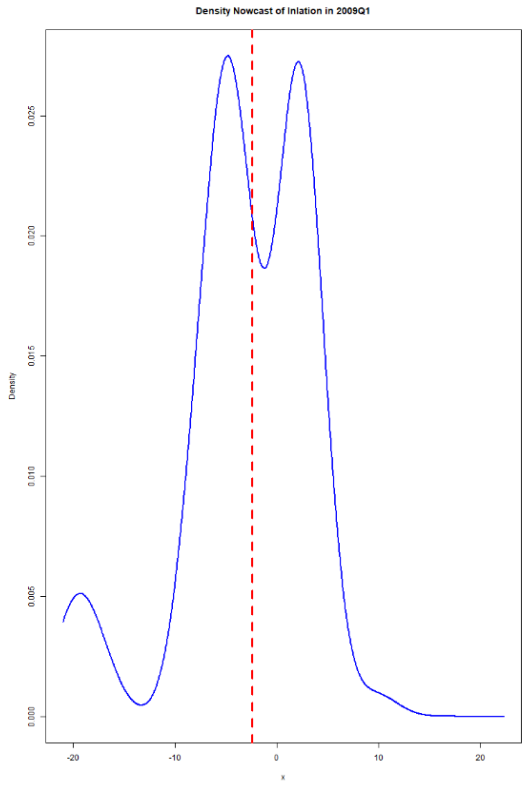
As exposed in the article by Korobilis, the Quantile Regression Bayesian Model Average does perform overall better than the Bayesian Model Average. However, we do not find the same results as in the paper. Our version of the model has more trouble forecasting inflation overall, and specially in periods of crisis. Also, both models have a better MLPS than the one indicated in in the paper for the AR(2). But, again, even if this is encouraging, we cannot infer much, given that the periods on which we estimated the models are different.

Below (1) is the density forecast of 2009Q1 at two different dates. The first one corresponds to the forecast done during the fourth quarter of 2008 (2008Q4) and is a one step ahead forecast. The second corresponds to a nowcasting, forecast during the 2009Q1.

These two density predictions are informative of how our model behaves. In the first one, we notice that our model is not able to predict the decrease of inflation in 2009Q1 one quarter ahead, whereas Korobilis has better results for the same forecast. In the second one, we notice that our model is able to nowcast the decrease of inflation. This is due to the fact that in 2009Q1, the data from 2008Q4 is available, which is when inflation first decreased. Therefore, we conclude that our version of the model does take into account properly new data, but still pales in comparison to the one presented in the paper. Moreover, when predicting the density at different dates and different horizons, we notice that the density is usually symmetric, of the same shape and variance, but its mean changes. This can be seen in the two plots above. However, the results presented in the paper showcase different shapes and variance for the densities predicted at different dates. In order to clarify the difference in results with the paper, we have tried different variations of the model.



(a) Forecast of Inlfation 1-step ahead, 2008Q4



(b) Nowcast of Inflation, 2009Q1

Figure 1: Forecast and Nowcat of Inflation during 2009Q1

Discussion

As a first step, we tried running the Gibbs sampler with a larger number of iterations (100 000). It seems the sampler converges quite fast to a stationary distribution and the results when increasing the number of iterations do not change significantly. We compared different numbers of iterations and found that 5000 iterations and 500 burn are enough. Doing more iterations than 5500 does not change quantile estimates considerably.

Moreover, as done in the paper, we try using only every 40th iteration to avoid auto-correlation between the samples. This reduces the auto-correlation of the sample but, again, does not change the results.

Furthermore, in forecasting, it is usual, specially when using linear models, to sub-

tract the mean and divide by the standard deviation (standardize data). We tried using non-standardized data, as well as standardized data, and the best results are obtained when using the non-standardized data. It is surprising because the outputs of these cases are very different, and when using standardized data, quantiles estimates are decreasing in the quantile level. Therefore, we use the original stationary data and do not perform any additional transformations. This might be due to the scale of the asymmetric Laplace distribution being adjusted to the scale of the stationary data and not to the standardized data.

Indeed, another specification that deserves some attention is the asymmetric Laplace distribution specified for the error term. Usually, when doing mean bayesian regressions, the variance of the error term (of the likelihood), is not fixed, but rather, it is given a prior. In this case, the asymmetric Laplace variance is fixed and is not given the chance to adapt to the data and a prior. We find this rather surprising because this is not mentioned in the paper and we think it deserves some attention.

Lastly, we try different sets of priors, from non-informative priors to very informative priors. We find that, even if the results do not showcase a stark difference when using different priors, the best ones given this dataset are a high variance δ_p and a low value for π_0. The former allows for the β_p to have access to a wider range of values when running the Gibbs sampler, and the latter decreases the value of γ, making the model more restrictive when selecting variables in the model. Even though we tried several sets of priors [1], performing a grid-search over both priors might have been a better option. However, this approach is computationally intensive and time-consuming, as it would require running the Gibbs sampler for all possible variations. Additionally, we consistently apply the same priors across different quantile-level predictions for a given date. It may, however, be beneficial to assign distinct priors to each quantile level to enhance the performance of the quantile forecasts. Both of these improvements could be explored in future research. We further study how the model behaves when given different priors in the next section with simulated data.

[1] (0.1, 0.1), (2, 0.1), (2, 1), (1, 10) for the hyper-parameters of the γ prior and (1, 1), (5, 1), (1, 5), (1, 20) for the hyper-parameters of the π_0

4 Simulated Data

To study the properties of the QR-BMA, we simulated data according to the following procedure:

Model Specification We consider a linear model of the form:

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta} + \epsilon_t, \quad t = 1, 2, \ldots, T$$

where y_t is the dependent variable at time t, $\mathbf{x}_t$ is a k-dimensional vector of regressors, $\boldsymbol{\beta}$ is a k-dimensional vector of coefficients, and ϵ_t is an error term. The error term ϵ_t is assumed to follow a normal distribution with mean zero and variance σ^2, i.e., $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

Parameters and Data Generation Process

- **Number of Observations (T)**: We set the number of observations to $T = 200$, which reflects a moderate sample size for time series analysis.
- **Number of Regressors (k)**: The dimension of the regressor vector $\mathbf{x}_t$ is set to $k = 100$. This relatively high dimensionality is chosen to explore the behavior of the model in the presence of a large number of predictors.
- **Correlation Structure of Regressors**: The regressors $\mathbf{x}_t$ are generated from a multivariate normal distribution with a mean vector of zeros and a specific correlation structure. The covariance matrix of the regressors is defined as a Toeplitz matrix, where the correlation between two regressors $x_{t,i}$ and $x_{t,j}$ is given by $\text{corr}(x_{t,i}, x_{t,j}) = \rho^{|i-j|}$, with $\rho = 0.75$. This structure introduces a decaying correlation between regressors as the distance between their indices increases.
- **Coefficient Vector ($\boldsymbol{\beta}$)**: The coefficient vector $\boldsymbol{\beta}$ is sparse. Specifically, $k - s$ elements of $\boldsymbol{\beta}$ are set to zero, while the remaining s elements are drawn from independent standard normal distributions, i.e., $\mathcal{N}(0, 1)$. This sparsity in the coefficient vector is intended to reflect a situation where only a subset of the regressors have a meaningful impact on the dependent variable.
- **Error Term (ϵ_t)**: The error term ϵ_t is independently drawn from a normal distribution with mean zero and variance σ^2. The standard deviation σ is set to

1, providing a controlled level of noise in the simulated data.

This simulation setup provides a dataset that reflects the characteristics of a high-dimensional linear regression problem with correlated predictors and sparsity in the true model coefficients.

4.1 Study of priors

We first start by studying how the model behaves when changing the prior hyper-parameters. The QR-BMA model relies on two sets of hyper-parameters. The first set corresponds to the a_1 and a_2 that define the prior on the Inverse Gamma distribution of $\delta_{p,i}$, which is the variance of the coefficients $\beta_{p,i}$. The values chosen for these hyper-parameters deeply affect the model's behaviour as it allows $\beta_{p,i}$ coefficients to take higher or lower values. The second set of hyper-parameters on which the model heavily relies are b_1 and b_2, the coefficients of π_0, which follows a Beta distribution. pi_0 is a key parameter in the model as it is the expectation of the Bernouilli distribution of the spike and slab parameter $\gamma_{p.i}$. A lower π_0 leads to a lower $\gamma_{p,i}$, which leads to a higher probability that the associated $\beta_{p,i}$ will be very close to zero. Said otherwise, the hyper-parameters of the π_0 distribution define how restrictive the model is when selecting variables for prediction.

In order to study both these priors, we use a single simulated dataset of 50 observations and 20 predictors from which only 10 truly predict the target variable (the true coefficient β_i is set to zero for the remaining 10 variables). The variance of the error term is fixed to one.

We treat the last 10 observations of the sample as vintage data and study the performance of the model by simulating a real data forecast. For all these pseudo-vintages, we take the available data (between 40 and 49 observations) and run the Gibbs sampler for 5500 iterations (500 are burned) for each decile (0.1, 0.2, ..., 0.9). To evaluate the model we use the following observation as the test or out-of-sample sample. For example, the first sample we train the model on corresponds to the first 40 observations, and we evaluate the performance by computing the SFE and LPS on the 41st observation. These are the different hyper-parameter values we test for:

- non-informative (n-i) variance: $\delta_{i,p}^{-2} \sim \mathcal{G}amma(0.1, 0.1)$
- high variance (hv): $\delta_{i,p}^{-2} \sim \mathcal{G}amma(1, 10)$

- low variance (lv): $\delta_{i,p}^{-2} \sim \mathcal{G}amma(2, 0.1)$
- non-informative (n-i) prior on selection of variables: $\pi_0 \sim \mathcal{B}eta(1, 1)$
- restrictive (r) selection of variables: $\pi_0 \sim \mathcal{B}eta(1, 5)$
- loose (l) selection of variables: $\pi_0 \sim \mathcal{B}eta(5, 1)$

Prior on $\gamma_{p,i}$	**Prior on π_0**	**MSFE**	**MLPS**
n-i	n-i	1.7	-0.169
hv	n-i	3.2	-0.28
lv	n-i	1.75	-0.16
n-i	l	1.98	-0.29
n-i	r	1.34	0.43
hv	r	2.8	-0.07
hv	l	3.19	-0.16
lv	r	1.38	0.76
lv	l	1.8	0.08

Table 3: Results across different priors

We can gather some insights from this results. First of all, we may notice that the worst results are always given when the $\gamma_{p,i}$ is given high values (hv), i.e. when we give a high variance to the coefficients $\beta_{p,i}$. This may be due to the fact that the true $\beta_{p,i}$ have low values, between -1.7 and 2.5, but also that the given simulated data is not challenging. Indeed, with a standard normal distribution, the error term does not induce much volatility into the target, and thus the target is easily well explained by the linear regression (giving a higher R^2 coefficient).

On the other hand, we also notice that when using a "restrictive" prior for the π_0, the model consistently performs better than when using loose "prior". This makes sense since only half of the regressors are really used in the true regression. If we look at the mean of the sampled $\gamma_{p,i}$ when we use a "loose" prior, all of them have a mean around 0.9. This means that the model is using the associated variable for prediction 90 percent of the time. If we do the same for the "restrictive" case, most means are

slightly below 0.5 apart from the true regressors which have a mean around 0.7 for the sampled $\gamma_{p,i}$.

Also, it seems that the non-informative priors do not give bad results. This means that the model itself, without priors, may give satisfactory results but that it is only able to reach a certain level of precision when forecasting. Giving the model an informative prior that is consistent with the true regression improves the results.

The best priors for this set of simulated data is the low variance and restrictive one. As explained above, this is not surprising given the true characteristics of the data itself (a low variance of the error term and low values of the true β, and only 10 true regressors).

A better way to study this would be to do a grid-search on the hyper-parameters but not enough time for it as computations are very cumbersome.

4.2 Study on scarce data

We are interested in studying how the model behaves when confronting different data. We are particularly interested in evaluating the model when data is scarce, meaning the sample is small. For that purpose, we run the QR-BMA on two different datasets. The first one has 200 observations, and the second one has 50 (dataset used in previous section). We keep all other parameters of the simulated data fixed across the two samples so that we are sure differences between estimates are due to the size of the dataset. We use 20 regressors, from which only 10 truly predict the target, and the other 10 have zero valued coefficients in the true regression. We set the variance of the error term to one.

Dataset size	**Priors**	**MSFE**	**MLPS**
50 observations	lv | r	1.7	-0.169
200 observations	lv | r	0.9	1.72

Table 4: Results across different datasets

For simplicity we compare directly the performance of both models using the same priors which are the best performing priors in both cases. We can quickly notice that the model performs much better when given a bigger sample with an MSFE that is almost divided by two and a much higher MLPS.

4.3 Study on challenging data

As a last study, we explore how the model performs when it is confronted by a more challenging dataset. For that purpose, we keep a low sample size of 50, as used previously. We also use 30 variables as regressors, instead of 20, from which only 5 are true predictors, instead of 10. We also increase the variance of the error term so that it is equal to 5. Finally, we also change the covariance that defines how regressors are simulated. We reduce the 0.75 correlation coefficient to be equal to 0.65, so that we reduce correlation between variables.

We proceed as previously, we simulate a real data procedure with vintages with multiple training samples and tests samples. We obtain the following results when using different priors:

Prior on $\gamma_{p,i}$	**Prior on π_0**	**MSFE**	**MLPS**
hv	vr	14.6	-1.2
hv	r	10.85	-1.84
hv	l	10.35	-1.36
lv	r	8.9	-2.4
lv	l	8.45	-2.48
lv	vr	15.15	-1.13

Table 5: Results across different priors for challenging dataset

The first thing we can notice is that both metric are much worse than in the previously studied cases. Not only is the MSFE much higher, but also the MLPS is much lower.

How the priors affect the performance of the model with this data is less clear than in the previous case. This is probably due to the fact that the target is difficult to predict, and therefore the model is having a hard time finding the right distribution and point forecasts. Both best density predictions are made by two opposite prior specifications (hv and l, and lv and vr). This may tell us that when the data is very hard to predict, not even specifying the right prior helps improve the forecast. Indeed, the restrictive prior and very restrictive prior should work better than than the other cases given that only 5 out of 30 variables are true predictors.

One final insight that we may gather from these results, is that, after increasing the sigma considerably to 5, volatilty of the target has increased and so has its scale.

However, the quantiles estimated have kept a lower variance overall, very similar to the variance of the previous cases. It seems changing the prior of Z, as discussed in the first section, might be a subject for further research.

5 Conclusion and further research

In this master thesis, I tried to replicate the results found by Korobilis on "Quantile Regression forecasts of inflation under model uncertainty". Even though I do not find the exact same results as him, I provide further insights on how the model he introduces behaves when confronting different situations by simulating data hat has different characteristics. Although the data simulations I provide are quite simple, they allow to enlighten specially how the model reacts when we change the priors, and how the right specification (dependent of the true regression properties) of hyper-parameters allows for a better prediction. however, one missing component from these data simulations is auto-correlation. One could further study the properties of this model by using it on simulated data that contains auto-correlation in the regressors and in the target, which would be more representative of real times series macroeconomics of financial data. Another interesting subject to study could be to see how the model performs when used on non-linear simulated data. This could not only provide insights on the model properties but also give room for a new version of the QR-BMA where a non-linear form is used instead of a linear quantile regression. Indeed, this model creates a framework for other models to be used for density forecasting. For instance, Garch models have been known to work well when forecasting inflation. A new version of the QR-BMA could be a Garch QR-BMA that would maybe perform better when forecasting indicators with high volatility. Finally, it might be interesting the test the model on more recent data to evaluate how it performs and see if it is able to predict the inflation sure of 2022.

References

[1] Wagner Piazza Gaglianone and Luiz Renato Lima. *Constructing Density Forecasts from Quantile Regressions*. 2012.

[2] Aubrey Poon James Mitchell and Dan Zhu. *Constructing Density Forecasts from Quantile Regressions: Multimodality in Macroeconomic Dynamics*. 2012.

[3] Dimitriv Korobilis. *Quantile Regression forecasts of inflation under model uncertainty*. 2017.

[4] Kozubowski T. J. Podgórski K. Kotz, S. *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance*. 2001.

[5] Nina Boyarchenko Tobias Adrian and Domenico Giannone. *Vulnerable Growth*. 2019.

[6] Jonathan H. Wright. *Forecasting U.S. Inflation by Bayesian Model Averaging*. 2003.

Appendix

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

Table 6: Description of variables and their necessary transformations

- Tcode 1: no transformation (levels)
- Tcode 2: first difference of level $x_{i,t} = z_{i,t} - z_{i,t-1}$
- Tcode 4: logarithm, $x_{i,t} = ln(z_{i,t})$
- Tcode 5: first difference of logarithm, annualized, $x_{i,t} = 400(ln(z_{i,t}) - ln(z_{i,t-1}))$

Acknowledgments

I would like to thank Anna Simoni, my supervisor, for providing not only the idea for the subject of this thesis, but also for guiding me through the study.