International Journal of Forecasting 33 (2017) 11–20

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Quantile regression forecasts of inflation under model uncertainty

Dimitris Korobilis

Adam Smith Business School, University of Glasgow, Gilbert Scott Building, G12 8QQ Glasgow, UK

ARTICLE INFO

Keywords:
Bayesian model averaging
Quantile regression
Inflation forecasts
Fan charts

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 to and better calibrated than those from BMA in the traditional regression model. In addition, QR-BMA methods also compare favorably to popular nonlinear specifications for US inflation.

1. Introduction

Quantile regression generalizes traditional least squares regression by fitting a distinct regression line for each quantile of the distribution of the variable of interest. Least squares regression only produces coefficients that allow us 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 the forecast uncertainty. At the same time, quantile regression estimates are more robust to outliers in the dependent variable. Therefore, quantile regression can be used to investigate predictive relationships between the dependent and exogenous variables, when typical regression modelling fails to indicate the existence of predictability in these exogenous variables; see Koenker (2005).

This paper examines the forecasting performance of Bayesian quantile regression. The final aim is to produce quantile forecasts for inflation using several potential explanatory variables, and to examine the role of model uncertainty in quantile forecasts. Bayesian model averaging (BMA) and selection (BMS) methods have been used traditionally to deal with model uncertainty in forecasting regressions. Following Alhamzawi and Yu (2012) and Yu, Chen, Reed, and Dunson (2013), it is shown that applying BMA to the quantile regression model allows each quantile of inflation to be forecast using a different set of predictors, whereas estimation using Bayesian methods is quite straightforward. Using model selection and averaging in a quantile regression setting allows us to approximate complex forms of the posterior predictive density of inflation, in spite of the fact that the quantile regression model specified in this paper is inherently linear.[1] Although a large body of empirical literature using quantile regression exists, applications of (Bayesian) model averaging are scarce. The only exception is the study by Crespo-Cuaresma, Foster, and Stehrer (2011); however, they approximated Bayesian inference by using least squares and the Bayesian information criterion (BIC).

This paper integrates two vastly expanding bodies of literature. On the one hand, there are several studies that have developed estimation, inference and forecasting in (Bayesian) quantile regression models, such as those by

E-mail address: Dimitris.Korobilis@glasgow.ac.uk.

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

http://dx.doi.org/10.1016/j.ijforecast.2016.07.005

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 to other alternatives; see Koop and Korobilis (2012) and Wright (2008), among others.

Our empirical evaluation of the quantile regression BMA method is based on the real-time forecasting of quarterly US consumer price index inflation, observed over 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 compare our results to Bayesian model averaging in the mean regression specification, as well as to various popular nonlinear regression specifications that have been shown to forecast inflation well. Based on predictive likelihoods (Geweke & Amisano, 2010), the quantile regression BMA provides density forecasts that are superior to those from either regular regression BMA or naive quantile regression methods without BMA.

The next section presents the Bayesian quantile dynamic regression model and the BMA prior, and Section 3 presents the empirical results. Section 4 concludes the paper and discusses further extensions.

2. Bayesian quantile regression

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

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

where x_t is a $n \times 1$ vector of explanatory variables and own lags, and β_p is a vector of coefficients that depends on the pth 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 the estimation of β_p is the solution to the minimization problem

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

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

The main contribution of Yu and Moyeed (2001) was to show that the minimization problem shown in Eq. (2) is equivalent to maximizing a likelihood function under the asymmetric Laplace error distribution; see also Tsionas (2003). Reed and Yu (2011) recently established, both theoretically and empirically, that the asymmetric Laplace likelihood provides accurate approximations of the true quantiles of many distributions with different properties. At the same time, Kotz, Kozubowski, and Podgórski (2001) showed that the asymmetric Laplace distribution can admit various mixture representations. In Bayesian analysis, one popular representation is that of a scale mixture of normals, with the 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 a complex form (as is the case when the likelihood is asymmetric Laplace, no matter what the prior is), one can sample from these conditional posteriors by relying on the Gibbs sampler (Reed & Yu, 2011). When the conditional likelihood admits a normal or a mixture of normals form, these conditional posteriors belong to known distributions, and thus, are easy to draw samples from; see the technical appendix for details.

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

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

where $z_t \sim \textit{Exponential}\,(1)$, that is, a variate from an exponential distribution with rate parameter one, and u_t has a standard normal distribution. In this formulation it holds that $\theta = (1-2p)/p(1-p)$ and $\tau^2 = 2/p(1-p)$, for a given quantile $p \in [0, 1]$. Substituting the formula for ε_t into Eq. (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 of the form

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

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

Given this likelihood formulation, we can now define the 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 of the posterior moments of β_p exist when the prior for β_p is normal. In this paper, we consider the conditionally normal prior

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

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

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

$$\left(\pi_0\right) \sim \textit{Beta}\left(b_1, b_2\right), \tag{8}$$

where $\underline{c} \to 0$ is a fixed hyperparameter. This multilevel prior specification for $\beta_{i,p}$, $i = 1, \ldots, n$, is a mixture of normals prior. Whenever the indicator variable $\gamma_{i,p} = 1$, $\beta_{i,p}$ has a normal prior with variance $\delta_{i,p}^2$. When $\gamma_{i,p} = 0$, $\beta_{i,p}$ 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 in this paper, $\underline{c} = 0.00001$). Such an extremely informative prior means that the predictor $x_{i,t}$ is not relevant for the pth quantile. The indicators $\gamma_{i,p}$ are estimated from the data, and therefore have their own Bernoulli prior with probability π_0. In addition, we avoid selecting the hyperparameters π_0 and $\delta^2_{i,p}$ subjectively by introducing 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 be sampled from easily using Markov Chain Monte Carlo methods. In particular, we sample from the posteriors of each unknown parameter sequentially, conditional on all other parameters, using a standard Gibbs sampler algorithm, which is provided in the technical appendix. For the results in the next section that refer to the full sample of data, we have used 50,000 Monte Carlo iterations: 10,000 iterations are discarded for convergence, and only every 40th draw from the remaining 40,000 is retained; for a justification of this approach and the selection of draws, see Appendix B. For the more demanding recursive forecasting exercise, which is also outlined in the next section, we use 22,000 iterations, of which 2000 are discarded, and only every 20th draw is stored (giving a total of 1000 draws from the posterior densities used for inference). In both cases, the 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

This section examines whether QR-BMA can provide point and density forecasts that are superior to those from popular linear and nonlinear specifications that have been considered successful for forecasting inflation. We consider real-time CPI data 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 measured in real time if possible; otherwise, their final vintage is used if they are not subject to revisions (e.g., interest rates). Other important variables could also be used as predictors (e.g., surveys); however, either they 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 forecasting purposes, the model in Eq. (1) is rewritten as

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

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

As a comparison, we also estimate and forecast using several alternative specifications that have been shown to forecast inflation well (see also Pettenuzzo & Timmermann, 2015):

1. BMA regression (Wright, 2008): The standard "mean" regression model is of the form
   $$y_{t+h} = x_t'\beta + \varepsilon_{t+h},$$
   where there is now 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 Eqs. (5)–(6), which is now applied to β instead of β_p (and posterior expressions are a special case of the ones derived for the BMA quantile regression). The 16 predictors that are defined for the benchmark QR-BMA are also those used in the estimation of the BMA regression.
2. UCSV regression (Stock & Watson, 2007): The unobserved components stochastic volatility model is already used as a benchmark model for forecasting inflation. It takes the form
   $$y_{t+h} = c_t + \varepsilon_{t+h}$$
   $$c_t = c_{t-1} + u_t,$$
   where c_t is trend inflation and the disturbance terms ε_{t+h} and u_{t+h} have stochastic volatilities (such that the log-variances follow random walks).
3. CPS regression (Cogley, Primiceri, & Sargent, 2010): This is the UC-SV model of Stock and Watson but with one autoregressive term for inflation, y_t, which also has a time-varying coefficient.
4. TVP-DMA regression with dynamic model averaging (Koop & Korobilis, 2012): This model generalizes the UC-SV and CPS models by allowing the inflation to depend on further predictors. The same 16 real-time predictors are used for the estimation of this model as for the benchmark QR-BMA and BMA regression models. Koop and Korobilis (2012) suggest that overparameterization concerns (especially compared to the parsimonious UC-SV model) be dealt with by performing Bayesian model averaging at each point in time, leading to a dynamic model averaging (DMA) scheme.

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

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

Predictor	1st lag	IPM	HSTARTS	CUM	RINVBF	ROUTPUT	RUC	ULC	WSD	DYS	NAPMNOI
	Predictors at horizon $h = 0$ (nowcasting)										
$p = 0.05$		•		•	•	•		•	•		•
$p = 0.25$	•									•	
$p = 0.50$	•									•	
$p = 0.75$	•									•	
$p = 0.95$		•					•				•
	Predictors at horizon $h = 4$										
$p = 0.05$	•	•		•	•	•		•	•		•
$p = 0.25$	•									•	
$p = 0.50$	•									•	
$p = 0.75$	•										
$p = 0.95$	•	•					•				•
	Predictors at horizon $h = 8$										
$p = 0.05$	•	•		•		•		•	•		
$p = 0.25$	•		•			•				•	
$p = 0.50$	•		•							•	
$p = 0.75$	•		•							•	
$p = 0.95$	•	•					•				

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

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

3.2. How does QR-BMA work?

This subsection clarifies certain features of the QR-BMA algorithm, and explains why it may be potentially useful for forecasting inflation. The first interesting exercise is to pin down the relevant predictor variables that are 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). In addition, 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. In particular, Koop and Korobilis (2012) showed that the predictors of inflation are extremely unstable over time, so that the predictors that are relevant for different samples are expected to be quite different.

The table shows which predictors have a probability of inclusion in the final regression of more 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, George, & McCulloch, 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.

There are two main messages to be garnered from Table 1: (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 that by 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 become relevant as the forecast horizon increases. The second message, though, is a novel one in this literature, and an encouraging one. It says that several variables potentially have predictive ability in times of extreme changes in inflation rates, that is, changes that are beyond the "median expectation" of consumers and/or the central bank.

In order to understand why such is 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 that followed the collapse of the global banking system and the turbulence in commodity markets. Note that the largest drop in monthly data 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 of these events are very important for policy-makers, so that accurate predictions of their occurrence are of paramount interest.

The top and bottom panels of Fig. 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), together with the predictive distribution of the QR-BMA,

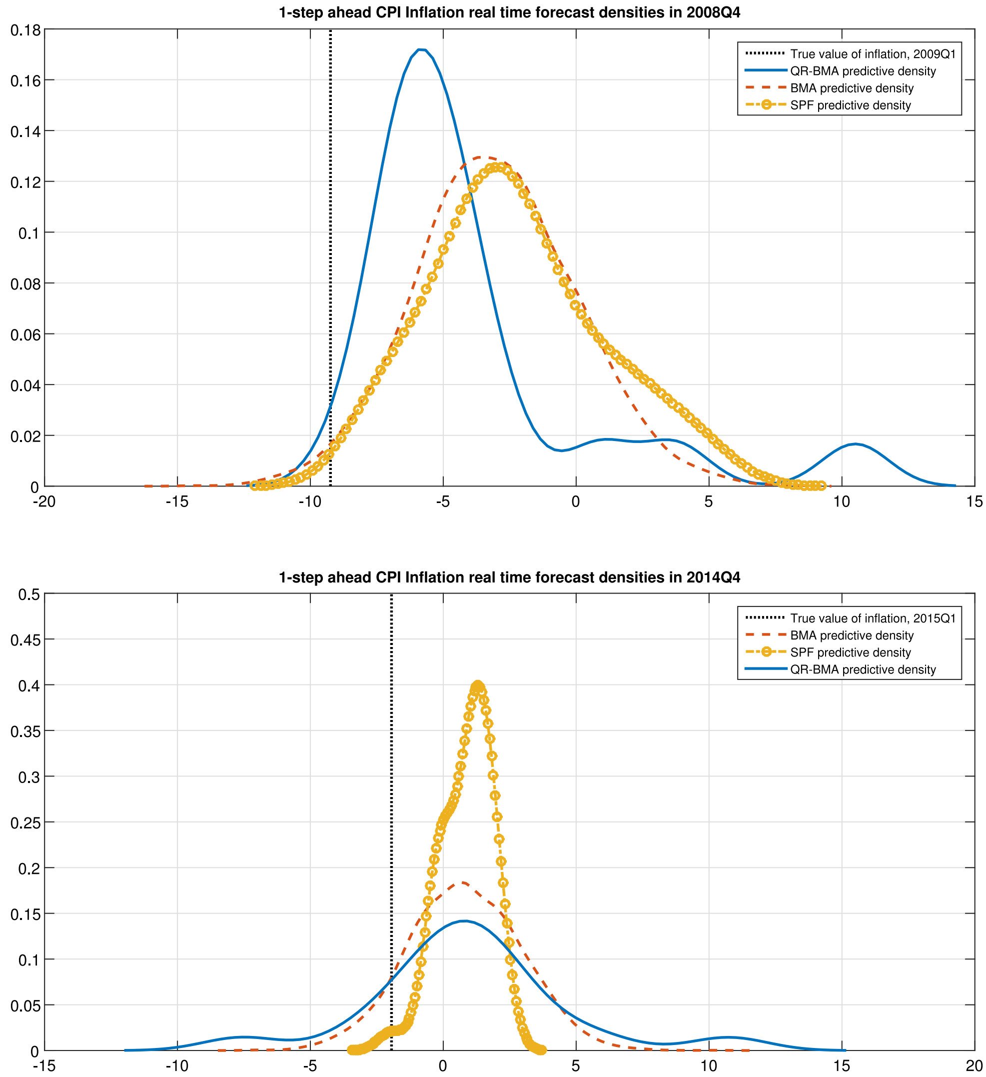

Fig. 1. Predictive densities for two abruptly occurring deflation events in US data.

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 that of the mean regression BMA very closely. In this specific case, the QR-BMA places quite a lot of weight on highly negative outcomes, and a considerably lower weight on positive outcomes for inflation. Allowing this kind of multimodality means that the QR-BMA distribution places 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 to placing equal weights on highly negative and highly positive outcomes.

A slightly different story is observed in 2014Q4. Again, the QR-BMA distribution is multimodal, but this time it looks symmetric. However, strikingly, while the SPF distribution is a mixture with possibly three modes, it is quite concentrated around a certain value, 1%. The forecast disagreement is quite low, but the vast majority of forecasters failed to account for the possibility of disinflation. In contrast, there is a mode with a 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 to the true realization of inflation than the SPF distribution does.

3.3. Forecasting results

We evaluate the forecast performance of each model by considering a recursive pseudo-out-of-sample (poos) procedure: we start by estimating model parameters for the

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

Table 2
Forecasting results, 1975q1–2015q3.

	$h=0$	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$
	Mean squared forecast error								
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								
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: The 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 show those of competing models relative to that of the AR(2). Values lower (higher) than one mean that the respective model performs better (worse) than the AR(2). The 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 for equal predictive accuracy in terms of both 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 MSFE (MLPS) differentials indicate that the point (density) forecast performance of the respective model is significantly better than that of the AR(2) at the 1% (***), 5% (**), and 10% (*) levels.

1975Q1 vintage (sample is 1947Q1–1974Q4), nowcasting the 1975Q1 observation and forecasting out-of-sample for each horizons $h = 1, 2, 3, 4, 5, 6, 7, 8$. We then use the 1975Q2 vintage (the sample is 1947Q1–1975Q1) 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, namely 2015Q4 (for the sample 1947Q1–2015Q3).

An initial assessment of the forecasting performance can be provided by looking at point forecasts. While the quantile regression model is naturally designed to allow for more complex predictive densities, point forecasts are still the most popular way for policy-makers to communicate to the public their expectations about inflation. The top panel of Table 2 presents mean squared forecast errors (MSFEs) for these. In particular, the first row presents the performance of an AR(2) model for inflation, and the subsequent rows provide the MSFEs relative to that of the AR(2). Values lower than one show that the respective model generates point forecasts that are superior to those of the AR(2). Asterisks next to the relative MSFEs show that the respective model performs significantly better than the AR(2) at the 1% (***), 5% (**), and 10% (*) levels, 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 on traditional BMA to the extent that it is comparable to the nonlinear UCSV specification of Stock and Watson (2007). Regarding the three time-varying parameter specifications, the results differ slightly from those found by Koop and Korobilis (2012) using a shorter sample (which did not include the period 2009–2015) and slightly different set of predictors (they also included measures such as inflation expectations). In particular, the UCSV specification seems to perform the best for longer term forecasts ($h > 6$), even though this specification has no explanatory variables. In addition, the nonlinear CPS and TVP-DMA specifications do not perform that well compared to the benchmark AR(2), even though previous evidence (Koop & Korobilis, 2012) suggests that it is nonlinearity that matters for short-term forecasts of inflation (while there is scant evidence that inflation is affected by the predictors).

A natural second step in the analysis is the evaluation of 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 between each model and the AR(2). As with the MSFEs, asterisks show significant differentials at the 1% (***), 5% (**), and 10% (*) levels, based on the Diebold–Mariano statistic. The results clearly indicate that QR-BMA offers substantial improvements when considering the whole distribution of forecasts. The improvements are particularly evident at short horizons (nowcasting and $h = 1$), where surprise movements of inflation can result in the failure of forecasting models (see also the discussion in the previous subsection and Fig. 1).

Additional insights can be provided by the probability integral transform (PIT), which is used here 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)

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

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: The entries are p-values of the respective statistics for the probability integral transforms (PITs) of the predictive densities: Kolmogorov–Smirnov (KS), Doornik–Hansen (DH), the Ljung–Box test of the mean (LB1), and the Ljung–Box test of the variance (LB2); see Rossi and Sekhposyan (2014) for more information.

evaluated at the realization y_{t+h}:

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

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. 1975Q1–2015 Q3–h) is independent and identically distributed (i.i.d.) uniform (0,1), and its cumulative distribution function is the 45° line; see Diebold, Gunther, and Tay (1998).

An initial way to evaluate predictive densities involves a visual assessment by means of plotting histograms of the PITs: the closer the PITs look to a continuous uniform distribution, the better calibrated they are. However, there are also more formal metrics that allow formal tests of the 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. In addition, we can also use the result of Berkowitz (2001) that if $z_{t+h} \overset{\text{i.i.d.}}{\sim} U(0,1)$ then $\zeta_{t+h} \equiv \Phi^{-1}(z_{t+h}) \overset{\text{i.i.d.}}{\sim} N(0,1)$, where $\Phi^{-1}(\cdot)$ is the inverse of the normal CDF, denoted by $\Phi(\cdot)$. The Doornik–Hansen (DH) test is used to assess the 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 the LB statistic for testing independence of the mean by LB1, and the statistic for testing independence of the variance of the PITs by LB2.

Table 3 provides diagnostics related to the calibration of predictive densities, and is similar to Table 1 of Metaxoglou, Pettenuzzo, and Smith (2016). The entries are p-values, and values lower than 0.05 indicate rejection of the null hypothesis of the test at the 5% level. In the case of BMA, the KS test rejects the null of uniformity of the PITs for both horizons. For the other models, there is evidence that the PITs are distributed uniformly, even though this evidence is weaker (i.e., p-values that are 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 model that is misspecified 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 the PITs for all models are serially correlated, with the only exception being the variance of the PIT for the UCSV at $h = 0$. These statistics are not illuminating for comparing the various models, but the KS and DH statistics suggest that the quantile regression model generates predictive densities that are as well calibrated as the predictive densities of popular nonlinear specifications for inflation.

Finally, if the policy-maker is interested in a certain quantile of inflation, as is typically the case with Value-at-Risk (VaR) forecasting in finance, several measures are available, such as the DQ test of Engle and Manganelli (2004). In a recent paper, Gerlach, Chen, and Lin (in press) show how to implement Bayesian variants of popular tests proposed in the quantile regression literature. However, 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 the implementation of 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. The results indicate that the quantile regression BMA approach does indeed find that different predictors are relevant for each quantile of inflation, and that superior predictive distributions can be obtained by taking this feature into account.

Appendix A. Data appendix

The real time data are from Philadelphia Fed (https://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data), and the variables that are not revised (final vintages) are from FRED (http://research.stlouisfed.org/fred2/). The dependent variable is CPI (consumer price index, quarterly vintages). All of the variables are transformed to be approximately stationary, implying that inflation rates (which are close to being random walk) are forecast instead of second differences of CPI. 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

Appendix B. Technical appendix

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

The transformed quantile regression model is given in Eq. (4), which we rewrite here for convenience:

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

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

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

$$\left(\delta_{i,p}^{-2}\right) \sim Gamma\left(\underline{a}_1,\underline{a}_2\right) \tag{B.3}$$

$$\left(\gamma_{i,p}|\pi_0\right) \sim Bernoulli\left(\pi_0\right) \tag{B.4}$$

$$\left(\pi_0\right) \sim Beta\left(\underline{b}_1,\underline{b}_2\right), \tag{B.5}$$

where $\left(\underline{a}_1,\underline{a}_2,\underline{b}_1,\underline{b}_2\right)$ are prior hyperparameters that are chosen by the researcher, and $\underline{c}$ is a fixed parameter that is set to be very close to zero. We obtain draws from the posteriors of all the unknown parameters by sampling sequentially from the following conditional distributions:

1. Sample $\beta(p)$ conditional on a knowledge of all other parameters (including 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 the 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 ith 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 is 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

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

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

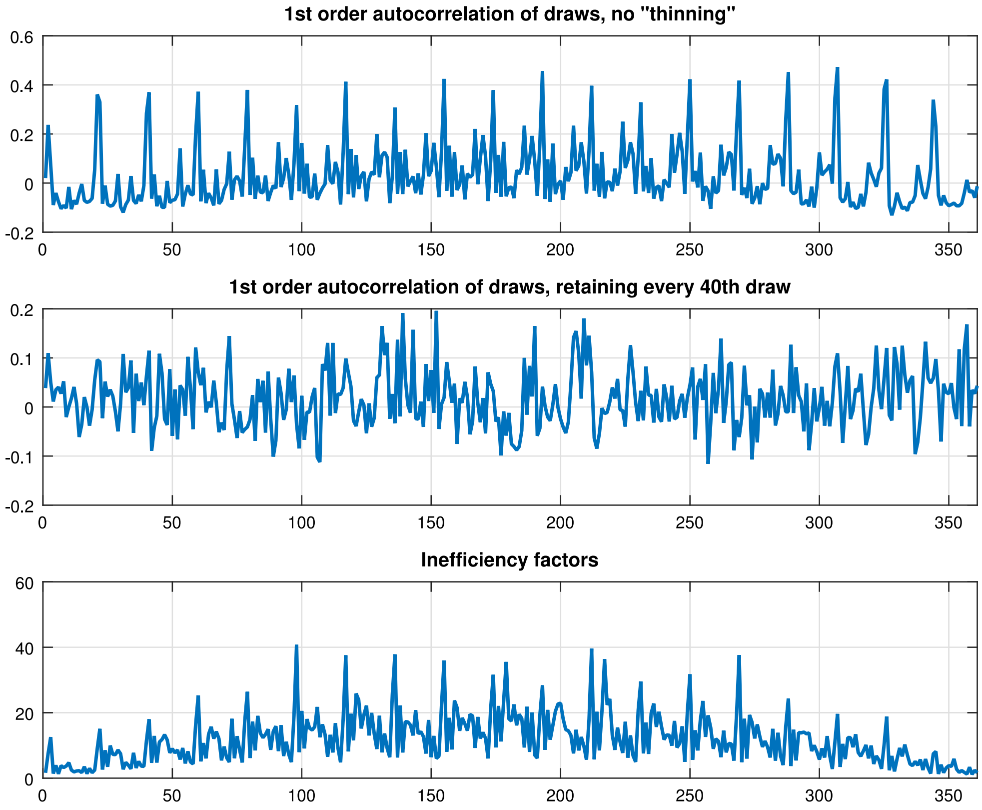

Fig. B.1. MCMC diagnostics, β_p coefficients only.

B.2. Convergence diagnostics

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

The convergence of the posterior sampling algorithm is relatively straightforward for the case of the univariate regression models that 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 perform thinning; that is, retain only the *n*th draw from the MCMC chain. The top and middle panels of Fig. B.1 show the first order autocorrelation of the draws of β_p for all $p \in [0.05, 0.10, \ldots, 0.90, 0.95]$. This figure reveals that if we save a single draw from every 40 iterations of the Gibbs sampler, this is enough to ensure that the autocorrelation among draws is reasonably low.

The bottom panel of Fig. 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 Geweke (1992); that is, the IF is an estimate of $\left(1 + 2\sum_{k=1}^{\infty} \rho_k\right)$, 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 that are equal to or lower than 20 are considered satisfactory. One can increase the length of the MCMC chain further in order to achieve accuracy; however, this is computationally costly in a recursive forecasting exercise, and a balance between precision and computation has to be achieved.

References

Alhamzawi, R., & Yu, K. (2012). Variable selection in quantile regression via Gibbs sampling. *Journal of Applied Statistics*, *39*(4), 799–813.
Barbieri, M. M., & Berger, J. O. (2004). Optimal predictive model selection. *The Annals of Statistics*, *32*, 870–897.
Berkowitz, J. (2001). Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics*, *19*(4), 465–474.
Bernardi, M., Casarin, R., & Petrella, L. (2016). Dynamic model averaging for quantile regression. arXiv:1602.00856.
Chipman, H., George, E. I., & McCulloch, R. E. (2001). The practical implementation of Bayesian model selection. In P. Lahiri (Ed.), *Lecture notes –monograph series*: *Vol. 38*. *Model selection* (pp. 67–116). Beachwood, OH: Institute of Mathematical Statistics.
Cogley, T., Primiceri, G., & Sargent, T. (2010). Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, *2*, 43–69.
Cogley, T., & Sargent, T. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*(2), 262–302.
Crespo-Cuaresma, J., Foster, N., & Stehrer, R. (2011). Determinants of regional economic growth by quantile. *Regional Studies*, *45*(6), 809–826.
Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.

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

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

Geraci, M., & Bottai, M. (2007). Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, *8*, 140–154.

Gerlach, R., Chen, C. W. S., & 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.

Gerlach, R., Chen, C. W. S., & Lin, E. M. H. (2016). Bayesian assessment of dynamic quantile forecasts. *Journal of Forecasting*, http://dx.doi.org/10.1002/for.2408, (in press).

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian statistics 4* (pp. 169–193). Oxford: Oxford University Press.

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

Geyer, C. (2011). Introduction to Markov chain Monte Carlo. In S. Brooks, A. Gelman, & X.-L. Meng (Eds.), *Handbook of Markov chain Monte Carlo*. Boca Raton: Chapman & Hall/CRC.

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

Koenker, R. (2005). *Quantile regression*. New York: Cambridge University Press.

Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, *53*(3), 867–886.

Kotz, S., Kozubowski, T. J., & Podgórski, K. (2001). *The Laplace distribution and generalizations: A revisit with applications to communications, economics, engineering, and finance*. Boston: Birkhauser.

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

Lancaster, T., & Jun, S. J. (2010). Bayesian quantile regression methods. *Journal of Applied Econometrics*, *25*, 287–307.

Meligkotsidou, L., Vrontos, I. D., & Vrontos, S. D. (2009). Quantile regression analysis of hedge fund strategies. *Journal of Empirical Finance*, *16*, 264–279.

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

Pettenuzzo, D., & Timmermann, A. (2015). Forecasting macroeconomic variables under model instability. *Journal of Business and Economic Statistics*, http://dx.doi.org/10.1080/07350015.2015.1051183.

Reed, C., & Yu, K. M. (2011). Efficient Gibbs sampling for Bayesian quantile regression. Manuscript.

Rossi, B., & Sekhposyan, T. (2014). Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set. *International Journal of Forecasting*, *30*, 662–682.

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.

Stock, J. H., & Watson, M. W. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking*, *39*, 3–33.

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

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

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

Yu, K. M., & Moyeed, R. A. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, *54*, 437–447.

Dimitris Korobilis is Associate Professor (Reader) in Economics at the Adam Smith Business School, University of Glasgow. His research interests include time series analysis, Bayesian and computational econometrics, and applied macroeconomics. He is visiting researcher at the Deutsche Bundesbank and has vast experience as a consultant for several central banks, financial companies, and the Scottish Government. He research has been published in *Journal of Econometrics*, *International Economic Review*, and *European Economic Review*.