Quantile density combination: An application to US GDP forecasts*

Knut Are Aastveit[†] Saskia ter Ellen[‡] Giulia Mantoan[§]

September 15, 2022

Abstract

In this paper, we combine density forecasts from Bayesian quantile regressions. We develop a forecast combination scheme that assigns weights to the individual predictive density forecasts based on quantile scores. Compared to standard combination schemes, our approach has the advantage of assigning a different set of combination weights to the various quantiles of the predictive distribution. We apply our approach to US GDP growth forecasts based on quantile regressions using a set of common leading indicators. The results show that density forecasts from our quantile combination approach outperforms forecasts from commonly used combination approaches such as Bayesian Model Averaging, optimal combination, combinations based on recursive logarithmic score weights and equal weights. In particular, our quantile combination approach provides more accurate forecasts for the lower tail of the GDP distribution, measuring downside macroeconomic risk.

Keywords: *Density Forecasts; Forecast Combinations; Quantile Regressions*

JEL classification: *C11, C32, C53, E37*

*This paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of the Norges Bank. The authors would like to thank Laurent Ferrara, Gergerly Gánics, Malte Knüppel, James Mitchell, Tatevik Sekhposyan, Herman van Dijk and seminar and conference participants at Norges Bank, the 2020 Joint Statistical Meeting, the 2020 SNDE symposium, the 2020 International Symposium on Forecasting, the 2021 IAAE conference, the fall 2021 Federal Reserve Bank of St. Louis Applied Time Series Econometrics Workshop and IIF MacroFor. This paper is part of the research activities at the Centre for Applied Macroeconomics and Commodity Prices (CAMP) at the BI Norwegian Business School.

[†]Norges Bank & Centre for Applied Macroeconomics and Commodity Prices (CAMP), BI Norwegian Business School

[‡]Norges Bank Norges Bank & BI Norwegian Business School

[§]The Alan Turing Institute and Warwick Business School, University of Warwick.

1 Introduction

Uncertainty and downside risk plays a prominent role for economic forecasts and policy decisions. To reflect uncertainty around mean predictions, it has become popular for economic forecasters' and policy institutions, particularly for central banks, to provide probabilistic (density) forecasts, see e.g., monetary policy reports of the Bank of England, Norges Bank, Swedish Riksbank, and the Federal Reserve Bank. In recent years, policymakers have shifted their focus from forecasts uncertainty in general to being particularly interested in quantifying macroeconomic downside tail risk, often referred to as GDP-at-risk.[1] Motivated by the concept of value-at-risk from the finance literature, Adrian et al. (2019) study the distribution of macroeconomic risk by estimating a quantile forecast regression of real GDP growth over the next year for various quantiles. Adrian et al. (2019) argue that financial conditions are particularly informative above future downside macroeconomic risk. This has led to a surge of interest in growth-at-risk (e.g. Coe and Vahey, 2020; Reichlin et al., 2020; Carriero et al., 2020; Clark et al., 2021; Brownlees and Souza, 2021).[2] On the other hand, a vast amount of research has shown that a variety of economic and financial variables, such as various survey data, labour market variables, housing related variables, oil prices, stocks prices and the slope of the terms structure, contain predictive information about future economic recessions and downturns, thus being informative of downside macroeconomic risk, see e.g. Marcellino (2006) and Liu and Moench (2016) for an overview.

Concerned with accurate and useful forecasts, forecasters and policymakers routinely rely on multiple sources, employing multiple models and predictors, to produce forecasts. This has spurred a recent resurgence in interest in combination of density forecasts in macroeconomics and econometrics. These new developments range from combining predictive densities using weighted linear combinations of prediction models, evaluated using various scoring rules (e.g. Hall and Mitchell, 2007; Amisano and Giacomini, 2007; Jore et al., 2010; Hoogerheide et al., 2010; Kascha and Ravazzolo, 2010; Geweke and Amisano, 2011, 2012; Gneiting and Ranjan, 2013; Aastveit et al., 2014; Kapetanios et al., 2015a; Ganics, 2017; Ganics et al., 2020), to more complex combination approaches that allows for time-varying weights with possibly both learning and model set incompleteness (e.g. Billio et al., 2013; Casarin et al., 2015; Pettenuzzo and Ravazzolo, 2016; Del Negro et al.,

[1] In the US, the Federal Open Market Committee (FOMC) commonly discusses downside risks to growth in FOMC statements, with the relative prominence of this discussion fluctuating with the business cycle. More generally, macroeconomic downside risk has also been the focus of recent publications and speeches by policy institutions such as the International Monetary Fund (IMF), Bank of Canada and Bank of England.

[2] Note that all these papers uses a current vintage of NFCI in a pseudo out-of-sample framework to form predictions GDP growth. In a recent paper, Amburgey and McCracken (2022) construct real-time vintages of the NFCI. They find additional gains in the predictive content of NFCI for quantiles of GDP growth when using real-time vintages of NFCI, particularly leading up to recessions.

2016; Aastveit et al., 2018; McAlinn and West, 2019a; McAlinn et al., 2020).[3] Common to all these combination approaches, a single weight is attach to the entire predictive distribution for each model. This is limiting as it ignores that some models may be good at forecasting the mean of the distribution but poor in the tails, while other models may provide accurate forecasts in the tail but put less accurate forecasts for the mean of the distribution. The need for a coherent methodology that gives policy makers the flexibility to construct density forecasts that incorporates the heterogeneity in accuracy across regions of the forecast distribution from multiple sources cannot be understated.

In this paper, we address the issues above, by proposing a new alternative forecast combination approach, which aims to obtain overall more accurate density forecasts, by assigning a set of combination weights to the various quantities of the individual density forecasts. To achieve this goal, we first produce individual density forecasts using Bayesian quantile regression models as in Kozumi and Kobayashi (2011). Then we combine the various individual forecasts using a novel quantile combination approach, where each quantile of the combined density forecast is constructed as a weighted combination of the individual forecasts for the corresponding quantile. To account for the heterogeneity in forecast accuracy from the various models across the various parts of the distribution, we allocate the quantile-specific weights from each model using the quantile score by Gneiting and Ranjan (2011). As highlighted by Gneiting and Ranjan (2011), the quantile score is a strictly proper scoring rule, which is a weighted version (decomposition) of the continuous ranked probability score (CRPS).

In an empirical application, we demonstrate the usefulness of our novel quantile combination approach to forecasting real GDP growth rate for the Unites States for the period 1993Q1-2020Q1 using a real-time dataset. We combine predictive distributions from $K = 5$ quantile regression models. Each quantile regression model consists of lagged GDP growth and one additional predictor (with lags). Motivated by the recent paper by Adrian et al. (2019) and the vast literature on predicting economic recessions, we include the following predictors, the National Financial Condition Index (NFCI), the University of Michigan Consumer Sentiment Index (ICS), a credit spread that measures the difference between BAA corporate bond yield and the 10 year treasury yield, residential investments and the unemployment rate.

Our novel quantile combination approach extends the findings of earlier forecast combination and GDP-at-risk literature in several ways.

First, we show that density forecasts from our quantile combination approach outperforms forecasts from commonly used combination approaches such as Bayesian Model Averaging (BMA), the optimal combination of density forecasts (OptComb) suggested by Hall and Mitchell (2007) and Geweke and Amisano (2011), recursive logarithmic score weights as in Jore et al. (2010) and equal weights. This holds irrespective of using the

[3] See Aastveit et al. (2019) for a recent survey on the advances in forecast density combinations.

CRPS or any threshold or quantile weighted version of the CRPS, that emphasize performance in either the centre, left or right tail of the distribution, as a measure of forecast accuracy. The latter therefore indicates that the relative gains in terms of forecasting performance from our model is not specific to observations in a certain region of the distribution or to specific subperiods in our forecasting sample. Instead, we find a steady improvements over time and in all quantiles of the GDP distribution.

Second, for each individual model, we show that forecasts from a quantile regressions outperform forecasts from the linear regression. This complements findings in Korobilis (2017) and Mazzi and Mitchell (2019) that quantile regression methods can be useful for macroeconomic forecasting.

Third, while Adrian et al. (2019) argue that financial conditions are particularly informative about future downside macroeconomic risk, we show that quantile regressions that include variables, such as residential investments and stock prices, provide somewhat more accurate forecasts for the lower left quantile of the GDP distribution than quantile regressions that include the NFCI. This suggests that also other variables than the NFCI are informative about future downside macroeconomic risk and supplements findings in Reichlin et al. (2020) who finds that financial conditions have little predictive content for quantiles conditional on other macroeconomic information.

Finally, our paper is also related to Opschoor et al. (2017) that assess the merits of density forecast combination schemes that assign weights to individual density forecasts based on the censored likelihood scoring rule of Diks et al. (2011) and the CRPS of Gneiting and Ranjan (2011). While in their paper, they use this approach in the context of measuring downside risk (Value-at-Risk) in equity markets using recently developed individual volatility models, our paper differ in three important aspects. First, our combination approach differs as we assign weights to individual density forecasts based on quantile scores. Second, our goal is different as we aim to obtain density forecasts that are overall more accurate for all parts of the distribution and not only for the lower tail. Finally, we focus on forecasting GDP growth, arguably the most important macroeconomic variable, instead of measuring downside risk in equity markets.

The rest of the paper is organized as follows: Section 2 presents our quantile combination approach and the individual quantile regression models; Section 3 presents the data set we use and results from our empirical application. The limitations of our empirical exercise are relaxed in simulated scenarios in Section (4), showing the performance of quantile combinations in controlled environments. Finally Section 5 concludes.

2 Econometric framework

In this section we describe our novel quantile forecast combination approach. Our combination approach aims to obtain overall more accurate density forecasts, by assigning a set of combination weights to the various quantities of the individual density forecasts. To achieve this goal, we first produce individual density forecasts using Bayesian quantile regression models, outlined in section (2.1). Then we combine the various individual forecasts using a novel quantile combination approach, where each quantile of the combined density forecast is constructed as a weighted combination of the individual forecasts for the corresponding quantile, detailed in section (2.2).

2.1 Quantile regression models

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 variables. In that respect, quantile regression is more appropriate for making inferences about predictive distributions and assessing the forecast uncertainty. In principle, we would like to know the entire conditional distribution function that relates the dependent variable with the predictors. However, in practice quantile regression is based on minimizing sums of asymmetrically weighted absolute residuals.

Consider the quantile regression model given by

$$y_t = x_t\beta_q + \varepsilon_t \tag{1}$$

for $t = 1, \ldots, T$. Here ε_t is the error term whose distribution (with density, $f(q(\cdot))$ is restricted to have the qth quantile equal to zero, that is, $\int_{-\infty}^{0} f(q(t))dt = q$. Traditionally, quantile regression estimation for β_q proceeds by minimizing:

$$\sum_{t=1}^{T} \rho_q(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_q), \tag{2}$$

where $\rho_q(\cdot)$ is the check (or loss) function defined by

$$\rho_q(u) = u\{q - I(q < 0)\}. \tag{3}$$

and $I(\cdot)$ denotes the usual indicator function. The set of quantiles provides a more complete description of the response distribution than the mean, making the quantile regression an important alternative to classical mean regression.

Since, however, the check function is not differentiable at zero, we cannot derive explicit solutions to the minimization problem. To solve this issue, we follow Kozumi and Kobayashi (2011) approach to Bayesian quantile regression models using the asymmetric Laplace distribution for the error term.

2.1.1 Predictive quantile function for GDP

For each variable $k = 1, \ldots, K$, a predictive distribution for GDP growth is obtained using a ARDL model:

$$y_{t+h,q,k} = \boldsymbol{x}'_{t,k}\boldsymbol{\beta}_q + \sigma\theta z_{t+h} + \sigma\tau\sqrt{z_{t+h}}u_{t+h} \tag{4}$$

where $\boldsymbol{x}'_{t,k}$ is the vector of lagged values of y_t (with maximum lag r) and of one of the K predictors (with maximum lag p). In our empirical application, the number of lags p and r are selected using BIC selection criterion with a maximum of four lags. Moreover, $q = 1, \ldots, 5$ denotes the respective quantile, set to 10, 25, 50, 75 and 90 in our empirical application. The error term takes the form $\varepsilon_{t+1} = \sigma\theta z_{t+h} + \sigma\tau\sqrt{z_{t+h}}u_{t+h}$ as in Kozumi and Kobayashi (2011), where $z_{t+h} \sim Exponential(1)$, u_{t+h} has a standard normal distribution and $\sigma \sim IG(n_0/2, s_0/2)$. Finally, we will focus on forecast horizons $h = \{1, 4\}$ in our empirical application.

2.1.2 Bayesian Inference

We consider the linear model given by:

$$y_t = \boldsymbol{x}'_t\boldsymbol{\beta}_p + \varepsilon_t \tag{5}$$

where p denotes the quantile and assume that ε_t has the asymmetric Laplace distribution with density:

$$f_p(\varepsilon_t) = p(1-p)\exp\{-\rho_p(\varepsilon_t)\} \tag{6}$$

where $\rho_p(\varepsilon_t) = \varepsilon_t\{p - I(\varepsilon_t < 0)\}$. The mean and variance of the asymmetric Laplace distribution are given by:

$$\mathbf{E}(\varepsilon_t) = \frac{1-2p}{p(1-p)} \quad Var(\varepsilon_t) = \frac{1-2p+2p^2}{p^2(1-p)^2} \tag{7}$$

To develop a Gibbs sampling algorithm for the quantile regression model, we use a mixture representation based on exponential and normal distribution by Kotz et al. (2012). Following Kozumi and Kobayashi (2011) we can present the error term ε_t as:

$$\varepsilon_t = \sigma\theta z_t + \sigma\tau\sqrt{z_t}u_t \tag{8}$$

where σ is the scale parameter, $z_t \sim Exponential(1)$ and $u_t \sim N(0,1)$ are mutually independent, and:

$$\theta = \frac{1-2p}{p(1-p)} \qquad \tau^2 = \frac{2}{p(1-p)} \tag{9}$$

From this we can rewrite y_t as:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_p + \sigma\theta z_t + \sigma\tau\sqrt{z_t}u_t \tag{10}$$

To facilitate the inference, we adopt the reparametrization by Kozumi and Kobayashi (2011):

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_p + \theta\upsilon_t + \tau\sqrt{\upsilon\sigma}u_t \tag{11}$$

where $\upsilon_t = \sigma z_t$. We assume that $\boldsymbol{\beta}_p \sim \mathcal{N}(\boldsymbol{\beta}_{p0}, \boldsymbol{B}_{p0}$ and $\sigma \sim \mathcal{IG}(n_0 72, s_0/2)$, where $\mathcal{IG}(a,b)$ denotes an inverse Gamma distribution with parameters a and b. The conditional distribution of y_t given z_t is normal with mean $\boldsymbol{x}_t'\boldsymbol{\beta}_p + \theta\upsilon_t$ and variance $\tau^2\upsilon_t$. The joint density of $\boldsymbol{y} = (y_1, \ldots, y_T)'$ is given by:

$$f(\boldsymbol{y}|\boldsymbol{\beta}_p, \boldsymbol{z}, \sigma) \propto \left(\prod_{t=1}^{T} \upsilon_t^{-1/2}\right) exp\left\{ -\sum_{t=1}^{T} \frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_p - \theta\upsilon_t)^2}{2\tau^2\upsilon_t}\right\}, \tag{12}$$

We need to sample $\boldsymbol{\beta}_p$, $\boldsymbol{v} = (\upsilon_1, \ldots, \upsilon_t)'$ and σ from their conditionals distributions: The full conditional density of $\boldsymbol{\beta}_p$ is given by:

$$\boldsymbol{\beta}_p|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{v}, \sigma \sim \mathcal{N}(\bar{\beta}_p, \bar{V}_\beta), \tag{13}$$

where:

$$\bar{V}_\beta^{-1} = \left(\sum_{t=1}^{T} \frac{x_t' x_t}{\tau^2\sigma\upsilon_t} + \boldsymbol{B}_{p_0}^{-1}\right) \quad \bar{\beta}_p = \bar{V}_\beta\left[\sum_{t=1}^{T} \frac{x_t(y_t - \theta\upsilon_t)}{\tau^2\sigma\upsilon_t} + \boldsymbol{B}_{p0}^{-1}\boldsymbol{\beta}_{p_0}\right] \tag{14}$$

and assuming a normal prior

$$\boldsymbol{\beta}_p \sim \mathcal{N}(\boldsymbol{\beta}_{p_0}, \boldsymbol{B}_{p_0}) \tag{15}$$

where $\boldsymbol{\beta}_{p_0}$ and $\boldsymbol{B}_{p_0}$ are the prior mean and variance covariance matrix of $\boldsymbol{\beta}_p$. Priors for the quantile betas have been chosen to have mean zero and variance 1000.

The full conditional distribution of υ_t is proportional to:

$$\upsilon_t|\boldsymbol{y}_t, \boldsymbol{x}_t\boldsymbol{\beta}_p\sigma \sim GIG(1/2, \delta_t, \gamma_t) \tag{16}$$

where:

$$\delta_t = (y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2/\tau^2\sigma \quad \gamma_t^2 = 2/\sigma + \theta^2/\tau^2\sigma \tag{17}$$

and where GIG denotes the Generalized Inverse Gaussian distribution which pdf for the general case $GIG(v, a, b)$ is:

$$f(x|v, a, b) = \frac{(b/a)^v}{2K_v(ab)} x^{v-1} exp\left\{ -1/2(a^2x^{-1} + b^2x)\right\}, \quad x > 0, \quad -\infty < v < \infty, \quad a, b \leq 0 \tag{18}$$

and K_v is a modified Bessel function of the third kind. By noting that $\upsilon_t \sim \mathcal{E}(\sigma)$, the full conditional density of σ is proportional to:

$$\sigma|\boldsymbol{y}_t, \boldsymbol{x}_t\boldsymbol{\beta}_p\boldsymbol{\upsilon} \sim IG(n/2, s/2) \tag{19}$$

where $n = n_0 + 3n$ and $s = s_0 + 2\sum_{t=1}^{T} \upsilon_t + (y_t + \boldsymbol{x}_t'\boldsymbol{\beta}_p + \theta\upsilon_t)^{\prime 2}/\tau^2\upsilon_t$

The posterior distribution is calculated using 8000 replications after a burn-in period of 4000 replications. For each replication, a quantile predictive function is calculated for the following:

$$p(y_{t+h}|\boldsymbol{y}_t, \boldsymbol{\vartheta}) = \int p(y_{t+h}|\boldsymbol{\vartheta}, \boldsymbol{y}_t)p(\boldsymbol{\vartheta}|\boldsymbol{y}_t)d\boldsymbol{\vartheta} \tag{20}$$

where $p(\boldsymbol{\vartheta}|\boldsymbol{y}_t)$ corresponds to the posterior distributions of the parameters' set $\boldsymbol{\vartheta} = \{\boldsymbol{\beta}_p, \theta, \tau, \upsilon_t\}$, and $p(\boldsymbol{y}_{t+h}|\boldsymbol{\vartheta})$ corresponds to:

$$p(\boldsymbol{y}_{p,t+h}|\boldsymbol{\vartheta}) = \int \mathcal{N}(y_{t+1}|x_T'\boldsymbol{\beta}_p + \theta z_T\upsilon_T, \tau^2 z_T\upsilon_T)d\boldsymbol{\beta}_p d\upsilon_t \tag{21}$$

2.2 Quantile Combination of Density Forecasts

Combining density forecasts has recently become a well-known practice for macroeconomic forecasting, see Aastveit et al. (2019) for a recent survey of the literature. The most common approach for combining predictive densities is to use weighted linear combinations of prediction models, evaluated using a type of scoring rule (e.g. Hall and Mitchell, 2007; Amisano and Giacomini, 2007; Jore et al., 2010; Hoogerheide et al., 2010; Kascha and Ravazzolo, 2010; Geweke and Amisano, 2011, 2012; Gneiting and Ranjan, 2013; Aastveit et al., 2014; Kapetanios et al., 2015b). However, recent advances also include more complex combination approaches that allows for time-varying weights with possibly both learning and model set incompleteness (e.g. Billio et al., 2013; Casarin et al., 2015; Pettenuzzo and Ravazzolo, 2016; Del Negro et al., 2016; Aastveit et al., 2018; McAlinn and West, 2019b; McAlinn et al., 2020).

Common to all the aforementioned approaches is that a single weight is attached to the entire predictive distribution for each model, assuming the predictive ability is constant across the various regions of the distribution. Suppose that a set of $k = 1, \ldots, K$ predictive distributions $f_{t+h,k}$ for the same variable of interest y_t at horizon h are available. Standard combination methods apply a unique combination weight to the entire predictive distribution, i.e.:

$$\underset{1\times Q}{y_{t+h}} = \underset{1\times K}{w_k}\, \underset{K\times Q}{f_{t+h,k,q}} \tag{22}$$

where $q = 1, \ldots, Q$ indicates the quantiles or bins of the density distribution. However, this procedure implicitly overlooks superior forecast accuracy of some $f_{t+h,k}$ over a specific region of the distribution. In this case, it would be ideal to use a more flexible approach to combination. Suppose indeed that a subset of this set is more accurate in predicting the mean (tails) of the distribution, while they perform poorly in the tails (mean); It would be desirable then to consider this heterogeneity when constructing the combined density i.e.:

$$\underset{1\times Q}{y_{t+h}} = diag\left(\underset{Q,K}{w_{q,k}} \underset{K\times Q}{f_{t+h,k,q}}\right) \tag{23}$$

Since here weights are quantile-specific, density forecasts are now multiplied by a vector of combination weights instead of a scalar as in equation (22).

2.2.1 Evaluation of quantiles' forecast accuracy: the quantile scores

From equation (4) we obtain K (set to 6 in our empirical application) density forecasts for y_{t+h}, distributed over Q (set to 5 in our empirical application) quantiles. The purpose of this is paper is to combine them taking into consideration the forecast accuracy at a quantile level. In order to do so, we need an evaluation method that helps us discriminating not only the accuracy of the k^{th} forecast but also its accuracy at specific quantiles.

A common scoring rule for evaluating density forecasts is the Continuous Ranked Probability Score (CRPS). According to a loss function, the density forecast is evaluated by computing the distance at each point of the distribution to the realisation. It is defined by:

$$CRPS(f_{t+h,k}, y_{t+h}) = -\int_{-\infty}^{\infty} (F_{t+h,k} - \mathbb{I}(F_{t+h,k} \geq y_{t+h}))^2 dy \tag{24}$$

where $F_{t+h,k}$ represents the cdf of forecast $f_{t+h,k}$ and y_{t+h} the corresponding realization. The CRPS corresponds to the integral of the Brier scores for the probability forecasts at all real-valued thresholds (Matheson and Winkler (1976), Hersbach (2000), Gneiting and Raftery (2007)). While this score is the average "error" across the domain of the distribution, Gneiting and Ranjan (2011) propose a quantile decomposition of the CRPS (24). It is represented by:

$$CRPS_{t+h,k} = \int_0^1 QS_{t+h,k}(q)dq \tag{25}$$

where $QS_{t+h,k}(q)$ is called Quantile Score:

$$\begin{aligned} QS_{t+h,k}(q) &= \frac{1}{n-h+1} \sum_{t=m}^{m+n-h} QS_q(F^{-1}_{t+h,k}(q), y_{t+h}) \\ QS_q(F^{-1}_{t+h,k}(q), y_{t+h}) &= 2\left(\mathbb{I}\{y_{t+h} \leq F^{-1}_{t+h,k}(q)\} - q\right)(F^{-1}_{t+h,k}(q) - y_{t+h}) \end{aligned} \tag{26}$$

where n and m are defined by the in-sample and out-of-sample periods, $F^{-1}_{t+h,k}(q)$ is the value the inverse of the cdf of $f_{t+h,k}$ taken at quantile q. We suggest to use this decomposition, $(QS_q(F^{-1}_{t+h}(q), y_{t+h}))$, to evaluate the accuracy of predictive distributions at each quantile.

We would like to highlight a couple of properties of function QS_q in equation (26). First, notice that the closer q is to zero, the lower are the probabilities that $F^{-1}_{t+h}(q)$ will have a value lower than y_{t+h}; at the same time, the closer q is to one, the lower are the probabilities that $F^{-1}_{t+h}(q)$ will have a value greater than y_{t+h} (Laio and Tamea (2007). The quantile score based on equation (26) therefore has a concave shape. Second, since QS is a measure of loss accuracy, the density forecast k that obtains the lowest QS curve is preferred to the other alternatives. The CRPS-quantile decomposition QS is a proper scoring rule as proven by Friederichs and Hense (2008).

2.2.2 Quantile-specific combination weights

We propose to use the quantile scores as loss function to build quantile-specific combination weights:

$$\underset{K\times Q}{w_{t+h}}(k,q) = \frac{\sum_{t=m}^{m+n-h} 1/QS_{t,k,q}}{\sum_{k=1}^{K}\sum_{t=m}^{m+n-h} 1/QS_{t,k,q}} \tag{27}$$

where $k = 1, \ldots, K =$ denotes the individual forecasts, q the quantiles and $t = m, \ldots, m+n-h$ the forecast origins. $w(t+h,k,q)$ is the matrix $K \times Q$ of combination weights for forecast ϕ^*. The recursive weights are then a function of past performance of each model k known at the time the forecast is made (t). We need to impose the following constrain on the combination weight $w(k,q)$:

$$\sum_{k=1}^{K} w(t,k,q) = 1$$

The combined density forecast y^c_{t+h} is obtained by multiplying the matrix of combination weights computed according to (27) with the matrix of quantile forecasts:

$$y^c_{t+h} = \text{diag}(w_{t+h,k,q} \times f_{t+h,q,k}) \tag{28}$$

The diagonal of this matrix corresponds to the match between the vector of weights and the corresponding model k.

2.2.3 Forecast evaluation

We measure forecasts accuracy using the CRPS. In addition, we also compare the forecasting performance of the various individual models and alternative combination approaches with versions of the score that penalizes the loss in accuracy at certain regions (centre, left tail, right tail and so on) of the target distribution. Gneiting and Ranjan (2011) uses the quantile scores decomposition to have weighted versions of the continuous ranked probability score (6) that emphasize regions of interest and retain propriety.

$$\text{emphCRPS}_{t+h,k} \int_0^1 QS_q(F^{-1}_{t+h,k}(q), y_{t+h})\nu(q)dq \tag{29}$$

where ν is a nonnegative weight function on the unit interval. For a constant weight function, equation (29) reduces to the unweighted score (25).

2.3 Alternative combination approaches

In the empirical application, we will compare our quantile combination approach with three alternative combination approaches.

2.3.1 Equal Weights

The first combination approach we considered is to apply a linear pooling scheme to the K predictive distributions obtained from the kernel smoothing over the quantiles and summing them applying a combination weight $w_k = 1/K$ to each predictive. The combined predictive distribution is the following:

$$f(y_{t+h}) = \sum_{k=1}^{K} w_k f_{t+h,k} \tag{30}$$

Combination weights $\omega_n = 1/N$ assure that the combined distribution is still a distribution since $0 \leq \omega_n \leq 1$ and $\sum_{n=1}^{N} \omega_n = 1$.

2.3.2 Optimal Weights

The second combination approach is the so called "Optimal Weights" proposed by Hall and Mitchell (2007) and Geweke and Amisano (2011), held from the idea of determining combination weights based on some objective criterion or cost function, such as the logarithmic score. Combination weights are obtained by maximizing a logarithmic score function:

$$w_k = \frac{1}{T-h} \sum_{t=1}^{T-h} ln(f_{t+1,k}) \quad s.t. \quad w_k > 0, \quad \sum_{k=1}^{K} w_k = 1 \tag{31}$$

which is known as the log predictive score. Given the size of K, the inference algorithm for w_k in Conflitti et al. (2015) is used.

2.3.3 Log score weights

The third combination approach, proposed by Jore et al. (2010), is to choose recursive combination weights for each model based on their logartitmic score. The weights are then constructed as follow:

$$w_{t+h}(k) = \frac{\sum_{t=m}^{m+n-h} LS_{t,k}}{\sum_{k=1}^{K} \sum_{t=m}^{m+n-h} LS_{t,k}} \tag{32}$$

2.3.4 CRPS weights

An alternative to the linear pooling scheme with log-score weights, is to use CRPS instead as the scoring rule. The weights will then take the following form:

$$w_{t+h}(k) = \frac{\sum_{t=m}^{m+n-h} 1/CRPS_{t,k}}{\sum_{k=1}^{K} \sum_{t=m}^{m+n-h} 1/CRPS_{t,k}} \tag{33}$$

2.3.5 Bayesian Model Averaging

For Bayesian Model Averaging (BMA, henceforth) the individual predictive densities are combined into a composite-weighted predictive distribution $p(y_{t+h}|I_K)$, given by

$$p(y_{t+h}|I_K) = \sum_{k=1}^{K} P\left(M_k\right) p(\widetilde{y}_{t+h}|k) \tag{34}$$

where $P\left(M_k\right)$ is the posterior probability of model k, based on the predictive likelihood for model k. Mitchell and Hall (2005) discuss the analogy of the log score in a frequentistic framework to the log predictive likelihood in a Bayesian framework, and how it relates to the Kullback-Leibler divergence. See also Hall and Mitchell (2007), Jore et al. (2010), and Geweke and Amisano (2010) for a discussion on the use of the log score as a ranking device for the forecast ability of different models and Hoeting et al. (1999) for a review on BMA.

3 Empirical Application

In this section, we analyse the performance of our quantile combination approach for forecasting US real GDP growth using real time data. The main goal of the exercise is to examine the nowcasting performance of our quantile combination approach, to compare its performance to commonly used alternative combination approaches and to analyse what are the most informative predictors for the various parts of the predictive GDP growth distribution.

3.1 Data

We consider in total $K = 5$ different predictors. These are leading indicators that cover a broad range of the macroeconomy and that earlier studies have found to be useful for predicting GDP growth and recessions. A vast amount of research has shown that a variety of economic and financial variables contain predictive information about future economic recessions and downturns. While Estrella and Hardouvelis (1991) and Estrella and Mishkin (1998) have documented that the slope of the term structure has strong predictive power for US recessions, a recent study by Adrian et al. (2019) argue that financial conditions are particularly informative above future downside macroeconomic risk. In addition to these studies, several other variables have also been regarded as leading recession indicators for GDP growth and recessions, including stock prices (Estrella and Mishkin (1998) and Stock and W Watson (2003)), the index of leading economic indicators (Berge and Jordà (2011) and Stock and Watson (1989)), oil prices (Hamilton (1983, 1996) and Ravazzolo and Rothman (2013, 2016)), survey data (Hansson et al. (2005), Claveria et al. (2007)); and residential investments (Aastveit et al. (2019)).

We include the following 5 variables: the National Financial Condition Index (NFCI), the University of Michigan Consumer Sentiment Index (ICS), a credit spread that measures the difference between BAA corporate bond yield and the 10 year treasury yield, residential investments, building permits, the unemployment rate. Detailed information about the various series, including data source and data transformation, is provided in the online appendix, Table (A.1). Our data sample covers the period 1973Q1-2020Q2.

3.2 In-sample quantile coefficient estimates

To highlight the advantage of using quantile regressions in our setting, Figures A.1 - A.5 in the online appendix show the change in $\boldsymbol{\beta}_q$ estimates across quantiles from various specifications of equation (4). Each model specification consists of lagged GDP growth and one additional predictor (with lags). The figures provide estimation results for the full sample, 1973Q-2020Q2. For each figure, the $Q = 5$ $\boldsymbol{\beta}_q$ estimates (one for each quantile) are plotted against the OLS estimate of β (dashed blue line). The red solid line shows the median estimates of $\boldsymbol{\beta}_q$ and the shaded areas show the 68 and 84 percent probability bands, respectively. There are two interesting observations from the figures.

First, for many model specifications, such as those including NFCI and ICS, the $\boldsymbol{\beta}_q$ estimates across quantiles are very different. For instance in the model with NFCI, the $\boldsymbol{\beta}_q$ estimates for the lower quantiles are strongly negative, but are positive for the upper quantiles. This indicates that the informativeness of various predictors varies for the different parts of the predictive GDP growth distribution, favouring the use of quantile regressions.

Second, while for some of the model specifications, such as those credit spreads the probability bands of the various $\boldsymbol{\beta}_q$ estimates includes the OLS estimate, for others, such as those including NFCI and ICS, the probability bands of the $\boldsymbol{\beta}_q$ estimates for some of the quantiles do not include the OLS estimate. This indicates potential benefits from using quantile regressions.

3.3 Out-of-sample density forecasts for US GDP growth

Our out-of-sample forecasting evaluation period runs from 1993Q1-2020Q2. Predictions are updated recursively for forecast horizons $H = 1$ (one quarter ahead) and $H = 4$ (one year ahead) based on models that are estimated using an expanding window.

Table 1: Average CRPS values with emphasis on specific regions of the distribution, one-quarter-ahead forecasts.

Emphasis	$\nu(q)$	EQ	OPT	BMA	Log Score	CRPS	Q-comb
Uniform	$\nu_0(q) = 1$	0.604***	0.923	0.903	0.952	0.988	**0.336**
Centre	$\nu_1(q) = q(1-q)$	0.602***	0.919	0.906	0.957	**1**	**0.068**
Tails	$\nu_2(q) = (2q-1)^2$	0.604***	0.927	0.888	0.941	0.955	**0.064**
Right Tail	$\nu_3(q) = q^2$	0.489***	0.834	0.873	0.881	0.923	**0.096**
Left Tail	$\nu_4(q) = (1-q)^2$	0.776	1.029	0.937	1.019	**1.04**	0.104
Heavy Tails	$\nu_5(q) = (2q-1)^4$	0.617	0.913	0.913	0.954	0.954	**0.021**

Note: The table reports average CRPS values, with emphasis on specific regions of the distribution, for various forecast combination approaches. The alternative combination models EQ, OPT, BMA, Log Score, CRPS combines linear models, while Q-comb combines quantile models. For the alternative models, we report the relative performance compared to Q-comb. Thus, values > 1 denotes higher forecast accuracy than our quantile combination. Stars indicate significance levels for Diebold-Mariano Test of Q-Comb versus alternative approaches combinations.

Table 2: Average CRPS values with emphasis on specific regions of the distribution, one-year ahead forecasts.

Emphasis	$\nu(q)$	EQ	OPT	BMA	Log Score	CRPS	Q-comb
Uniform	$\nu_0(q) = 1$	0.561***	0.793***	1.000	0.996	0.846	**0.319**
Centre	$\nu_1(q) = q(1-q)$	0.579***	0.795***	1.000	1.000	0.833	**0.066**
Tails	$\nu_2(q) = (2q-1)^2$	0.504***	0.778***	**1.018**	0.833	0.857	0.056
Right Tail	$\nu_3(q) = q^2$	0.452***	0.833***	**1.021**	0.989	0.842	0.095
Left Tail	$\nu_4(q) = (1-q)^2$	0.707***	0.748***	0.978	0.989	0.858	**0.092**
Heavy Tails	$\nu_5(q) = (2q-1)^4$	0.500***	0.783***	1.200	1.000	0.833	0.018

Note: The table reports average CRPS values, with emphasis on specific regions of the distribution, for various forecast combination approaches. The alternative combination models EQ, OPT, BMA, Log Score, CRPS combines linear models, while Q-comb combines quantile models. For the alternative models, we report the relative performance compared to Q-comb. Thus, values > 1 denotes higher forecast accuracy than our quantile combination. Stars indicate significance levels for Diebold-Mariano Test of Q-Comb versus alternative approaches combinations.

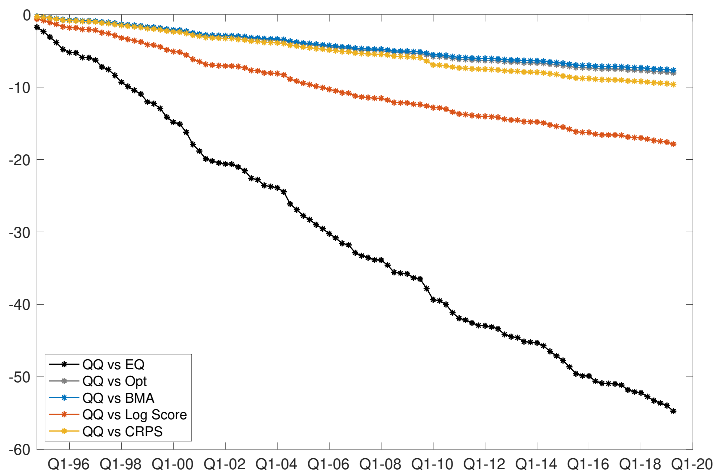
We compare the forecasting performance from our quantile combination approach with commonly used alternative combination approaches. Tables 1) and (2 compare density forecasting accuracy from our quantile combination approach (labelled Q-comb) with equal weighting combination of linear forecasts (EQ), optimal weighting combination of linear forecasts (OPT), Bayesian model averaging (BMA), combination of linear forecasts using CRPS weights (CRPS) and combinations of linear forecasts using log score weights (Log Score). We report density forecasting accuracy, measured by standard CRPS and various CRPS versions that penalizes the loss accuracy of at certain regions (centre, left tail, right tail and so on) of the target distribution. The tables report averages over the evaluation periods.

The tables reveal that our quantile combination approach outperforms the other alternative combination approaches for both forecasting horizons ($H = 1$ and $H = 4$). This holds irrespective of using the CRPS or any quantile weighted version of the CRPS, that emphasize performance in either the centre, left or right tail of the distribution, as forecast accuracy measure. This indicates that the relative gains in terms of forecasting performance from our combination approach is not specific to observations in a certain region of the distribution. Stars in the tables represent the significance level of Diebold-Mariano Test for superior forecast ability of quantile combination versus the alternative linear combination approaches.

To address whether our improved out-of-sample forecast performance is limited to a certain time period or driven by some outliers, we report in Figures 1) and (2 the cumulative CRPS of the alternative combination approaches relative to our quantile combination. While the various individual models show considerable instabilities in predictive performance over time, the performance from our quantile combination approach, is far more robust, yielding a steady improvement over the various alternative combination approaches over the different time periods.

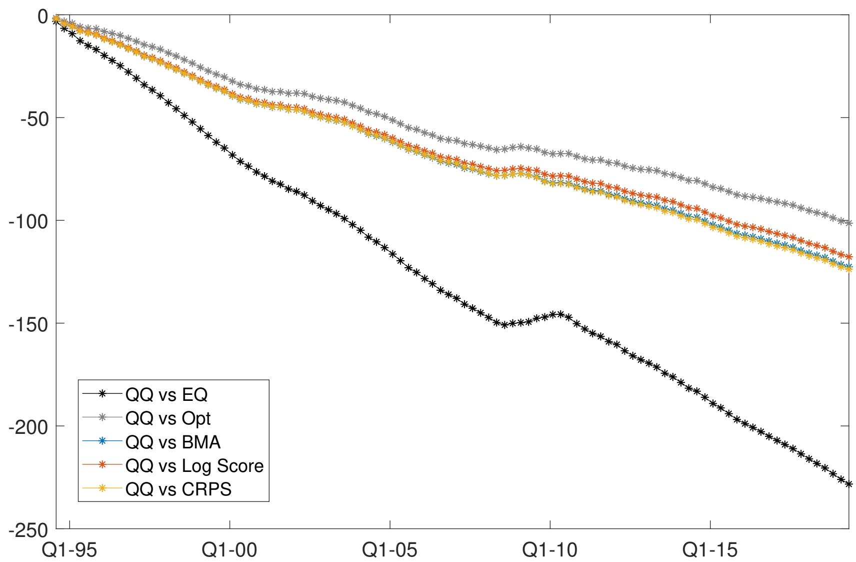
Tables 3 and 4 report the forecasting performance for all the individual quantile regression models. First, the tables show that our quantile combination performs well compared to the individual models. In fact, for forecast horizon $H = 4$ forecast from our quantile combination approach outperforms all of the individual models, both for the tails and the centre of the distribution. Second, the table reveal that the quantile regression model that includes credit spread as a predictor is the best performing model for forecasting horizon $H = 1$, both in terms of average performance as well as in the tails. On the other hand, for forecasting horizon $H = 4$ the model that includes the residential investment as a predictor is the best performing model both in terms of average performances as well as for the tails. While earlier studies, such as Aastveit et al. (2019) and Estrella and Hardouvelis (1991) and Estrella and Mishkin (1998), have found residential investments and the credit spread, respectively, to be informative about future economic recessions, a recent study by Adrian et al. (2019) argue that financial conditions are particularly informative about future downside macroeconomic risk. Although the model that includes financial conditions also is among the best performing models, particularly for the left

Figure 1: Cumulative CRPS for alternative approaches to forecast and combination for one-quarter ahead.



Note: Combination weights used are: equal weight (EQ), optimal weight (Opt), BMA, weights proportional to Log Scores and CRPS. QQ denotes the quantile combination of quantile forecasts.

Figure 2: Cumulative CRPS for alternative approaches to forecast and combination for one-year ahead.



Note: Combination weights used are: equal weight (EQ), optimal weight (Opt), BMA, weights proportional to Log Scores and CRPS. QQ denotes the quantile combination of quantile forecasts.

tail, our results suggests that also other variables than the NFCI are informative about future downside macroeconomic risk.

Table 3: Average CRPS values with emphasis on specific regions of the distribution, one-quarter ahead forecasts.

Emphasis $\nu(q)$	Uniform $\nu_0 = 1$	Centre $\nu_1 = q(1-q)$	Tails $\nu_2 = (2q-1)^2$	Right Tail $\nu_3 = q^2$	Left Tail $\nu_4 = (1-q)^2$	Heavy Tails $\nu_5 = (2q-1)^4$
Q comb	0.336	0.068	0.064	0.096	0.104	0.021
GDP	0.342	0.07	0.063	0.098	0.104	0.02
NFCI	0.361	0.073	0.067	0.099	0.115	0.022
ICS	0.334	0.068	0.063	0.095	0.103	0.02
U	0.367	0.075	0.069	0.104	0.114	0.022
CR Spread	**0.317**	**0.065**	**0.059**	**0.093**	**0.095**	**0.019**
ResInv	0.343	0.07	0.064	0.103	0.1	0.021

Note: The table reports average CRPS values, with emphasis on specific regions of the distribution, for quantile combination and all individual models.

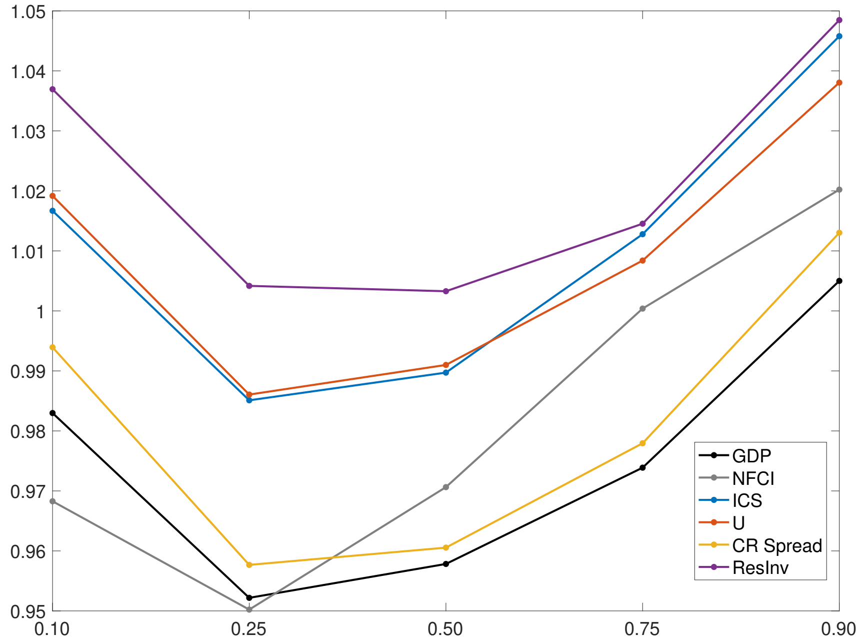
Table 4: Average CRPS values with emphasis on specific regions of the distribution. four-quarter ahead forecasts.

Emphasis $\nu(q)$	Uniform $\nu_0 = 1$	Centre $\nu_1 = q(1-q)$	Tails $\nu_2 = (2q-1)^2$	Right Tail $\nu_3 = q^2$	Left Tail $\nu_4 = (1-q)^2$	Heavy Tails $\nu_5 = (2q-1)^4$
Q comb	**0.319**	**0.066**	**0.056**	**0.095**	**0.092**	**0.018**
GDP	0.393	0.079	0.076	0.117	0.118	0.025
NFCI	0.35	0.071	0.066	0.095	0.113	0.021
ICS	0.344	0.07	0.065	0.096	0.108	0.021
U	0.393	0.08	0.074	0.112	0.121	0.024
CR Spread	0.352	0.072	0.066	0.099	0.109	0.021
ResInv	0.329	0.067	0.062	0.097	0.099	0.02

Note: The table reports average CRPS values, with emphasis on specific regions of the distribution, for quantile combination and all individual models.

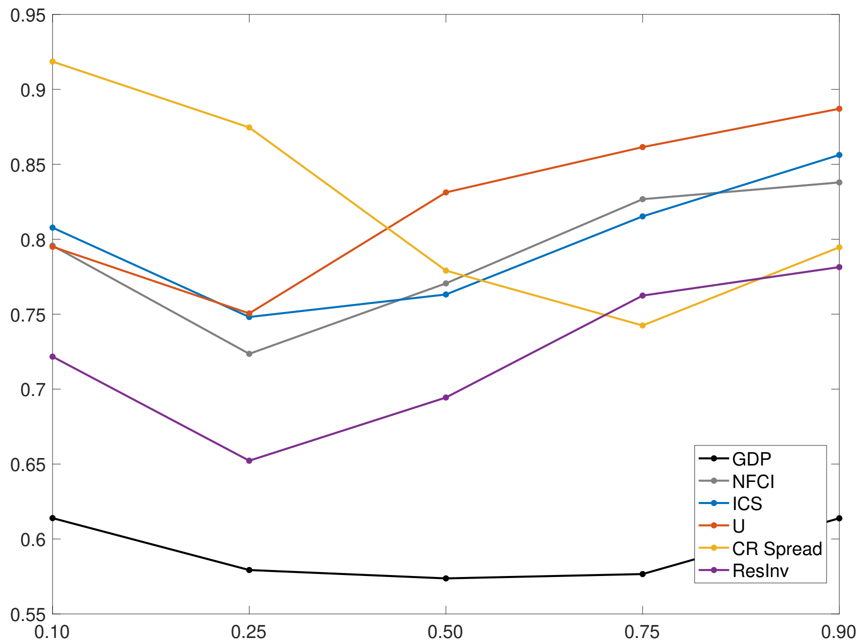
Figures (3-4) report the average quantile score for each of the individual quantile regression models. The figures reveal a substantial heterogeneity in relative score performance, where for instance some of the models have a relative low score for the lower quantiles, but a relative high score for the higher quantiles. This indicates that the relative model performance varies over the forecast distribution.

Figure 3: Average quantile Scores for all $K = 6$ forecasting models for one-quarter ahead over forecasts. Average score over the last 1000 MC replications.



Note: QS is a measure of loss in accuracy, it has to be preferred the density forecast model k that exhibits a lower QS curve than the alternatives.

Figure 4: Average quantile Scores for all $K = 6$ forecasting models for one-year ahead over forecasts. Average score over the last 1000 MC replications.



Note: QS is a measure of loss in accuracy, it has to be preferred the density forecast model k that exhibits a lower QS curve than the alternatives.

4 Simulation Experiments with Data

In this section, a set of simulation exercises have been run to explore the validity of our results and the characteristics of quantile combination in controlled environments. First, we assessed the impact of sample size on forecast accuracy for quantile combination. As discussed at the end of previous section, a lower sample size corresponds to a small number of tail events which undermines the accuracy of our forecast evaluation. For this reason, we draw 1,000 realisations from data-driven distributions and compare the combination approaches (Section 4.1). Second, we use the larger sample size generated in (4.1) to relax the constrain of $Q = 5$ we assumed in application. In the second exercise in Section (4.2), we estimate $Q = 10$ quantiles in the quantile regression. Third, in Section (4.3), we add to the previous setup further complication of increasing the number of individual models to be combined.

In all those simulation, the accuracy gained from using quantile combination forecast increases further compared to other traditional combinations.

4.1 Simulation Setup: higher sample size

The first simulation aims to increase simple size in our application exercise by simulating realisations. We consider two foresting models: an AR model and an ARDL model with NFCI and past values of GDP growth as predictor. The sample size is increased to $T = 1000$. The realisations are drawn from the empirical distributions of GDP and NFCI using the dataset presented in Section . In particular simulated realisations are drawn from:

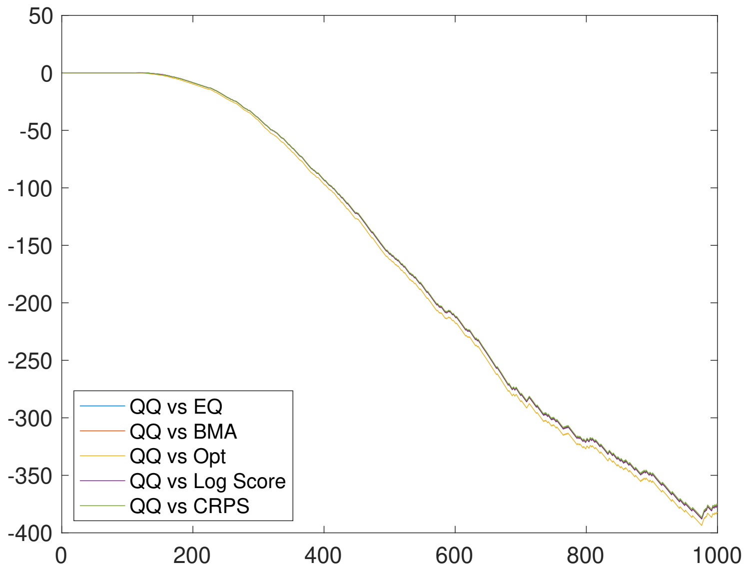
$$\begin{aligned} y_{hs} &\sim p(y_t | y_{hs-r}, NFCI_{hs-v}, \boldsymbol{\theta}) \\ NFCI_{hs} &\sim p(NFCI_h s | NFCI_{hs-p}, \boldsymbol{\theta}) \end{aligned} \tag{35}$$

where $y_{t=hs}$ identifies GDP growth simulated for $t = hs$, y_{hs-r} the GDP growth lagged by r and $NFCI_{hs-v}$ the values lagged by v of simulated data; $\boldsymbol{\theta}$ are a vector of parameters estimated from quantile regression using "real data" for $hs = 1, \ldots, T$ and simulated from for the remaining part of the sample (i.e. $hs = T + 1, \ldots, 1000$). Predictive distributions using simulated data for GDP and NFCI are then estimated and combined as in the previous application. From Table (5) we can see that, on average, quantile combination is more accurate compared to linear combinations when the sample size is larger. This result is quite encouraging: under the scenario that the researcher has access to a large dataset, the quantile combination delivers a much more accurate density forecast. However, quite rarely macroeconomic time-series have such a big size. The plot of cumulative CRPS scores in Figure (5), shows the marginal gain in accuracy with the increase in sample size. From its inspection we can infer that the accuracy increases from a sample size as big as 200 observations; after which the quantile combination produces a exponentially more accurate density forecast than the linear combinations. This result is consistent with our empirical exercise, where the sample size is around 190 observations, and the gain of using quantile combination is quite marginal. However, this results is quite promising since it displays that the quantile combination method performs increasingly better even at a realistic sample size.

Table 5: Simulation (1) with T=1000 forecast origins. CRPS scores for Quantile combination (Q-comb) and relative performance of alternative models compared to Q-comb.

	Uniform	Centre	Tails	Right Tail	Left Tail	Heavy Tails
Q comb	0.584	0.118	0.11	0.185	0.162	0.036
EQ	0.996	0.204	0.182	0.235	0.354	0.058
BMA	0.996	0.204	0.182	0.235	0.354	0.058
Opt	0.994	0.203	0.182	0.235	0.353	0.058
LS	0.996	0.204	0.182	0.235	0.354	0.058
CRPS	0.996	0.204	0.182	0.235	0.354	0.058

Figure 5: Simulation (1) with T=1000 forecast origins: Cumulative Scores for Linear and Quantile Combinations



Note: Combination weights used are: equal weight (EQ), optimal weight (Opt), BMA, Log Score and CRPS. QQ denotes the quantile combination of quantile forecasts.

4.2 Simulation Setup: Higher number of quantiles

The higher sample size allows as to increase the number of quantiles to $Q = 10$. The number of individual models to be combined is $K = 2$ as in the previous exercise. In the empirical application we had to impose a quite small number of quantiles to estimate, not enough (or none) observations for GDP growth was falling into the specific bins of quantile forecasts, making impossible to evaluate the forecast accuracy at specific quantiles (i.e. computing the quantile score). However, with simulated data in the previous exercise, we can relax this assumption and estimate Q=10 quantiles. Once again, we can see from Table (6) that, on average, quantile combination is more accurate compared to linear combinations. However, the quantile combination needs a larger sample size than before to be more accurate than the linear combinations, i.e. at more than 350.

Table 6: Simulation (2) with T=1000 forecast origins and Q=10 quantiles. CRPS scores for Quantile combination (Q-comb) and relative performance of alternative models compared to Q-comb.

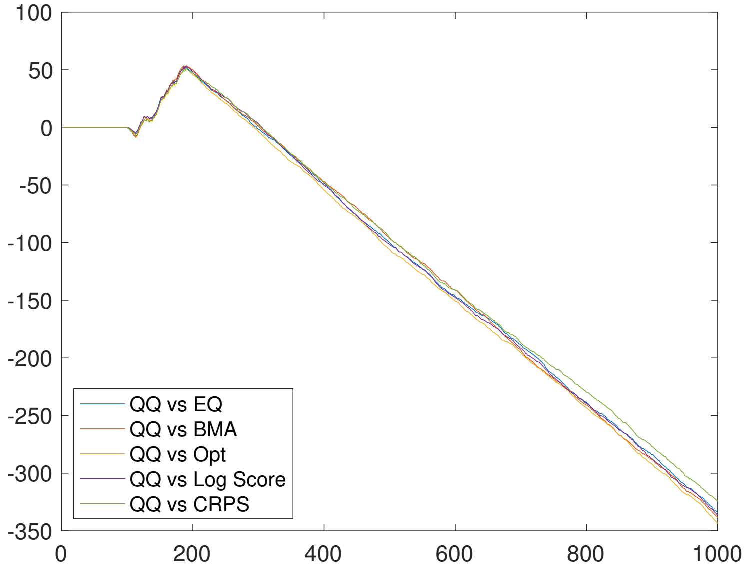
	Uniform	Centre	Tails	Right Tail	Left Tail	Heavy Tails
Q comb	0.563	0.108	0.132	0.183	0.164	0.058
EQ	0.972	0.191	0.206	0.24	0.348	0.088
BMA	0.971	0.191	0.206	0.24	0.348	0.088
Opt	0.987	0.195	0.207	0.244	0.353	0.088
LS	0.971	0.191	0.206	0.24	0.348	0.088
CRPS	0.973	0.192	0.206	0.241	0.348	0.088

4.3 Simulation Setup: Higher number of individual models to be combined

In this exercise we increase the number of combined models, from $K = 5$ to $K = 10$. The additional 5 models are obtaining using as predictors the following variables:

- CFNAI: Chicago Fed National Activity Index (source:Chicago Fed), period: 1967:M3-2020M3
- S&P500: Stock Market Index (source: FRED), period 1959Q1-2020Q1
- OIL: Spot Crude Oil Price WTI (source: FRED), period 1946:Q12020Q1
- PERMSA: New Private Housing Units Permits (source: FRED), period 1960:M1-2020M3
- BANKCRg:Total Credit to Private non-fin sector (source: FRED), period 1970Q1-2020M3

Figure 6: Simulation (2) with T=1000 forecast origins and Q=10 quantiles: Cumulative Scores for Linear and Quantile Combinations



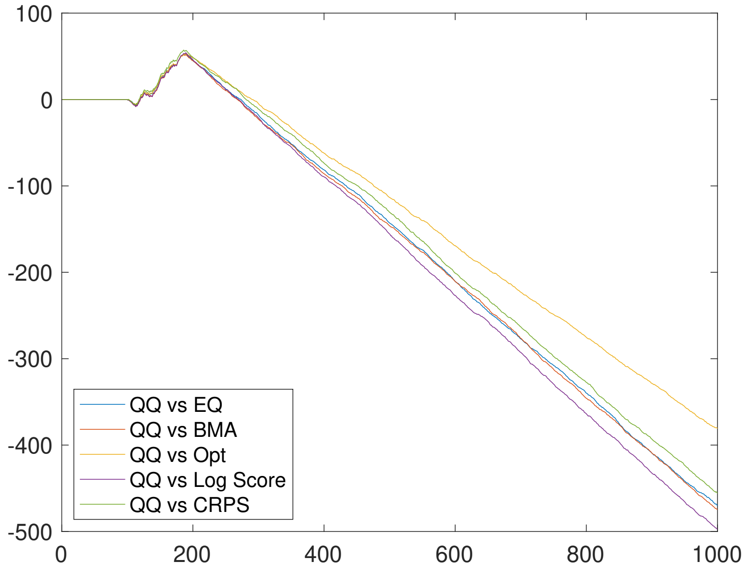
Note: Combination weights used are: equal weight (EQ), optimal weight (Opt), BMA, Log Score and CRPS. QQ denotes the quantile combination of quantile forecasts.

Data are simulated from these variables in the same way as in the first simulation set up. With this set up we want to explore the impact of higher number of components to combine on our combination approach. It is well known that some combination approaches loose accuracy when combining a high number of components. Here we see similar results as in the previous simulation: quantile combination is the most accurate approach when the sample size is big enough to compensate for the estimation error.

Table 7: Simulation (3): T=1000 forecast origins, Q=10 quantiles and K=10 models. CRPS scores for Quantile combination (Q-comb) and relative performance of alternative models compared to Q-comb.

	Uniform	Centre	Tails	Right Tail	Left Tail	Heavy Tails
Q comb	0.532	0.102	0.124	0.173	0.155	0.055
EQ	1.059	0.212	0.21	0.263	0.371	0.088
BMA	1.059	0.212	0.21	0.263	0.371	0.088
Opt	0.946	0.187	0.196	0.233	0.337	0.083
LS	1.086	0.218	0.214	0.271	0.379	0.089
CRPS	1.027	0.205	0.206	0.255	0.362	0.086

Figure 7: Simulation (3) with T=1000 forecast origins Q=10 quantiles and K=10 models: Cumulative Scores for Linear and Quantile Combinations



Note: Combination weights used are: equal weight (EQ), optimal weight (Opt), BMA, Log Score and CRPS. QQ denotes the quantile combination of quantile forecasts.

5 Conclusions

In this paper, we propose a new forecast combination approach, which aims to obtain overall more accurate density forecasts, by assigning a set of combination weights to the various quantities of the individual density forecasts. To achieve this goal, we first produce individual density forecasts using Bayesian quantile regression models as in Kozumi and Kobayashi (2011). Then we combine the various individual forecasts using a novel quantile combination approach, where each quantile of the combined density forecast is constructed as a weighted combination of the individual forecasts for the corresponding quantile. To account for the heterogeneity in forecast accuracy from the various models across the various parts of the distribution, we allocate the quantile-specific weights from each model using the quantile score by Gneiting and Ranjan (2011).

In an empirical application, we demonstrate the usefulness of our novel quantile combination approach to forecasting real GDP growth rate for the Unites States for the period 1993Q1-2020Q2, combining predictive distributions from $K = 5$ quantile regression models.

As a main result, we show that density forecasts from our quantile combination approach outperforms forecasts from commonly used combination approaches such as Bayesian Model Averaging (BMA), the optimal combination of density forecasts (Opt-

Comb) suggested by Hall and Mitchell (2007) and Geweke and Amisano (2011), recursive logarithmic score weights as in Jore et al. (2010) and equal weights. This holds irrespective of using the CRPS or any threshold or quantile weighted version of the CRPS, that emphasizes performance in either the centre, left or right tail of the distribution, as a measure of forecast accuracy. The latter indicates that the relative gain in terms of forecasting performance from our model is not specific to observations in a certain region of the distribution or to specific subperiods in our forecasting sample. Instead, we find a steady improvements over time and in all quantiles of the GDP distribution.

Furthermore, while Adrian et al. (2019) argue that financial conditions are particularly informative about future downside macroeconomic risk, we show that quantile regressions that include variables, such as residential investments and stock prices, provide somewhat more accurate forecasts for the lower left quantile of the GDP distribution than quantile regressions that include the NFCI. This suggests that also other variables than the NFCI are informative about future downside macroeconomic risk.

References

Aastveit, K. A., A. K. Anundsen, and E. I. Herstad (2019). Residential investment and recession predictability. *International Journal of Forecasting 35*(4), 1790–1799.

Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud (2014). Nowcasting GDP in real time: A density combination approach. *Journal of Business & Economic Statistics 32*, 48–68.

Aastveit, K. A., J. Mitchell, F. Ravazzolo, and H. van Dijk (2019). The Evolution of Forecast Density Combinations in Economics. In *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Aastveit, K. A., F. Ravazzolo, and H. K. van Dijk (2018). Combined density nowcasting in an uncertain economic environment. *Journal of Business & Economic Statistics 36*, 131–145.

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–89.

Amburgey, A. and M. W. McCracken (2022, January). On the Real-Time Predictive Content of Financial Conditions Indices for Growth. Working Papers 2022-003, Federal Reserve Bank of St. Louis.

Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics 25*(2), 177–190.

Berge, T. J. and Ò. Jordà (2011). Evaluating the classification of economic activity into recessions and expansions. *American Economic Journal: Macroeconomics 3*(2), 246–77.

Billio, M., R. Casarin, F. Ravazzolo, and H. K. van Dijk (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics 177*, 213–232.

Brownlees, C. and A. B. Souza (2021). Backtesting global Growth-at-Risk. *Journal of Monetary Economics 118*(C), 312–330.

Carriero, A., T. E. Clark, and M. Massimiliano (2020, May). Nowcasting Tail Risks to Economic Activity with Many Indicators. Working Papers 20-13R2, Federal Reserve Bank of Cleveland.

Casarin, R., S. Grassi, F. Ravazzolo, and H. K. van Dijk (2015). Parallel sequential Monte Carlo for efficient density combination: The DeCo MATLAB toolbox. *Journal of Statistical Software, Articles 68*, 1–30.

Clark, T. E., F. Huber, G. Koop, M. Marcellino, and M. Pfarrhofer (2021, March). Tail Forecasting with Multivariate Bayesian Additive Regression Trees. Working Papers 21-08R, Federal Reserve Bank of Cleveland.

Claveria, O., E. Pons, and R. Ramos (2007). Business and consumer expectations and macroeconomic forecasts. *International Journal of Forecasting 23*(1), 47–69.

Coe, P. J. and S. P. Vahey (2020, April). Financial conditions and the risks to economic growth in the United States since 1875. CAMA Working Papers 2020-36, Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University.

Conflitti, C., C. De Mol, and D. Giannone (2015). Optimal combination of survey forecasts. *International Journal of Forecasting 31*(4), 1096–1103.

Del Negro, M., R. B. Hasegawa, and F. Schorfheide (2016). Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics 192*(2), 391–405.

Diks, C., V. Panchenko, and D. van Dijk (2011). Likelihood-based scoring rules for comparing density forecasts in tails. *Journal of Econometrics 163*(2), 215–230.

Estrella, A. and G. A. Hardouvelis (1991). The term structure as a predictor of real economic activity. *The journal of Finance 46*(2), 555–576.

Estrella, A. and F. S. Mishkin (1998). Predicting us recessions: Financial variables as leading indicators. *Review of Economics and Statistics 80*(1), 45–61.

Friederichs, P. and A. Hense (2008). A probabilistic forecast approach for daily precipitation totals. *Weather and Forecasting 23*(4), 659–673.

Ganics, G., B. Rossi, and T. Sekhposyan (2020, January). From Fixed-event to Fixed-horizon Density Forecasts: Obtaining Measures of Multi-horizon Uncertainty from Survey Density Forecasts. CEPR Discussion Papers 14267, C.E.P.R. Discussion Papers.

Ganics, G. A. (2017, December). Optimal density forecast combinations. Working Papers 1751, Banco de España.

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

Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics 164*(1), 130–141.

Geweke, J. and G. G. Amisano (2012). Prediction with misspecified models. *The American Economic Review 102*, 482–486.

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

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

Gneiting, T. and R. Ranjan (2013). Combining predictive distributions. *Electronic Journal of Statistics 7*, 1747–1782.

Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting 23*(1), 1–13.

Hamilton, J. D. (1983). Oil and the macroeconomy since world war ii. *Journal of political economy 91*(2), 228–248.

Hamilton, J. D. (1996). This is what happened to the oil price-macroeconomy relationship. *Journal of Monetary Economics 38*(2), 215–220.

Hansson, J., P. Jansson, and M. Löf (2005). Business survey data: Do they help in forecasting gdp growth? *International Journal of Forecasting 21*(2), 377–389.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting 15*(5), 559–570.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382–401.

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

Jore, A. S., J. Mitchell, and S. P. Vahey (2010). Combining forecast densities from vars with uncertain instabilities. *Journal of Applied Econometrics 25*(4), 621–634.

Kapetanios, G., J. Mitchell, S. Price, and N. Fawcett (2015a). Generalised density forecast combinations. *Journal of Econometrics 188*(1), 150–165.

Kapetanios, G., J. Mitchell, S. Price, and N. Fawcett (2015b). Generalised density forecast combinations. *Journal of Econometrics 188*(1), 150–165.

Kascha, C. and F. Ravazzolo (2010). Combining inflation density forecasts. *Journal of forecasting 29*(1-2), 231–250.

Korobilis, D. (2017). Quantile regression forecasts of inflation under model uncertainty. *International Journal of Forecasting 33*(1), 11–20.

Kotz, S., T. Kozubowski, and K. Podgorski (2012). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media.

Kozumi, H. and G. Kobayashi (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation 81*(11), 1565–1578.

Laio, F. and S. Tamea (2007). Verification tools for probabilistic forecasts of continuous hydrological variables. *Hydrology and Earth System Sciences 11*(4), 1267–1277.

Liu, W. and E. Moench (2016). What predicts us recessions? *International Journal of Forecasting 32*(4), 1138–1150.

Marcellino, M. (2006). Leading indicators. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 879–960. Amsterdam: Elsevier.

Matheson, J. E. and R. L. Winkler (1976). Scoring rules for continuous probability distributions. *Management science 22*(10), 1087–1096.

Mazzi, G. L. and J. Mitchell (2019). Nowcasting Euro Area GDP Growth Using Quantile Regression. *mimeo*.

McAlinn, K., K. A. Aastveit, J. Nakajima, and M. West (2020). Multivariate Bayesian Predictive Synthesis in Macroeconomic Forecasting. *Journal of the American Statistical Association*. Forthcoming.

McAlinn, K. and M. West (2019a). Dynamic bayesian predictive synthesis for time series forecasting. *Journal of Econometrics 210*(1), 155–169.

McAlinn, K. and M. West (2019b). Dynamic bayesian predictive synthesis in time series forecasting. *Journal of econometrics 210*(1), 155–169.

Mitchell, J. and S. G. Hall (2005). Evaluating, comparing and combining density forecasts using the klic with an application to the bank of england and niesr 'fan' charts of inflation. *Oxford Bulletin of Economics and Statistics 67*(s1), 995–1033.

Opschoor, A., D. Van Dijk, and M. van der Wel (2017). Combining density forecasts using focused scoring rules. *Journal of Applied Econometrics 32*(7), 1298–1313.

Pettenuzzo, D. and F. Ravazzolo (2016). Optimal portfolio choice under decision-based model combinations. *Journal of Applied Econometrics 31*, 1312–1332.

Ravazzolo, F. and P. Rothman (2013). Oil and us gdp: A real-time out-of-sample examination. *Journal of Money, Credit and Banking 45*(2-3), 449–463.

Ravazzolo, F. and P. Rothman (2016). Oil-price density forecasts of us gdp. *Studies in Nonlinear Dynamics & Econometrics 20*(4), 441–453.

Reichlin, L., G. Ricco, and T. Hasenzagl (2020). Financial variables as predictors of real growth vulnerability. Technical report.

Stock, J. H. and M. W Watson (2003). Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature 41*(3), 788–829.

Stock, J. H. and M. W. Watson (1989). New indexes of coincident and leading economic indicators. *NBER macroeconomics annual 4*, 351–394.

Appendix A Online Appendix

A.1 Description of data series

Table A.1: Description of Data Series

Label	Trans	Period	Real-Time	Description	Source
rgdp	Δln	59:Q1-20:Q1	73Q1-20Q1	Real GDP growth, sa	AL
NFCI	level	71:Q1-20:Q1	11Q2-20Q1	National Financial Conditions Index	Chicago Fed
ICS	level-100	60:Q1-20:Q1	98Q3-20Q1	Consumer Sentiment Index	AL
CreSpread	Level	53:Q1-20:Q1	none	Credit Spread: BAA corporate bond yield - 10-year treasury	F
U	Δ log	48:Q1-20:Q1	65Q4-20Q1	Unemployment rate	AL
ResInv	Δ%	47:Q2-20:Q1	65Q4-20Q1	Real Gross Private Domestic Investment: Fixed Investment: Residential	AL

Notes: Sources abbreviated as "F" denotes Federal Reserve Economic Data (FRED), as "AL" denotes Federal Reserve Economic Real-Data (ALFRED) dataset.

A.2 In-sample quantile regression coefficient

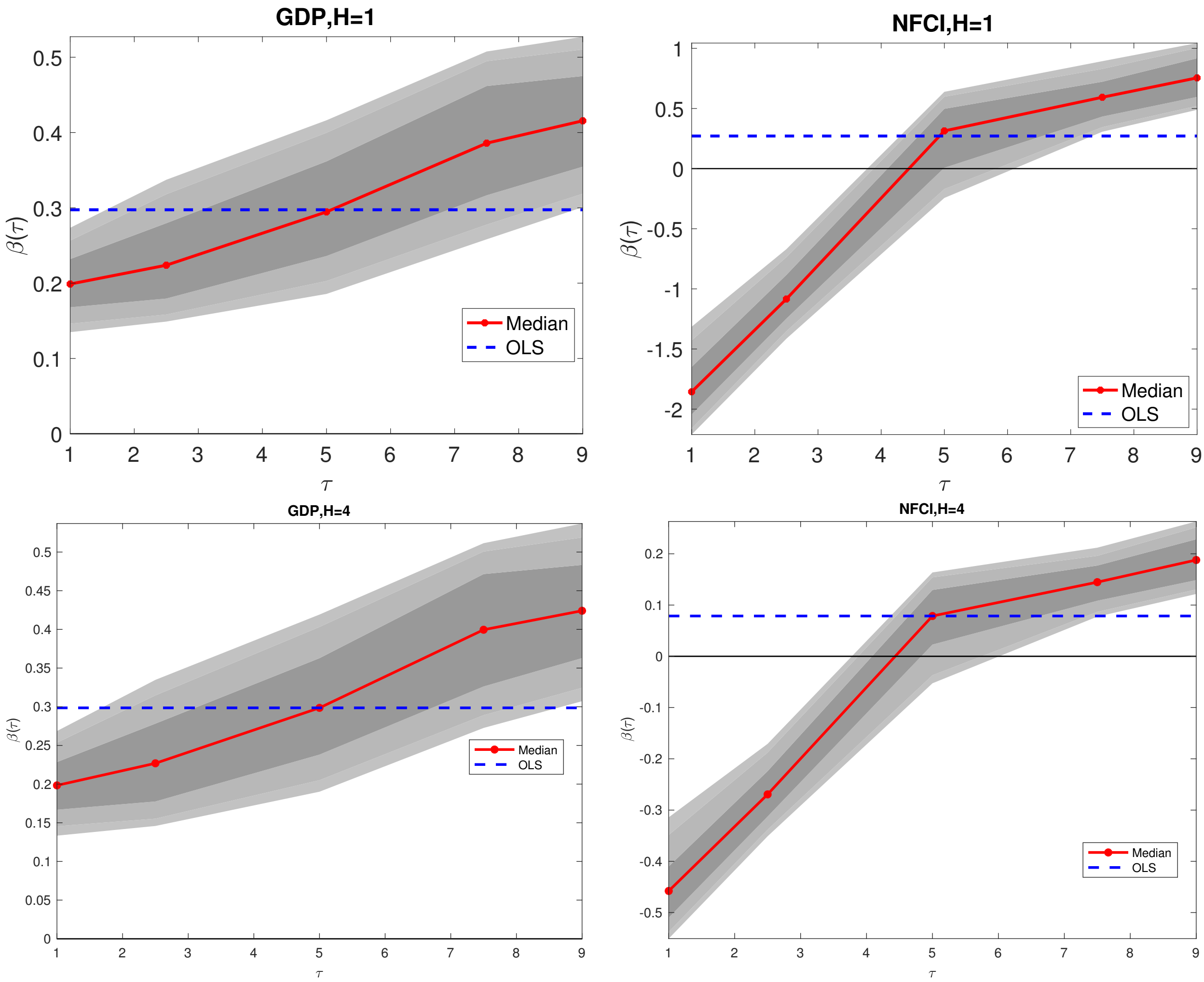


Figure A.1: Coefficients estimates from quantile ARDL model using GDP growth and NFCI as regressors. $H = 1$ denotes one-quarter ahead horizon, $H = 4$ one-year ahead.

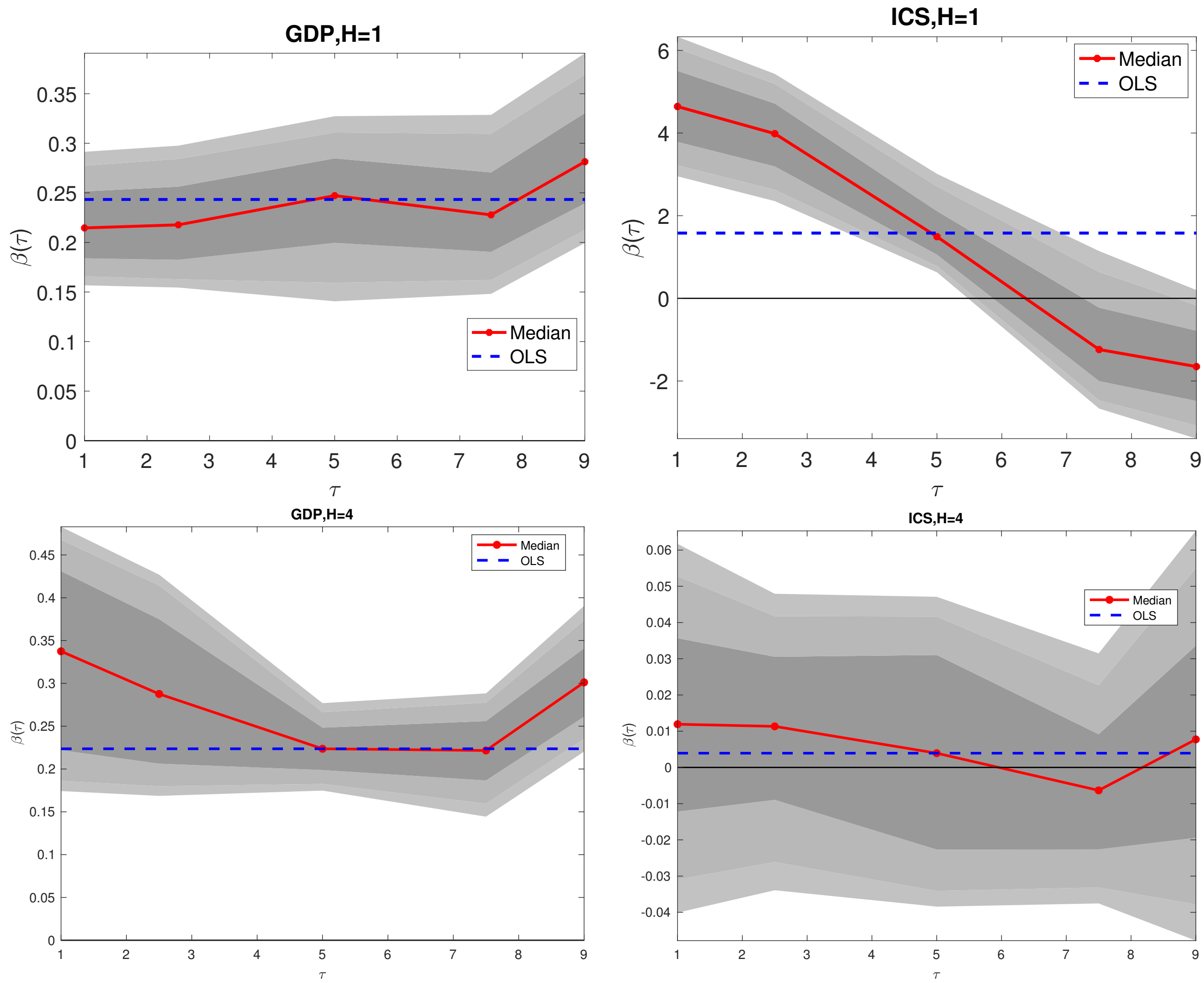


Figure A.2: Coefficients estimates from quantile ARDL model using GDP growth and ICS as regressors. $H = 1$ denotes one-quarter ahead horizon, $H = 4$ one-year ahead.

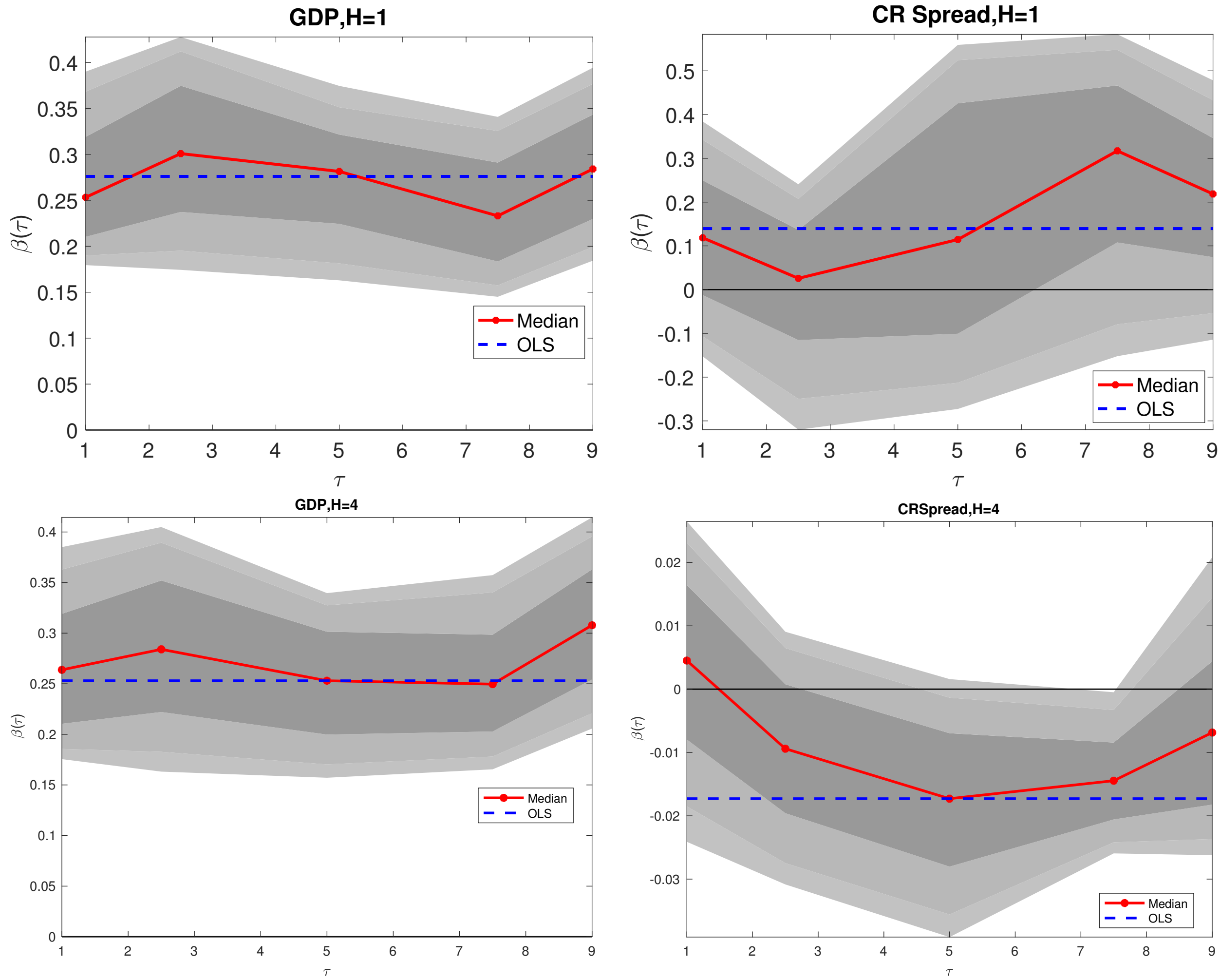


Figure A.3: Coefficients estimates from quantile ARDL model using GDP growth and Credit Spread as regressors. $H = 1$ denotes one-quarter ahead horizon, $H = 4$ one-year ahead.

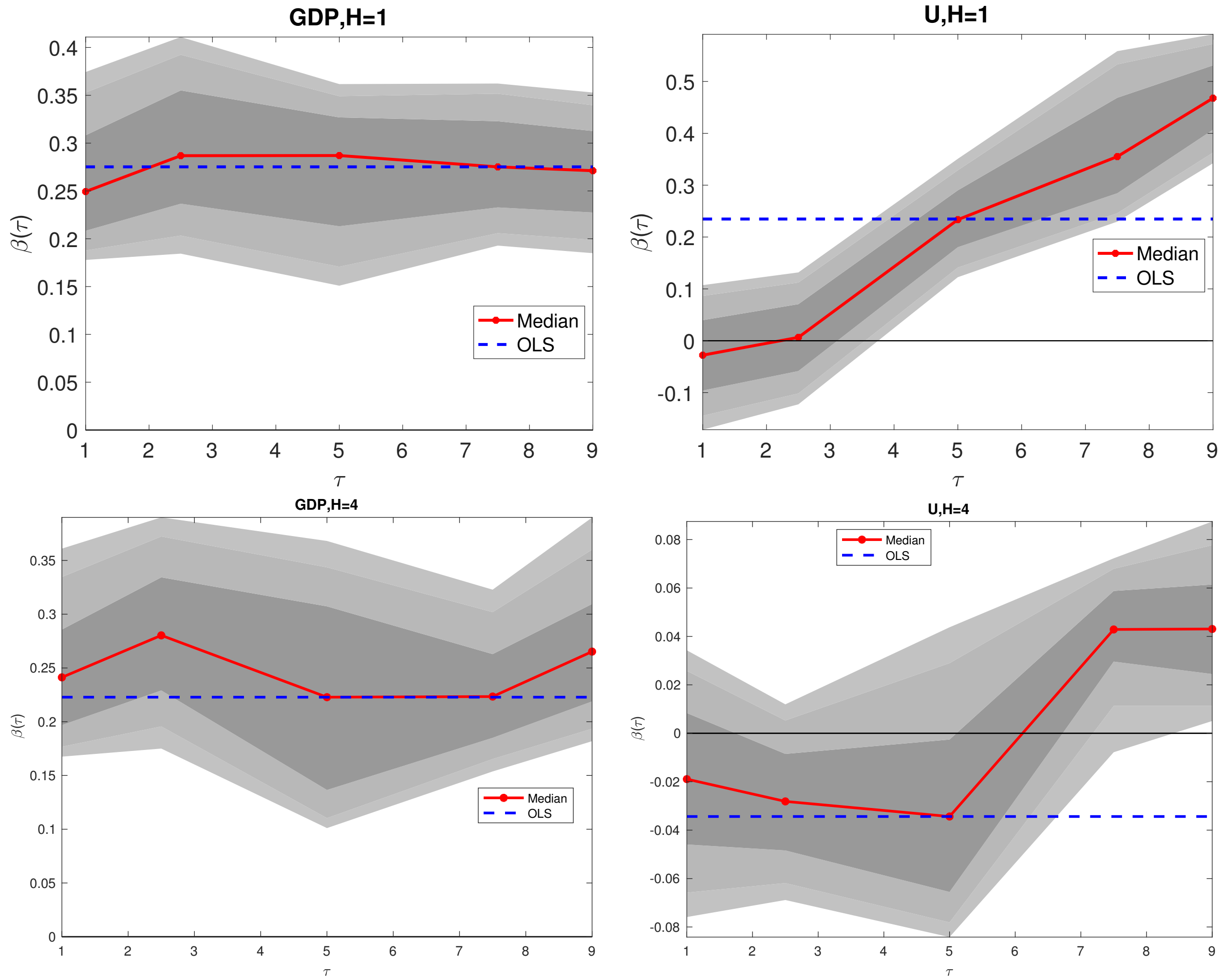


Figure A.4: Coefficients estimates from quantile ARDL model using GDP growth and Unemployment rate as regressors. $H = 1$ denotes one-quarter ahead horizon, $H = 4$ one-year ahead.

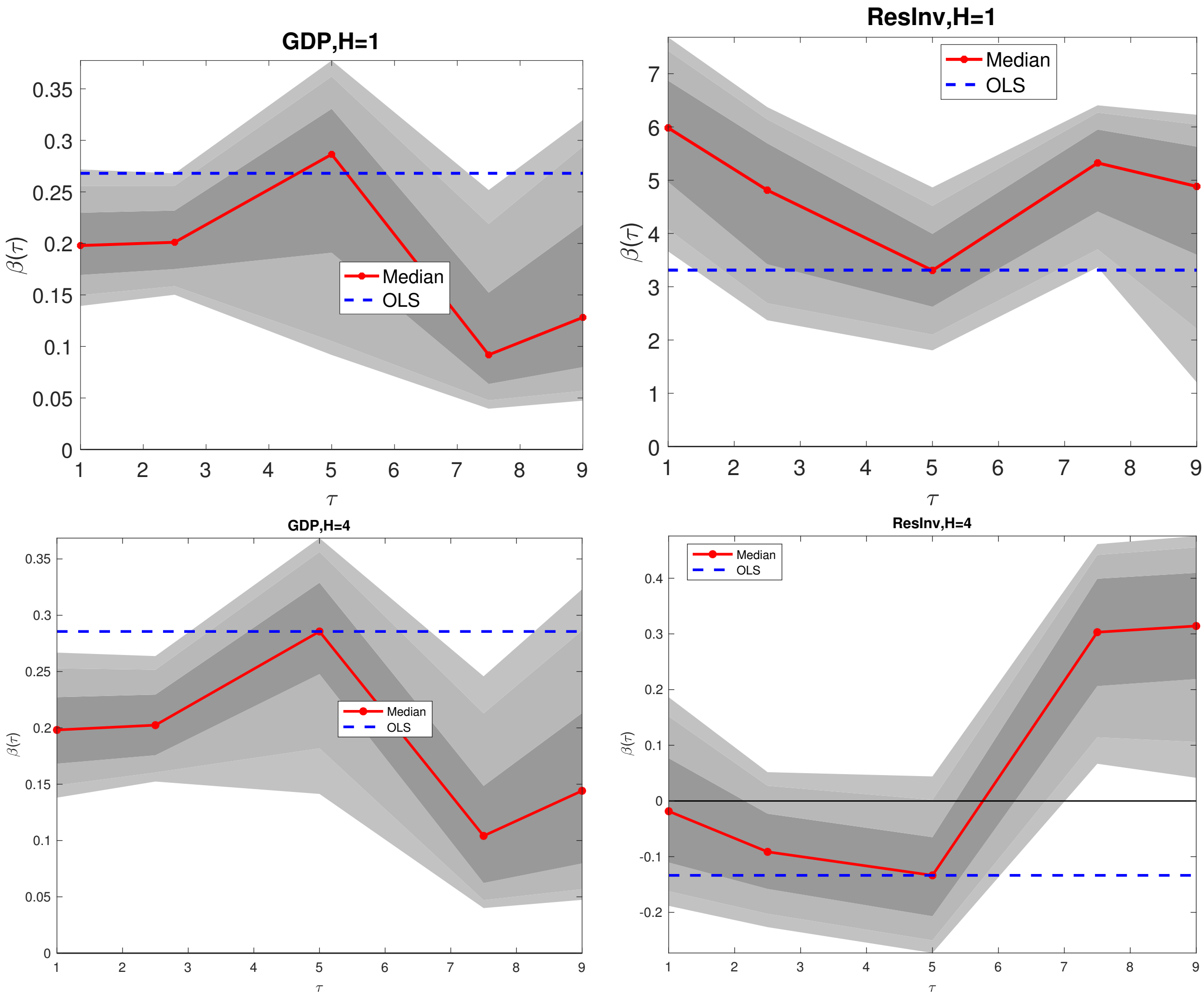


Figure A.5: Coefficients estimates from quantile ARDL model using GDP growth and Resident Investment growth rate as regressors. $H = 1$ denotes one-quarter ahead horizon, $H = 4$ one-year ahead.