Federal Reserve Bank of Cleveland Working Paper Series

Specification Choices in Quantile Regression for Empirical Macroeconomics

Andrea Carriero, Todd E. Clark, and Massimiliano Marcellino

Working Paper No. 22-25

August 2022

Suggested citation: Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino. 2022. "Specification Choices in Quantile Regression for Empirical Macroeconomics." Working Paper No. 22-25. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202225.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

Specification Choices in Quantile Regression for Empirical Macroeconomics*

Andrea Carriero,[1] Todd E. Clark,[2] and
Massimiliano Marcellino[3]

[1] *Queen Mary University of London and University of Bologna,* [2] *Federal Reserve Bank of Cleveland,*
[3] *Bocconi University, CEPR, IGIER, BIDSA, and BAFFI*

August 29, 2022

Abstract

Quantile regression has become widely used in empirical macroeconomics, in particular for estimating and forecasting tail risks to macroeconomic indicators. In this paper we examine various choices in the specification of quantile regressions for macro applications, for example, choices related to how and to what extent to include shrinkage, and whether to apply shrinkage in a classical or Bayesian framework. We focus on forecasting accuracy, using for evaluation both quantile scores and quantile-weighted continuous ranked probability scores at a range of quantiles spanning from the left to right tail. We find that shrinkage is generally helpful to tail forecast accuracy, with gains that are particularly large for GDP applications featuring large sets of predictors and unemployment and inflation applications, and with gains that increase with the forecast horizon.

Keywords: Quantile regression, tail forecasting, shrinkage, Bayesian methods, quantile scores
JEL classification codes: C53, E17, E37, F47

*We gratefully acknowledge research assistance from Matthew Gordon and helpful suggestions and comments received from Matthew Gordon, James Mitchell, and Aubrey Poon. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System. Corresponding author: Todd Clark, todd.clark@researchfed.org.

1 Introduction

Quantile regression has become widely used in empirical macroeconomics, in particular for estimating and forecasting tail risks to macroeconomic indicators. Several examples in a voluminous literature include Giglio, Kelly, and Pruitt (2016) and Adrian, Boyarchenko, and Giannone (2019) for output growth; Kiley (2022) for unemployment; Gaglianone and Lima (2012), Korobilis (2017), Lopez-Salido and Loria (2022), and Manzan and Zerom (2013) for inflation; and Manzan (2015) for a range of variables.

Several of the papers cited above use classical, unconstrained quantile estimation. Yet, in point and density forecasting in macroeconomics, practitioners and researchers now widely recognize the potential value of employing some form of shrinkage in the estimation of the parameters of forecasting models. Bayesian estimation is one common way to incorporate shrinkage, particularly with vector autoregressions, and also for structural analysis. Frequentist approaches to incorporating shrinkage commonly rely on loss functions with particular penalty terms, as in ridge, lasso, and elastic net formulations. Other tools such as forecast averaging and data reduction through common factors can also be seen as shrinkage methods.

In the tail risk and quantile regression setting, comparable options for incorporating shrinkage into estimation exist. These include Bayesian quantile regression as developed in sources such as Khare and Hobert (2012) and Yu and Moyeed (2001) and used in macroeconomic forecasting by studies including Korobilis (2017) and Mitchell, Poon, and Mazzi (2022). Studies have also considered averaging or combining quantile forecasts (e.g., Giacomini and Komunjer (2005) and Korobilis (2017)) and factor reduction (e.g., Carriero, Clark, and Marcellino (2022)) through the partial quantile regression approach developed in Giglio, Kelly, and Pruitt (2016).

The availability of multiple options raises the question of whether any particular approach offers consistent advantages such that it should be preferred. In point and density forecasting in macroeconomics, many studies have conducted examinations of alternative estimation approaches, covering Bayesian versus frequentist approaches, and the choice or estimation of prior settings in Bayesian VARs (e.g., Carriero, Clark, and Marcellino (2015), Chan (2021), and Giannone, Lenza, and Primiceri (2015)). In the quantile regression setting, no such examination of alternative specification choices has been conducted. Yet such shrinkage could be particularly helpful, given the effectively small sample sizes for estimation of tail risk models. In the quantile literature, it is well known that the estimation of extreme quantiles with small samples of observations can result in coefficient bias (see, e.g., Chernozhukov, Fernandez-Val, and Kaji (2017)). According to a rule of thumb summarized in Chernozhukov, Fernandez-Val, and Kaji (2017), extreme value methods — as opposed to standard quantile regression methods — should be used when $\tau T/k \leq 15$ to 20, where τ denotes the quantile, T is the sample size, and k is the number of regressors. In an application like that of Adrian, Boyarchenko, and Giannone (2019), with about 40 years of quarterly data, a quantile of 0.05, and 3 regressors, $\tau T/k \approx 2.7$. This suggests some challenges with extreme quantile estimation and forecasting in quarterly macroeconomic time series, as well as some likely benefit to shrinkage methods for forecasting.

Accordingly, in this paper we examine various choices in the specification of quantile regressions used for modeling and forecasting macroeconomic tail risks. Our analysis is based on applications to US data on GDP growth, the unemployment rate, and inflation, patterned on empirical work by Adrian, Boyarchenko, and Giannone (2019), Caldara, et al. (2021), and Plagborg-Moller, et al. (2020) for output growth; Kiley (2022) for unemployment; and Korobilis, et al. (2021) and Lopez-Salido and Loria (2022) for inflation. For each application, we compare the accuracy of quantile forecasts obtained from simple quantile regression, averages of quantile regression forecasts obtained with one indicator at a time, partial quantile regression, quantile regression with ridge penalty, and a few different specifications of priors for Bayesian quantile regression. These prior specifications include a fixed Minnesota-type prior (e.g., Carriero, Clark, and Marcellino (2022)) and a horseshoe prior (e.g., Kohns and Szendrei (2021) and Mitchell, Poon, and Mazzi (2022)). In comparing accuracy, we consider a range of quantiles — spanning from the left to right tail — and use quantile scores (e.g., Giacomini and Komunjer (2005)), as well as quantile-weighted continuous ranked probability scores developed in Gneiting and Ranjan (2011) that consider the entire predictive density.

Some related work has considered choices that bear on the accuracy of complete predictive densities obtained with quantile regression as an input. Specifically, some empirical work on tail risk has considered measures — such as expected shortfall — that require a complete predictive density and not just quantile estimates. In this empirical work, following the precedents of studies such as Korobilis (2017) and Adrian, Boyarchenko, and Giannone (2019), a two-step approach is common: In step 1, estimate selected quantiles using quantile regression; in step 2, use some smoothing approach to fit a complete predictive density to the quantiles. One approach follows Korobilis (2017) in empirically smoothing a wide range of quantiles. Another approach follows Adrian, Boyarchenko, and Giannone (2019) in fitting a parametric skewed-t density by matching population expressions of the moments of the density to several empirical quantiles. Mitchell, Poon, and Zhu (2022) and Mitchell, Poon, and Mazzi (2022) compare alternative specifications of this two-step approach to fitting a complete density, focusing on the choice of the number of quantiles and the choice of the second-step fitting or smoothing approach. Mitchell, Poon, and Mazzi (2022) recommend the empirical smoothing of Korobilis (2017) with 19 quantiles for accurate density estimation; Mitchell, Poon, and Zhu (2022) recommend an improved non-parametric smoothing approach that is capable of capturing multi-modalities in predictive distributions. As mentioned, in our evaluation we focus on the quantile forecasts themselves rather than on the resulting predictive density, as specific quantiles (e.g., in the left or right tails) are often of interest in empirical macro applications. As noted above, we evaluate these quantile forecasts using both quantile scores and quantile-weighted continuous ranked probability scores that use a range of quantiles to cover the entire predictive distribution but weight selected regions more than others.

To anticipate some of our main results, we find that shrinkage is generally helpful to tail forecast accuracy, yielding consistent gains relative to standard quantile regression. The gains are particularly large for GDP applications featuring large sets of predictors and unemployment and inflation applications, and the gains increase with the forecast horizon. This suggests that newer analyses of risks to GDP growth, inflation, and unemployment would be better based on partial quantile regression or Bayesian quantile regression rather than the classical quantile regression that is most commonly (although not exclusively) used in macroeconomic applications. Moreover, based on magnitudes of gains and consistency across applications, horizons, and quantiles, partial quantile regression may be seen as the best-performing method, though at long forecast horizons the partial quantile regression tail quantile estimates can be very volatile, in particular for unemployment but also for inflation. Furthermore, within the Bayesian quantile regression family of approaches, nothing clearly beats simply fixing a Minnesota-type prior at our baseline setting. Finally, while there is some evidence of temporal instability, it does not seem to affect qualitatively the model and method rankings obtained over the full evaluation sample.

The paper proceeds as follows. Section 2 details the models. Section 3 explains the applications and data used. Section 4 provides details on the forecast evaluation. Section 5 presents our empirical results. Section 6 summarizes the main findings and concludes. Additional results are gathered in the supplementary online appendix.

2 Models

This section details the models or shrinkage approaches considered, from quantile regression through Bayesian quantile regression. Throughout, our models take a direct multi-step form, relating the variable of interest h steps ahead to indicators in period t. We consider a few different forecast horizons, detailed in the applications and data section. We estimate the models and evaluate forecasts for a range of quantiles τ: $\tau = 0.05, 0.10, 0.2, 0.3, \ldots, 0.90, 0.95$. For some applications (GDP growth), the lower tail is mostly of interest. For others (e.g., unemployment), the upper tail may be of more interest.

2.1 Quantile regression (QR)

We take as a baseline forecasts obtained from simple (frequentist) quantile regression estimates. For a given variable y to be predicted and quantile τ, we estimate a regression model using a direct multi-step form as in Adrian, Boyarchenko, and Giannone (2019) and many others in the literature:

$$y_{t+h}^{(h)} = x_t'\beta_\tau + \epsilon_{\tau,t+h}, \tag{1}$$

where h is the forecast horizon and the coefficient vector and innovation term are specific to quantile τ. The $k \times 1$ vector of predictors x_t includes a constant and other indicators as described below, depending on the application.

The parameter vector β_τ is estimated with quantile regression:

$$\hat{\beta}_\tau = \underset{\beta_\tau}{\operatorname{argmin}} \sum_{t=1}^{T-h} \left(\tau \cdot \mathbf{1}_{(y_{t+h}^{(h)} \geq x_t'\beta_\tau)} |y_{t+h}^{(h)} - x_t'\beta_\tau| + (1-\tau) \cdot \mathbf{1}_{(y_{t+h}^{(h)} < x_t'\beta_\tau)} |y_{t+h}^{(h)} - x_t'\beta_\tau| \right). \tag{2}$$

We estimate the model for the range of quantiles indicated above.

In results not reported in the interest of brevity, we have also examined forecasts obtained by estimating quantile regressions with the extreme value methods of Chernozhukov, Fernandez-Val, and Kaji (2017). Although their bias adjustment did yield some adjustments of coefficient estimates, we obtained time series of quantile forecasts very similar to the simple QR forecasts. Effects could well be larger for a researcher looking at tails more extreme than those we consider.

2.2 Quantile regression average (QR-avg)

For one frequentist approach to shrinkage of quantile regression, we consider QR forecasts obtained as an (equally weighted) average of estimates from regressions estimated with one predictor at a time. In a different setting, Korobilis (2017) also uses a one-predictor-at-a-time approach, applying Bayesian model averaging to quantile regressions.

More specifically, let $x_{j,t}$ denote a set of regressors including a constant, a lag of the dependent variable, and just one of the other indicators in the full variable set x_t for a given application. For each j, quantile τ, and horizon h, we estimate a quantile regression including as predictors the vector $x_{j,t}$, of the form:

$$y_{t+h}^{(h)} = x'_{j,t}\beta_{j,\tau} + \epsilon_{j,\tau,t+h}. \tag{3}$$

For the regressor set $x_{j,t}$, we form the quantile forecast as the predicted component $x'_{j,t}\hat{\beta}_{j,\tau}$. Our QR-avg forecast for quantile τ and horizon h then takes the form of a simple average of these individual forecasts, from a total of J indicators: $J^{-1}\sum_{j=1}^{J} x'_{j,t}\hat{\beta}_{j,\tau}$.

2.3 Partial quantile regression (PQR)

As another shrinkage approach to frequentist quantile regression, we consider dimension reduction via the use of a common factor in the predictors. Specifically, we consider the partial quantile regression (PQR) method of Giglio, Kelly, and Pruitt (2016) (henceforth, GKP). GKP characterize partial quantile regression as an adaptation of partial least squares to a quantile regression framework. PQR is targeted to quantile regression in that it uses quantile regression in the factor estimation. With respect to using principal components for factor estimation, PQR summarizes the information to make it as useful as possible to predict the quantiles of the specific target variable of interest. In our implementation, we follow GKP in using a single factor specification.

At each forecast origin, consider the vector of variables x_t included as predictors in a given application. For each quantile τ and horizon h, we follow the quantile regression-based approach of GKP to obtain a time series of a scalar factor $f_{\tau,t}$ from the indicators of x_t.[1] Letting $\tilde{x}_t$ denote a vector containing an intercept, lagged dependent variable, and PQR factor estimate formed for the other variables of x_t, we then estimate the quantile regression

$$y_{t+h}^{(h)} = \tilde{x}_t' \beta_\tau + \epsilon_{\tau,t+h}, \tag{4}$$

and form the PQR forecast for quantile τ and horizon h with the resulting coefficient estimates.

2.4 Quantile regression with ridge penalty

With frequentist estimation, other options for imposing shrinkage on estimates of quantile regression follow from adding penalty terms to the tick loss function underlying quantile regression, so as to penalize the addition of more regressors and associated parameters. Following Bayer (2018), we consider quantile regression with a ridge penalty, which pushes coefficients toward zero.[2] For QR with the ridge penalty, the loss function takes the form

$$\begin{aligned} \hat{\beta}_\tau &= \underset{\beta_\tau}{\operatorname{argmin}} \sum_{t=1}^{T-h} \left(\tau \cdot \mathbf{1}_{(y_{t+h}^{(h)} \geq x_t'\beta_\tau)} |y_{t+h}^{(h)} - x_t'\beta_\tau| + (1-\tau) \cdot \mathbf{1}_{(y_{t+h}^{(h)} < x_t'\beta_\tau)} |y_{t+h}^{(h)} - x_t'\beta_\tau| \right) \\ &+ 0.5\lambda \sum_{i=1}^{k} \beta_{\tau,i}^2. \end{aligned} \tag{5}$$

For estimating the model, we rely on the algorithm of Yi and Huang (2017), also used in Bayer (2018); we set the shrinkage parameter λ so as to minimize in pseudo-real-time the quantile score up through the forecast origin.

2.5 Bayesian quantile regression (BQR)

Particularly in settings with a regressor set that can be large, Bayesian shrinkage is helpful in mitigating imprecision in coefficient estimates and associated noise in forecasts. Yu and Moyeed

[1] In the first-stage quantile regression used to obtain the factor, we include a constant and one of the components of x_t, on a one-at-a-time basis (omitting the lagged dependent variable).

[2] Following Lima, Meng, and Godeiro (2020), in additional results omitted in the interest of brevity, we consider quantile regression with an elastic net penalty, originally developed in Zou and Hastie (2005). The elastic net penalty encourages sparsity while working well with strongly correlated indicators. In the recursive scheme results we examined, QR with the elastic net penalty performed comparably to QR with ridge penalty.

(2001) established that quantile regression has a convenient mixture representation that enables Bayesian estimation. For a given variable y to be forecast for quantile τ, our BQR formulation takes the form

$$y_{t+h}^{(h)} = x_t'\beta_\tau + \epsilon_{\tau,t+h}, \tag{6}$$

where $\epsilon_{\tau,t+h}$ has a mixture representation. For each model at quantile τ and horizon h, the representation includes $z_{\tau,t+h}$, which is exponentially distributed with scale parameter $\sigma_{\tau,h}$. The mixture representation of the quantile regression model can be written as

$$y_{t+h}^{(h)} = x_t'\beta_\tau + \theta z_{\tau,t+h} + \kappa\sqrt{\sigma_{\tau,h} z_{\tau,t+h}}\, u_{\tau,t+h}, \tag{7}$$

where θ and κ are fixed parameters as functions of the quantile τ and $u_{\tau,t+h}$ is i.i.d. standard normal.

We estimate Bayesian quantile regressions with the Gibbs sampler of Khare and Hobert (2012) (with modifications as needed to add steps to accommodate some of the coefficient priors indicated below). As a baseline we use an independent Normal-Gamma prior, with a normal distribution for the regression coefficients β_τ and a Gamma distribution for the scale parameter $\sigma_{\tau,h}$. The first step of the Gibbs sampler draws the mixture state time series z from an inverse Gaussian distribution. The second step draws the scale parameter $\sigma_{\tau,w}$ from its inverse Gamma conditional posterior. In the subsequent step, the regression parameter vector $\beta_{\tau,w}$ is drawn from its Normal conditional posterior. For some of the priors described below, additional steps (also described below) are needed to draw parameters entering the prior on the regression parameter vector. For each quantile, forecast results from BQR specifications are based on samples of 5000 retained draws, obtained by sampling a total of 6000 draws and discarding the first 1000. Letting $\hat{\beta}_\tau$ denote the posterior mean of the coefficient vector, the quantile forecast is formed as $x_t'\hat{\beta}_\tau$. We have verified that these settings are adequate for the Markov chain to mix well and converge. For example, across coefficients, forecast horizons, quantiles, and samples, the medians of inefficiency factors fall within a range of about 2 to 5.

Drawing on other approaches in the recent literature, we consider three different specifications of the prior for the regression coefficients. In all cases, the priors' means are 0; the differences are related to the prior variances. In the first approach (labeled BQR-MN below), following studies such as Carriero, Clark, and Marcellino (2022), we fix the prior variance using a Minnesota-style form that takes account of the relative scales of variables and shrinks the coefficients on other variables more than those on the lag of the dependent variable. The shrinkage is controlled by two hyperparameters (smaller numbers mean more shrinkage): λ_1, which controls the overall rate of shrinkage, and λ_2, which controls the rate of shrinkage on variables other than lags of the dependent variable. At each forecast origin, the prior variance for the coefficient on the lagged dependent variable is simply λ_1. The prior variance associated with the coefficient on the i-th variable $x_{i,t}$ of x_t is specified as $\lambda_1 \lambda_2 \frac{\sigma_y^2}{\sigma_i^2}$. Finally, for the intercept, the prior is uninformative, with variance $1000\sigma_y^2$. In setting these components of the prior, for σ_y^2 and σ_i^2 we use simple variances estimated with the regression sample available as of the forecast origin.[3] In this case, we fix the hyperparameters at values that may be considered very common in Minnesota-type priors and forecasting: $\lambda_1 = 0.04$ and $\lambda_2 = 0.25$.

In results omitted in the interest of brevity, we examined the extent to which our BQR-MN results are driven by our particular prior settings (of course, our consideration of other priors described in the remainder of this section provides some additional evidence along these lines). In particular, we have compared results for other settings of these parameters, ranging from tighter to much looser priors, and found that, across variables and quantiles, no alternative is consistently better than the baseline. For example, a tighter prior is better for unemployment and inflation in most cases but makes accuracy worse in the small GDP application. We have also examined results for tighter (lower variance) priors on the intercept. Our findings do not appear to be sensitive to our baseline of a loose prior. Finally, we have conducted simple comparisons of posterior estimates to assess whether the Bayesian shrinkage more typically pushes coefficients toward the prior of 0 in tail quantiles as compared to quantiles more in the middle of the distribution. We did not find any such consistent evidence; in some cases, coefficient estimates are larger (in absolute value) and more precisely estimated for tail quantiles than others.

In the second approach to the BQR prior (labeled BQR-estMN), we adjust the Minnesota prior so that, except for the intercept for which the prior variance is set to be loose, the prior variance is

[3] In the vector autoregression literature, it is typical to use the variances of the residual of low-order AR models. With multi-step forecast horizons in a direct-multi-step setup, the obvious analogue is not so clear. We simplify the choice by using simple variances to capture scale differences in the volatilities of variables.

determined by a single parameter δ, and that parameter is estimated. More specifically, the prior variance for the intercept is $1000\sigma_y^2$. For all of the other coefficients $i = 2, \ldots, k$ of the coefficient vector β_τ, the prior takes the form $\beta_{\tau,i} \sim N(0, \delta C_{\tau,i})$, where $C_{\tau,i} = 1$ for the lag of the dependent variable and $C_{\tau,i} = \frac{\sigma_y^2}{\sigma_i^2}$ for other variables. Following Chan (2021), we specify a Gamma prior on δ parameterized to have mean consistent with our baseline Minnesota setting: $\delta \sim G(c_1, c_2)$, with the Gamma parameterized so that its mean is $c_1 c_2 = 0.04$. As established in Chan (2021), the conditional posterior for δ is generalized inverse Gaussian (GIG):

$$\delta \sim GIG\left(c_1 - (k-1)/2,\ 2/c_2,\ \sum_{i=2}^{k} \beta_{\tau,i}^2 / C_{\tau,i}\right).$$

The estimation algorithm includes a fourth step to sample δ from this conditional posterior distribution. The supplementary online appendix provides time series of estimates of δ for some of our applications (see Figures A23-A26). These estimates show sizable variation over time (across forecast origins) in the shrinkage estimate and considerable commonality of estimates across quantiles. In the case of inflation, for most quantiles the shrinkage estimates decline over time, whereas in the large GDP application, for some quantiles, the shrinkage estimate moved up in the later years of the sample.

In another specification (denoted BQR-HS), we consider a horseshoe prior for the regression parameters, following the quantile regression precedents of Mitchell, Poon, and Mazzi (2022) and Kohns and Szendrei (2021)). The horseshoe prior, originally developed by Carvalho, Polson, and Scott (2010), falls in the general class of global-local shrinkage priors, in which the prior variance has one global term that applies to all coefficients and another term that is specific to each coefficient. As summarized in Mitchell, Poon, and Mazzi (2022), these priors are flexible in the sense that they shrink the (presumably small) coefficients of less informative predictors to 0 but have fat tails so that the (presumably large, in relative terms) coefficients of more informative predictors are subjected to less shrinkage. Accordingly, the horseshoe prior might be expected to work well when a small number of predictors in a larger set of variables are useful. Of course, there are other options for global-local shrinkage priors (these include Normal-Gamma and Dirichlet-Laplace), but in this class, the horseshoe has the advantage that the researcher does not need to specify any hyperparameters. In addition, Mitchell, Poon, and Mazzi (2022) find that the horseshoe prior performs better than the Dirichlet-Laplace prior in a BQR forecasting setting, and Kohns and Szendrei (2021) find that the horseshoe prior works well in quantile regression forecasting with a large number of variables.

Our horseshoe implementation follows that of Mitchell, Poon, and Mazzi (2022), who in turn borrowed aspects of implementation from Cross, Hou, and Poon (2020). For all of the coefficients $i = 1, \ldots, k$ of the coefficient vector β_τ, the prior takes the form $\beta_{\tau,i} \sim N(0, \delta\lambda_i)$, where δ governs global shrinkage and λ_i governs the local shrinkage for coefficient i. In the general horseshoe setting, the priors for these shrinkage parameters are half-Cauchy distributions. Mitchell, Poon, and Mazzi (2022) and Cross, Hou, and Poon (2020) (see the latter for more details) exploit the scale-mixture representation of the half-Cauchy distribution to obtain conditional posterior distributions of known form for each coefficient. In turn, following these studies, we use conditional posterior distributions that are inverse Gamma (IG) to sample — in the fourth step of a Gibbs sampler — the following parameters that determine the prior variances of the coefficients in the horseshoe specification:

$$
\begin{aligned}
\lambda_i|\beta_{\tau,i}, \delta, \nu_i, \varepsilon &\sim IG\left(1, \frac{1}{\nu_i} + \frac{\beta_{\tau,i}^2}{2\delta}\right), i = 1, \ldots, k \\
\delta|\beta_{\tau,i}, \lambda_i, \nu_i, \varepsilon &\sim IG\left(\frac{k+1}{2}, \frac{1}{\varepsilon} + \frac{1}{2}\sum_{i=1}^{k} \frac{\beta_{\tau,i}^2}{\lambda_i}\right) \\
\nu_i|\beta_{\tau,i}, \lambda_i, \delta, \varepsilon &\sim IG\left(1, 1 + \frac{1}{\lambda_i}\right), i = 1, \ldots, k \\
\varepsilon|\beta_{\tau,i}, \lambda_i, \delta, \nu_i &\sim IG\left(1, 1 + \frac{1}{\delta}\right).
\end{aligned}
$$

In unreported results, we also consider an elastic net prior, following Mitchell, Poon, and Mazzi (2022), which encourages sparsity while working well with strongly correlated indicators. Since this prior has no advantage over the others in our results, we have omitted these results in the interest of brevity.

3 Applications and Data

This section first details the applications we use in examining the efficacy of alternative quantile forecasting specifications and then describes the data.

3.1 Applications

In total, we consider five different applications, largely drawn from the recent literature on the measurement and forecasting of macroeconomic tail risks. All use quarterly data. In all, we consider forecast horizons of 1, 4, and 12 quarters; in the interest of brevity, we focus on results for 1 and 4 quarters and more briefly refer to results for 12 quarters provided in the supplementary online appendix (see Figures A1-A3 and Tables A1-A2 in the appendix). For convenience, Table 1 lists the variables in each model, summarizing the explanations below and detailing transformations and lag orders.

Our first three applications focus on tail risks to real GDP growth, one with a small set of predictors, a second with a larger set, and a third with a quite large number of predictors. In these applications, y_t is annualized quarterly GDP growth computed as 400 times the log change. For h=1, $y_{t+h}^{(h)} \equiv y_{t+h}$. For h=4,12, we forecast the four-quarter average growth rate of GDP; at these horizons, $y_{t+h}^{(h)} = 100 \ln(\mathrm{GDP}_{t+h}/\mathrm{GDP}_{t+h-4})$.

In the *small GDP growth* application, based on the influential work of Adrian, Boyarchenko, and Giannone (2019), we examine tail risks to GDP growth with a model that relates growth to lagged growth and financial conditions as measured by the Chicago Fed's index of financial conditions (NFCI). Note that, in results for this application, we omit the QR-avg and PQR methods because the predictor set only includes one indicator apart from a constant and lagged GDP growth.

For the *large GDP growth* application, in view of some other work that has found a broader range of indicators to be helpful for capturing tail risks to economic activity, we examine a model that relates growth to lagged growth and a broad set of indicators. To have a relatively large model in this case, we pull in indicators from a range of studies, some that have considered tail risks to economic activity measures other than GDP growth. More specifically, in the large GDP growth application, our choice of broad indicators is informed by the results of applications in Caldara, et al. (2021) and Plagborg-Moller, et al. (2020), who find that broad factor indexes of economic and financial conditions have predictive content for growth-at-risk; in a range of studies in the forecasting literature that generally find credit spreads to be helpful for macroeconomic forecasting (e.g., Faust, et al. (2013)); and in Kiley (2022), who finds that a credit spread and medium-term changes in the credit-to-GDP ratio have predictive content for tail risks to economic activity as measured by the unemployment rate. Drawing on specifications from these studies, the broad set of indicators used for tail risk prediction in this large GDP growth application consists of the four NFCI subindexes for leverage, non-financial leverage, credit, and risk; the Baa corporate/10-year Treasury bond spread; the Aaa corporate/10-year Treasury bond spread; the four-year growth rate of the credit-to-nominal GDP ratio; and the Chicago Fed's national activity index (CFNAI) of the business cycle.

For the *very large GDP growth* application, we deliberately adjust the set of predictors of output growth to include even more variables but to have less overlap with the predictors in our other applications. We take the set of predictors from the variables included in a factor model designed for forecasting in Bok, et al. (2018), omitting the handful of their variables for which long time series are not available and adding as predictors the NFCI and the Baa corporate/10-year Treasury bond spread in order to include more financial indicators. The total number of predictors in the model (including lagged GDP growth) is 29. We transform the variables for stationarity following Bok, et al. (2018).

Following Kiley (2022), our fourth application examines tail risks to unemployment with a model that relates the change in the unemployment rate to the lagged level of the unemployment rate, the Baa corporate/10-year Treasury bond spread, the 10-year Treasury bond/federal funds rate term spread, the four-year growth rate of the credit-to-nominal GDP ratio, and the four-quarter rate of inflation in PCE prices excluding food and energy (henceforth, core PCE prices). In these results, our specification is patterned along the lines of Kiley (2022), to predict the change in unemployment rather than the level. In this case, the variable being predicted is $y_{t+h}^{(h)} = \mathrm{UR}_{t+h} - \mathrm{UR}_t$, where UR denotes the unemployment rate.

Finally, our fifth application considers tail risks to inflation, with a specification patterned after the model of Lopez-Salido and Loria (2022). Let y_t denote annualized quarterly inflation in the core PCE price index (computed as 400 times the log change), and let $\tilde{y}_{t+h}^{(h)} \equiv h^{-1} \sum_{i=1}^{h} y_{t+i}$ denote the h-period average inflation rate. In this application, the variable forecast is the h-period change in the $\tilde{h}$-period average inflation rate, where $\tilde{h}$ equals h if the horizon is four quarters or less but $\tilde{h}$ equals 4 otherwise: $y_{t+h}^{(h)} = \tilde{y}_{t+h}^{(h)} - \tilde{y}_t^{(h)}$ (so the 1-step-ahead predictand is the 1-period change in the annualized quarter-on-quarter inflation rate, whereas at the longer horizons the h-steps-ahead predictand is the h-period change in the year-on-year (four-quarter) inflation rate). The model relates the change in core PCE inflation (as just defined) to the lagged difference between core PCE inflation (measured on a four-quarter basis) and a long-run survey-based measure of inflation expectations taken from the Federal Reserve Board's FRB/US model (denoted PTR in the model); an unemployment rate gap measured as the actual rate less the Congressional Budget Office's (CBO's) estimate of the non-cyclical rate of unemployment; relative import price inflation; and the four-year growth rate of the credit-to-nominal GDP ratio.[4] We include the credit/GDP measure on the basis of evidence (for the euro area) in Korobilis, et al. (2021) that credit and money measures are helpful for quantile forecasts of inflation.

3.2 Data

All data, listed in Table 2, are from the FRED, FRED-MD, or FRED-QD databases maintained by the Federal Reserve Bank of St. Louis, with the exception of the measure of long-run inflation expectations (PTR) obtained from the public data files for the Federal Reserve Board's FRB/US model and indexes of manufacturing from the Institute for Supply Management obtained from Haver Analytics. We use data at the quarterly frequency. For those series available at higher frequencies, quarterly data are simple averages of the higher-frequency data.[5]

4 Forecast Evaluation

Although some other work has focused on the efficacy of alternative approaches to combining information from forecasts across quantiles to obtain density forecasts, in this paper we focus on the efficacy of forecasting the quantiles of interest themselves, as this is often the focus of macroeconomic applications. We do include some measures that summarize results across the predictive density, but our focus is not the densities per se.

To assess efficacy in a range of possible applications, we consider a range of 11 quantiles: $\tau =$

[4]To simplify estimation, our model slightly adjusts the timing used by Lopez-Salido and Loria (2022) and incorporates a linear transformation that puts the change in inflation on the left-hand side and makes the model linear. Specifically, Lopez-Salido and Loria (2022) put the level of inflation in period $t+h$ on the left-hand side and relate it to inflation in period $t-1$ and the inflation expectation in period t, with coefficients that sum to 1 — necessitating the imposition of the coefficient sum restriction in estimation. We instead date the lagged inflation term on the right-hand side as period t and impose the coefficient sum restriction, so as to put the change in inflation on the left-hand side and reduce the two terms for lagged inflation and the inflation to lagged inflation less the long-run inflation expectation.

[5]These series include the NFCI and its components, the bond yields and federal funds rate, and the CFNAI.

$0.05, 0.10, 0.2, 0.3, \ldots, 0.90, 0.95$. For some applications (GDP growth), the lower tail is mostly of interest. Of course, the tail quantile forecast corresponds to the value at risk (VaR) forecast.[6] For other applications (e.g., unemployment), it is the upper tail that may be of more interest.

Our evaluation of tail risk forecasts includes a few different forecast metrics. The first tail metric is the quantile score, commonly associated with the tick loss function (see, e.g., Giacomini and Komunjer (2005)). The quantile score (QS) is computed as

$$\mathrm{QS}_{\tau,t+h} = (y_{t+h}^{(h)} - Q_{\tau,t+h})(\tau - \mathbf{1}_{\{y_{t+h}^{(h)} \leq Q_{\tau,t+h}\}}), \tag{8}$$

where $y_{t+h}^{(h)}$ is the actual outcome of the variable of interest, $Q_{\tau,t+h}$ is the forecast of quantile τ, and the indicator function $\mathbf{1}_{\{y_{t+h}^{(h)} \leq Q_{\tau,t+h}\}}$ has a value of 1 if the outcome is at or below the forecast quantile and 0 otherwise.

We also evaluate forecasts with metrics based on the quantile-weighted continuous ranked probability scores (qwCRPS). Gneiting and Ranjan (2011) develop the qwCRPS as a proper scoring function of the entire predictive density with quantile weighting to emphasize either (as chosen with particular weights) the left tail, right tail, or the center of the distribution. The qwCRPS is computed as a weighted sum of quantile scores at a range of quantiles:

$$\mathrm{qwCRPS}_{t+h} = \frac{2}{J-1} \sum_{j=1}^{J-1} v(\tau_j) \mathrm{QS}_{\tau_j,t+h}, \tag{9}$$

with $\tau_j = j/J$. In all CRPS computations, we set $J = 10$ and compute the score using 9 evenly spaced quantiles of $0.10, 0.2, 0.3, \ldots, 0.90$. The left-tail-focused (CRPS-L) gives more weight to the left tail than the right, with a weighting function of $v(\tau_j) = (1 - \tau_j)^2$. The right-tail-focused (CRPS-R) instead assigns weights of $v(\tau_j) = \tau_j^2$, whereas the center-weighted (CRPS-C) uses $v(\tau_j) = \tau_j(1 - \tau_j)$.

For these metrics, we report results for each alternative forecast approach relative to a simple quantile regression benchmark for the same application.[7] In the tables and charts, entries of less (more) than 1 mean that a given specification is more (less) accurate than the simple QR baseline.

[6] Adrian, et al. (2022) coined the term "growth at risk" for GDP growth forecasts, and Denicolo and Lucchetta (2017) coined similar terms for industrial production and employment.

[7] See Amburgey and McCracken (2022) for a comparison of results from the small GDP growth model of Adrian, Boyarchenko, and Giannone (2019) to a quantile regression including just lagged GDP growth.

In our primary results, we compare out-of-sample forecast accuracy for predictions produced starting in 1985:Q1 — using data through the previous quarter for model estimation — and ending in either (for reasons given below) 2019 or 2021. In the primary results, the forecasts are produced with a recursive scheme, allowing the sample used for model estimation to expand as forecasting moves forward in time. The starting point of the sample used for model estimation varies by application, reflecting availability of data on the indicators used and adjustments for the forecast horizon. In the GDP applications for horizon h, the sample used for estimation starts with 1971:Q2+$h-1$. In the unemployment and inflation applications for horizon h, the sample used for estimation starts with 1963:Q2+$h-1$. As a robustness check of the main results, we also consider forecasts produced with a rolling scheme. In the first case, we keep the size of the sample used for model estimation constant at its initial size as forecasting moves forward in time; what we refer to below as the "rolling scheme" fixes the sample size at $55-h$ observations in the GDP applications and $87-h$ observations in the unemployment and inflation applications. In the second case, we set the rolling window size to 40 observations, for all applications.

5 Results

In this section we assess the forecasting performance of the various quantile regression methods introduced in Section 2 for the applications discussed in Section 3. We first provide results and then discuss, in turn, the quantile estimates, the variables with the largest predictive power, and temporal stability.

5.1 Main results

The main results on forecast accuracy — using the recursive scheme — are presented in Figures 1 and 2 and Tables 3 and 4, for the period ending in 2019. From these results we can draw both general conclusions, which apply across methods, quantiles, horizons, and applications, and more specific findings. To gauge the statistical significance of differences in scores, the tables include the results of Diebold and Mariano (1995)-West (1996) t-tests for equality of the average loss. We conduct these tests on a one-sided basis, such that the alternative hypothesis is that the indicated forecast is more accurate than the QR benchmark.[8] The tables also denote in blue the best-performing specification for a given forecast horizon and metric.

Starting with the general conclusions, four main ones are worth mentioning. First, shrinkage is generally helpful to tail forecast accuracy, yielding consistent gains relative to QR, with a few exceptions (mostly in the small GDP application in the left tail and in the right tail at longer horizons). The gains are large in the large GDP, very large GDP, unemployment, and inflation applications, and in all applications except small GDP, they generally increase with the forecast horizon. The gains are often, although not always, statistically significant. This finding suggests that newer analyses of risks to GDP growth with larger sets of predictors, inflation, and unemployment would be better based on PQR or Bayesian QR than classical QR, which has been most commonly used in the empirical macroeconomics literature to date. Moreover, in the GDP growth applications, the most accurate results in the left and right tails are obtained with the very large information set, confirming previous results in the forecasting literature that during uncommon periods a large information set can help.

Second, based on magnitudes of gains and consistency across applications, horizons, and quantiles, PQR may be seen as the best-performing method. Yet, some of the charts (reported below) of time series of quantile forecasts and factors raise some questions as to whether the performance of PQR is somewhat illusory, at least in that the score measures fail to reflect some possible problems with the PQR forecasts. At longer horizons the PQR tail quantile estimates can be very volatile, in particular for unemployment but also for inflation.

Third, other shrinkage approaches — frequentist and Bayesian (with neither having a clear edge over the other) — are also typically helpful to accuracy. In particular, QR-ridge performs relatively well, if not quite matching the overall accuracy of PQR.

Fourth, within the BQR family of approaches, nothing clearly beats simply fixing a Minnesota-type prior at our baseline setting. On balance, estimating the Minnesota tightness is not helpful relative to fixing tightness as in a standard Minnesota prior. Occasionally, BQR-estMN beats BQR-MN, but far more often than not, the reverse is true. Considering other BQR implementations, throughout BQR-HS does not seem to have a consistent advantage over BQR-MN. The simple

[8] We compute the variance underlying the test statistics using the pre-whitened quadratic spectral estimator of Andrews and Monahan (1992).

prior treatment used in BQR-MN is most often better, except in a few cases for the small GDP application.

Focusing now on patterns across quantiles, we can highlight three main ones. First, as mentioned, in most applications, PQR is best or close to best in the left tail, the middle, and the right tail. In the small GDP application, where PQR is not used due to the small data set, for the left and right tails there are only little differences across methods. Instead, in the large GDP application the gains from PQR (relative to the QR baseline from the small GDP variable set) reach 24 percent in the right tail for h=1 and 22 percent in the left tail for h=4; with the very large data set, the gains from PQR further increase to 26 percent in the right tail for h=1 and 30 percent in the left tail for h=4.

Second, for the small GDP application, it is interesting that in the middle quantiles QR-ridge is better than QR, by 6 to 9 percent, suggesting that shrinkage can help even with a small data set. PQR is best or generally among the best for middle quantiles in most of the other applications, though the gains with respect to the benchmark QR from methods such as QR-ridge remain.

Third, while for unemployment PQR is generally best across quantiles, with very large gains up to 40 percent in the right tail for h=4, in the inflation application there is substantial heterogeneity across quantiles. For h=1, BQR-MN is best in the left tail, PQR is best in the middle, and QR-ridge is best in the right tail, though the differences across these three methods are generally small across quantiles. For h=4, QR-ridge is best in the middle and high quantiles.[9] Overall, in the inflation application, it is hard to say one method is best for the left tail, but one consistent pattern is that BQR-MN offers a consistent improvement over the QR baseline.

Moving to patterns across forecast horizons, including the results for h=12 reported in the supplementary online appendix (see its Section 1), in most applications accuracy gains relative to QR increase with the forecast horizon. The small GDP application is an exception to this pattern. Moreover, PQR's advantage over other methods grows in the large GDP, very large GDP, and unemployment rate applications as compared to other applications. Related, except in the large and very large GDP applications, the prevalence of statistical significance is generally lower at longer horizons than at shorter ones.

[9]From the supplementary online appendix (see its Section 1), for h=12, the performance of BQR-MN varies widely across quantiles, from best in the left tail to middle of the pack (and beaten notably by PQR) in middle and high quantiles.

5.2 Selected quantile estimates

The 0.10 and 0.90 quantile estimates are reported in Figures 3 and 4 (sample ending before pandemic) for selected methods, the two forecast horizons, and the five applications.

It turns out that the quantile forecasts are more similar — across methods — for horizons of 1 and 4 quarters than for 12 quarters (Figure A2 in the supplementary online appendix reports the h=12 quantiles). For example, in the small GDP application, at the shorter horizons, the upper tail quantiles are very similar. Similarly, in the unemployment application, the quantile forecasts show fairly strong comovement across methods at the shorter horizons.[10]

To shed some more light on the PQR performance in forecasting, Figure 5 plots the estimated PQR factors at the 0.10 and 0.90 quantiles for the various forecast horizons and applications. It turns out that, especially in the unemployment and inflation applications where the estimated PQR quantiles were highly volatile, the PQR factor for the h=1 case is far more variable than the factors for the longer horizon of h=4 (and h=12, reported in Section 1 of the supplementary online appendix). Moreover, unreported results on correlations of the full-sample factors across horizons and quantiles, and with the variables included in each model, show that the factors for h=1 and h=4 are correlated with one another (more than with the h=12 factor estimate). However, for a fixed forecast horizon, factors are strongly correlated across quantiles.

5.3 What variables are most important to predictive content?

To tackle this question, we consider: (1) quantile scores of QR models with one indicator at a time relative to a QR model with just lags of the dependent variable to provide a rough measure of the predictive value of each indicator and (2) full-sample (pre-pandemic) coefficient estimates from the BQR-MN specification. The results (available upon request) point to the following.

First, for GDP growth, with some variation across left and right quantiles and forecast horizons, the variables with relatively strong predictive content are the CFNAI and a few of the NFCI

[10]The PQR-based forecast is quite variable at the 12-steps-ahead horizon for both unemployment and inflation. Moreover, in the small GDP application, at the 12-steps-ahead horizon, the quantile forecasts (particularly, the QR and BQR-MN forecasts) have a difficult time picking up sizable cyclical changes in downside risk. This likely reflects the limited predictability of GDP growth — as the horizon grows — in general (see, e.g., Breitung and Knüppel (2021)) and more specifically using just the NFCI. But even in the large and very large GDP applications, at the long horizon, the QR-ridge and BQR-MN forecasts are challenged to show much cyclicality in downside risk. In this case, the PQR-based quantile forecast shows more variation. But the PQR-based forecast more generally tends to be relatively volatile in the other applications as well.

components. In the right tail, the Aaa credit spread also has some predictive content, as well as the unemployment rate in the very large GDP application.

Second, for unemployment, although full-sample coefficient estimates show most variables to enter significantly (with posterior means/standard deviations above 2 in absolute value), out-of-sample predictive value is less clear. For h=1, the credit/GDP ratio consistently helps, especially in the right tail. The term and credit spreads also help a little in the right tail. For the longer horizons, there is not much that helps OOS predictive content in the tails, despite full-sample coefficients that appear significant.

Finally, for inflation, consistent with a broader literature on challenges in forecasting inflation, not many of the predictors considered clearly have significant predictive content for tail risks to inflation. For h=1, the inflation gap and import inflation helps in the left tail. But nothing makes much consistent difference in the right tail quantiles at this horizon. Most of the same pattern applies for h=4 and h=12.

5.4 Are the results stable over time?

To assess whether the results we have discussed so far are stable over time, we have conducted two robustness analyses. First, we have replaced recursive with rolling estimation of the forecasting models. Second, using recursive estimation, we have recursively conducted the model comparison over the evaluation sample, extending the latter to also cover the pandemic period of 2020-21, which is interesting by itself.

In the interest of brevity, figures and tables with the rolling window results are provided in the supplementary online appendix (see its Section 2; note that these omit the very large GDP application to avoid possible issues with estimating a very large model with a very small sample).[11] With few exceptions, the key general findings described above for the recursive scheme remain valid.

The results on performance over time, including over the pandemic, are also presented in the supplementary online appendix (see its Figures A8-A12). The figures provide score averages over

[11] As noted above, we consider two different rolling window sample sizes. In the first case, the rolling window size varies across the applications. Specifically, the rolling window sample size is set to be whatever it is for the first (1985:Q1) forecast origin. This makes the sample size smaller for the GDP growth applications than the other two (due to the NFCI data not starting until 1971). In the second case, the rolling window size is 40 observations.

recursively expanding windows, using forecasts that start with 1985.[12] For readability, the charts focus on a few of the best-performing methods (QR ridge, PQR, and BQR-MN) and four scores (quantile scores of 0.10 and 0.90 and the left- and right-tail-weighted CRPS).[13]

Relative to the baseline QR specification, forecast accuracy shows some variation over time, although not for all applications, horizons, and score metrics. Nonetheless, the relative rankings of the methods shown are mostly (although not entirely) stable over time. That is, the method that is best at the end of the sample (corresponding to the full sample, including the pandemic) is also best over earlier evaluation windows. For example, in the unemployment application, PQR is more accurate than other methods throughout the sample.

There is some tendency for time variation around the 2007-2009 Great Recession and the pandemic, with changes around the former usually sharper than in the pandemic. Around the Great Recession, score ratios (relative to QR) typically rise. However, changes in accuracy around the Great Recession are less evident for inflation than GDP growth or the unemployment rate. Around the pandemic, score ratios rise in some cases and fall in others.

Overall, while there is some evidence of instability, it does not seem to affect qualitatively the model and method rankings obtained over the full evaluation sample.

6 Conclusions

Quantile regression has become widely used in empirical macroeconomics, in particular for estimating and forecasting tail risks to macroeconomic indicators. In this paper we examine various choices in the specification of quantile regressions for macro applications. Specifically, focusing on applications to GDP growth, inflation, and unemployment, we compare the accuracy of quantile forecasts at different horizons, obtained from simple quantile regression, averages of QR forecasts obtained with one indicator at a time, partial quantile regression, quantile regression with ridge penalty, and a few different specifications of priors for Bayesian quantile regression, including a fixed Minnesota-type prior and a horseshoe prior. In comparing accuracy, we consider a range of quan-

[12] As noted before, the forecasts in the figures are dated by the forecast origin, not the outcome date. So, for example, forecasts for an outcome date of 2020:Q2 (the quarter of the sharp economic downturn induced by the pandemic shutdown) correspond to forecast origins of 2020:Q2$-h+1$ for h-steps-ahead forecasts.

[13] Similar figures in the supplementary online appendix (see its Figures A13-A20) provide corresponding results for forecasts obtained with the rolling window schemes described above; these results are qualitatively very similar to the recursive scheme results described here.

tiles — spanning from left to right tail — and use the quantile score as well as quantile-weighted continuous ranked probability scores that consider the entire predictive density.

We find that shrinkage is generally helpful to tail forecast accuracy, yielding consistent gains relative to standard QR. The gains are particularly large for GDP applications featuring large sets of predictors and unemployment and inflation applications, and the gains increase with the forecast horizon. Based on magnitudes of gains and consistency across applications, horizons, and quantiles, PQR may be seen as the best-performing method. Yet, at long horizons the PQR tail quantile estimates can be volatile, in particular for unemployment but also for inflation. Other shrinkage approaches — frequentist and Bayesian (with neither having a clear edge over the other) — are also typically helpful to accuracy, in particular QR-ridge. Moreover, within the Bayesian QR family of approaches, nothing clearly beats simply fixing a Minnesota-type prior at our baseline setting. Finally, while there is some evidence of instability, it does not seem to affect qualitatively the model and method rankings obtained over the full evaluation sample.

Therefore, we can conclude that newer analyses of risks to growth, inflation, and unemployment would be better based on PQR or Bayesian QR rather than classical QR, which to this point has been most commonly used in macroeconomic applications. We would recommend that, at a minimum, researchers using frequentist QR check the robustness of their results to the use of at least one estimation approach featuring shrinkage.

References

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone (2019), "Vulnerable growth," *American Economic Review*, 109, 1263–89, `https://doi.org/10.1257/aer.20161923`.

Adrian, Tobias, Federico Grinberg, Nellie Liang, Sheheryar Malik, and Jie Yu (2022), "The term structure of growth-at-risk," *American Economic Journal: Macroeconomics*, 14, 283–323, `https://doi.org/10.1257/mac.20180428`.

Amburgey, Aaron, and Michael W. McCracken (2022), "On the real-time predictive content of financial conditions indices for growth," Working Paper 2022-003, Federal Reserve Bank of St. Louis, `https://doi.org/10.20955/wp.2022.003`.

Andrews, Donald W.K., and J. Christopher Monahan (1992), "An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator," *Econometrica*, 60, 953–66, https://doi.org/10.2307/2951574.

Bayer, Sebastian (2018), "Combining value-at-risk forecasts using penalized quantile regressions," *Econometrics and Statistics*, 8, 56–77, https://doi.org/10.1016/j.ecosta.2017.08.001.

Bok, Brandyn, Daniele Caratelli, Domenico Giannone, Argia M. Sbordone, and Andrea Tambalotti (2018), "Macroeconomic nowcasting and forecasting with big data," *Annual Review of Economics*, 10, 615–643, https://doi.org/10.1146/annurev-economics-080217-053214.

Breitung, Jörg, and Malte Knüppel (2021), "How far can we forecast? Statistical tests of the predictive content," *Journal of Applied Econometrics*, 36, 369–392, https://doi.org/10.1002/jae.2817.

Caldara, Dario, Danilo Cascaldi-Garcia, Pablo Cuba-Borda, and Francesca Loria (2021), "Understanding growth-at-risk: A Markov switching approach," manuscript, Board of Governors of the Federal Reserve System, https://doi.org/10.2139/ssrn.3992793.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2015), "Bayesian VARs: Specification choices and forecast accuracy," *Journal of Applied Econometrics*, 30, 46–73, https://doi.org/10.1002/jae.2315.

——— (2022), "Nowcasting tail risk to economic activity at a weekly frequency," *Journal of Applied Econometrics*, 37, 843–866, https://doi.org/10.1002/jae.2903.

Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott (2010), "The horseshoe estimator for sparse signals," *Biometrika*, 97, 465–480, https://doi.org/10.1093/biomet/asq017.

Chan, Joshua C.C. (2021), "Minnesota-type adaptive hierarchical priors for large Bayesian VARs," *International Journal of Forecasting*, 37, 1212–1226, https://doi.org/10.1016/j.ijforecast.2021.01.002.

Chernozhukov, Victor, Ivan Fernandez-Val, and Tetsuya Kaji (2017), "Extremal quantile regression," in *Handbook of Quantile Regression* eds. by Roger Koenker, Victor Chernozhukov, Xuming He, and Limin Peng: Chapman and Hall/CRC: New York, https://doi.org/10.1201/9781315120256-18.

Cross, Jamie L., Chenghan Hou, and Aubrey Poon (2020), "Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity," *International Journal of Forecasting*, 36, 899–915, https://doi.org/https://doi.org/10.1016/j.ijforecast.2019.10.002.

Denicolo, Gianni, and Marcella Lucchetta (2017), "Forecasting tail risks," *Journal of Applied Econometrics*, 32, 159–170, https://doi.org/10.1002/jae.2509.

Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–63, https://doi.org/10.2307/1392185.

Faust, Jon, Simon Gilchrist, Jonathan H. Wright, and Egon Zakrajšsek (2013), "Credit spreads as predictors of real-time economic activity: A Bayesian model-averaging approach," *Review of Economics and Statistics*, 95, 1501–1519, https://doi.org/10.1162/REST_a_00376.

Gaglianone, Wagner Piazza, and Luiz Renato Lima (2012), "Constructing density forecasts from quantile regressions," *Journal of Money, Credit and Banking*, 44, 1589–1607, https://doi.org/10.1111/j.1538-4616.2012.00545.x.

Giacomini, Raffaella, and Ivana Komunjer (2005), "Evaluation and combination of conditional quantile forecasts," *Journal of Business and Economic Statistics*, 23, 416–431, https://doi.org/10.1198/073500105000000018.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri (2015), "Prior selection for vector autoregressions," *The Review of Economics and Statistics*, 97, 436–451, https://doi.org/10.1162/REST_a_00483.

Giglio, Stefano, Bryan Kelly, and Seth Pruitt (2016), "Systemic risk and the macroeconomy: An empirical evaluation," *Journal of Financial Economics*, 119, 457–471, https://doi.org/10.1016/j.jfineco.2016.01.010.

Gneiting, Tilmann, and Roopesh Ranjan (2011), "Comparing density forecasts using threshold- and quantile-weighted scoring rules," *Journal of Business and Economic Statistics*, 29, 411–422, https://doi.org/10.1198/jbes.2010.08110.

Khare, Kshitij, and James P. Hobert (2012), "Geometric ergodicity of the Gibbs sampler for Bayesian quantile regression," *Journal of Multivariate Analysis*, 112, 108–116, https://doi.org/10.1016/j.jmva.2012.05.004.

Kiley, Michael T. (2022), "Unemployment risk," *Journal of Money, Credit and Banking*, 54, 1407–1424, https://doi.org/10.1111/jmcb.12888.

Kohns, David, and Tibor Szendrei (2021), "Horseshoe prior Bayesian quantile regression," Papers 2006.07655, arXiv.org.

Korobilis, Dimitris (2017), "Quantile regression forecasts of inflation under model uncertainty," *International Journal of Forecasting*, 33, 11–20, https://doi.org/10.1016/j.ijforecast.2016.07.005.

Korobilis, Dimitris, Bettina Landau, Alberto Musso, and Anthoulla Phella (2021), "The time-varying evolution of inflation risks," Working Paper 2600, European Central Bank, https://doi.org/10.2139/ssrn.3938628.

Lima, Luiz Renato, Fanning Meng, and Lucas Godeiro (2020), "Quantile forecasting with mixed-frequency data," *International Journal of Forecasting*, 36, 1149–1162, https://doi.org/10.1016/j.ijforecast.2018.09.011.

Lopez-Salido, David, and Francesca Loria (2022), "Inflation at risk," *manuscript*, https://doi.org/10.2139/ssrn.4002673.

Manzan, Sebastiano (2015), "Forecasting the distribution of economic variables in a data-rich environment," *Journal of Business and Economic Statistics*, 33, 144–164, https://doi.org/10.1080/07350015.2014.937436.

Manzan, Sebastiano, and Dawit Zerom (2013), "Are macroeconomic variables useful for forecasting the distribution of U.S. inflation?" *International Journal of Forecasting*, 29, 469–478, https://doi.org/10.1016/j.ijforecast.2013.01.005.

Mitchell, James, Aubrey Poon, and Gian Luigi Mazzi (2022), "Nowcasting euro area GDP growth using Bayesian quantile regression," in *Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling* eds. by Alexander Chudik, Cheng Hsiao, and Allan Timmermann, Advances in Econometrics, vol. 43A: Emerald Publishing Limited, Bingley, 51–72, https://doi.org/10.1108/S0731-90532021000043A004.

Mitchell, James, Aubrey Poon, and Dan Zhu (2022), "Constructing density forecasts from quantile regressions: Multimodality in macro-financial dynamics," Working Paper 22-12, Federal Reserve Bank of Cleveland, https://doi.org/10.26509/frbc-wp-202212.

Plagborg-Moller, Mikkel, Lucrezia Reichlin, Giovanni Ricco, and Thomas Hasenzagl (2020), "When is growth at risk?" *Brookings Papers on Economic Activity*, Spring, 167–229, https://doi.org/10.1353/eca.2020.0002.

West, Kenneth D. (1996), "Asymptotic inference about predictive ability," *Econometrica*, 64, 1067–1084, https://doi.org/10.2307/2171956.

Yi, Congrui, and Jian Huang (2017), "Semismooth Newton coordinate descent algorithm for elastic-net penalized Huber loss regression and quantile regression," *Journal of Computational and Graphical Statistics*, 26, 547–557, https://doi.org/10.1080/10618600.2016.1256816.

Yu, Keming, and Rana A. Moyeed (2001), "Bayesian quantile regression," *Statistics and Probability Letters*, 54, 437–447, https://doi.org/10.1016/S0167-7152(01)00124-9.

Zou, Hui, and Trevor Hastie (2005), "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 301–320, https://doi.org/10.1111/j.1467-9868.2005.00503.x.

Table 1: Application specifications

Application	Dependent variable	Lagged dep. variable	Other predictors (common to all horizons)
Small GDP			
$h = 1$	$400 \ln \left(\mathrm{GDP}_{t+1}/\mathrm{GDP}_t\right)$	$400 \ln \left(\mathrm{GDP}_t/\mathrm{GDP}_{t-1}\right)$	NFCI_t
$h = 4, 12$	$100 \ln \left(\mathrm{GDP}_{t+h}/\mathrm{GDP}_{t+h-4}\right)$	$400 \ln \left(\mathrm{GDP}_t/\mathrm{GDP}_{t-1}\right)$	
Large GDP			
$h = 1$	$400 \ln \left(\mathrm{GDP}_{t+1}/\mathrm{GDP}_t\right)$	$400 \ln \left(\mathrm{GDP}_t/\mathrm{GDP}_{t-1}\right)$	NFCI:leverage$_t$, NFCI:credit$_t$, NFCI:nonfin. leverage$_t$, NFCI:risk$_t$
$h = 4, 12$	$100 \ln \left(\mathrm{GDP}_{t+h}/\mathrm{GDP}_{t+h-4}\right)$	$100 \ln \left(\mathrm{GDP}_t/\mathrm{GDP}_{t-4}\right)$	Baa$_t$ − 10 yr. Treasury$_t$, Aaa$_t$ − 10 yr. Treasury$_t$, $\ln(\mathrm{C/NGDP})_t - \ln(\mathrm{C/NGDP})_{t-16}$, CFNAI$_t$
Very Large GDP			
$h = 1$	$400 \ln \left(\mathrm{GDP}_{t+1}/\mathrm{GDP}_t\right)$	$400 \ln \left(\mathrm{GDP}_t/\mathrm{GDP}_{t-1}\right)$	period t values of 28 indicators listed in Table 2
$h = 4, 12$	$100 \ln \left(\mathrm{GDP}_{t+h}/\mathrm{GDP}_{t+h-4}\right)$	$100 \ln \left(\mathrm{GDP}_t/\mathrm{GDP}_{t-4}\right)$	period t values of 28 indicators listed in Table 2
Unemployment			
$h = 1$	$\mathrm{UR}_{t+1} - \mathrm{UR}_t$	UR_t	Baa$_t$ − 10 yr. Treasury$_t$, 10 yr. Treasury$_t$ − Fed. funds$_t$,
$h = 4, 12$	$\mathrm{UR}_{t+h} - \mathrm{UR}_t$	UR_t	$\ln(\mathrm{C/NGDP})_t - \ln(\mathrm{C/NGDP})_{t-16}$, $100 \ln \left(\mathrm{PCEXPI}_t/\mathrm{PCEXPI}_{t-4}\right)$
Inflation			
$h = 1$	$400 \ln \left(\mathrm{PCEXPI}_{t+1}/\mathrm{PCEXPI}_t\right) -$ $400 \ln \left(\mathrm{PCEXPI}_t/\mathrm{PCEXPI}_{t-1}\right)$	$100 \ln \left(\mathrm{PCEXPI}_t/\mathrm{PCEXPI}_{t-4}\right) -$ PTR_t	UR_t − CBO nat. rate$_t$,
$h = 4, 12$	$100 \ln \left(\mathrm{PCEXPI}_{t+h}/\mathrm{PCEXPI}_{t+h-4}\right) -$ $100 \ln \left(\mathrm{PCEXPI}_t/\mathrm{PCEXPI}_{t-4}\right)$	$100 \ln \left(\mathrm{PCEXPI}_t/\mathrm{PCEXPI}_{t-4}\right) -$ PTR_t	$100 \ln \left(\mathrm{PIMPORT}_t/\mathrm{PIMPORT}_{t-4}\right) - 100 \ln \left(\mathrm{PCEXPI}_t/\mathrm{PCEXPI}_{t-4}\right)$, $\ln(\mathrm{C/NGDP})_t - \ln(\mathrm{C/NGDP})_{t-16}$

Note: To save table space, C/NGDP refers to the ratio of the credit variable to nominal GDP, PCEXPI refers to the core PCE price index, and PIMPORT refers to the import price index.

Table 2: Data Series

Mnemonic	Series	Small GDP	Large GDP	Very large GDP	Unemp.	Inflation
GDPC1	Real Gross Domestic Product	Y	Y	Y		
UNRATE (FRED-MD)	Civilian Unemployment Rate			Y	Y	Y
NFCI	Chicago Fed National Financial Conditions Index	Y		Y		
GDP	Nominal Gross Domestic Product		Y		Y	Y
CRDQUSAPABIS	Total Credit to Private Non-Financial Sector		Y		Y	Y
JCXFE	Chain price index for Personal Consumption Expend. ex. food and energy			Y	Y	Y
NROU	Noncyclical Rate of unemployment, Congressional Budget Office					Y
B021RG3Q086SBEA	Chain price index for imports of goods and services					Y
NFCILEVERAGE	NFCI Leverage Subindex		Y			
NFCICREDIT	NFCI Credit Subindex		Y			
NFCINONFINLEVERAGE	NFCI Index Nonfinancial Leverage Subindex		Y			
NFCIRISK	NFCI Risk Subindex		Y			
AAA	Moody's Seasoned Aaa Corporate Bond Yield		Y			
DBAA	Moody's Seasoned Baa Corporate Bond Yield		Y	Y	Y	
DGS10	Yield on U.S. Treasury Securities at 10-Year Constant Maturity		Y	Y	Y	
CFNAI	Chicago Fed National Activity Index		Y			
DFF	Federal Funds Effective Rate				Y	
PTR (Board of Governors)	Long-run inflation expectation in the FRB/US model					Y
TCU	Capacity Utilization: Total Index			Y		
HSN1F	New One Family Houses Sold			Y		
GACDFSA066MSFRBPHI	Diffusion Index for Current General Activity, Philadelphia Fed			Y		
CPILFESL	Consumer Price Index: All Items Less Food and Energy			Y		
DSPIC96	Real Disposable Personal Income			Y		
A261RX1Q020SBEA	Real gross domestic income			Y		
PAYEMS (FRED-MD)	All Employees: Total nonfarm			Y		
CPIAUCSL (FRED-MD)	Consumer Price Index: All Items			Y		
RETAILx (FRED-MD)	Retail and Food Services Sales			Y		
HOUST (FRED-MD)	Housing Starts: Total New Privately Owned			Y		
INDPRO (FRED-MD)	IP Index			Y		
PERMIT (FRED-MD)	New Private Housing Permits			Y		
PCEPI (FRED-MD)	Personal Consumption Expenditures, Chain Price Index			Y		
WPSFD49207 (FRED-MD)	PPI: Finished Goods			Y		
BUSINVx (FRED-MD)	Total Business Inventories			Y		
AMDMNOx (FRED-MD)	New Orders for Durable Goods			Y		
HWI (FRED-MD)	Help-Wanted Index			Y		
DPCERA3M086SBEA (FRED-MD)	Real personal consumption expenditures			Y		
CMRMTSPLx (FRED-MD)	Real Manuf. and Trade Industries Sales			Y		
AMDMUOx (FRED-MD)	Unfilled Orders for Durable Goods			Y		
ULCNFB (FRED-QD)	Nonfarm Business Sector: Unit Labor Cost			Y		
ISM (Haver)	ISM Index for Manufacturing: Overall			Y		
ISM-emp (Haver)	ISM Index for Manufacturing: Employment			Y		
ISM-prices (Haver)	ISM Index for Manufacturing: Prices			Y		

Note: All data from FRED (except as noted), with mnemonics indicated in the first column.

Table 3: GDP applications

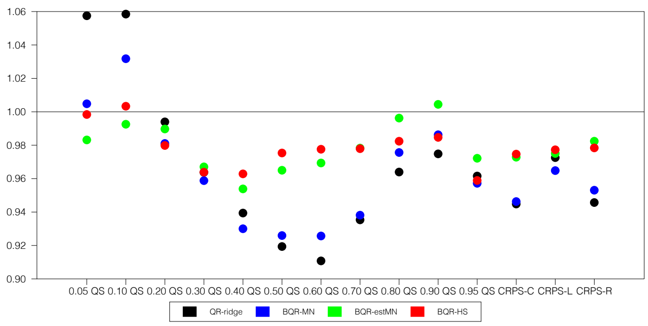
Metric	QR	QR ridge	BQR MN	BQR estMN	BQR HS	QR ridge	QR avg	PQR	BQR MN	BQR estMN	BQR HS	QR ridge	QR avg	PQR	BQR MN	BQR estMN	BQR HS
	Small GDP					Large GDP						Very Large GDP					
h=1																	
0.05 QS	0.25	1.06	1.00	0.98	1.00	1.08	1.01	0.88	1.01	0.89	0.92	1.03	0.92	0.74‡	0.91	0.95	0.89
0.10 QS	0.38	1.06	1.03	0.99	1.00	1.12	1.02	0.95	0.96	0.93	0.95	1.05	0.99	0.83†	0.89	0.92	0.91
0.20 QS	0.58	0.99	0.98	0.99	0.98‡	1.00	0.99	0.90†	0.91†	0.91	0.90	0.97	0.99	0.89†	0.89‡	0.92	0.95
0.30 QS	0.72	0.96	0.96	0.97‡	0.96*	0.94	0.96	0.87‡	0.88‡	0.91	0.89†	0.93†	0.96	0.91‡	0.89‡	0.92	0.94
0.40 QS	0.81	0.94‡	0.93*	0.95*	0.96*	0.91*	0.93*	0.87‡	0.87*	0.92	0.92	0.91*	0.92†	0.96	0.91†	0.94	0.94
0.50 QS	0.83	0.92*	0.93*	0.97*	0.98*	0.88*	0.90*	0.88‡	0.88*	0.93	0.93	0.89*	0.95	0.99	0.96	0.97	0.96
0.60 QS	0.82	0.91*	0.93*	0.97‡	0.98‡	0.89*	0.88*	0.91‡	0.88*	0.91‡	0.92†	0.90*	0.94	1.02	0.94	0.97	0.96
0.70 QS	0.78	0.94‡	0.94*	0.98†	0.98†	0.91*	0.91*	0.87*	0.87*	0.86*	0.89‡	0.91*	0.92†	1.01	0.89‡	0.92†	0.92†
0.80 QS	0.66	0.96‡	0.98†	1.00	0.98‡	0.92*	0.93*	0.81*	0.84*	0.84*	0.86*	0.92*	0.91‡	1.02	0.81*	0.83*	0.85*
0.90 QS	0.43	0.97‡	0.99	1.00	0.98†	0.92*	0.91*	0.82*	0.85*	0.84*	0.85*	0.92*	0.86*	0.99	0.77*	0.78*	0.82*
0.95 QS	0.27	0.96	0.96	0.97	0.96†	0.89*	0.92‡	0.76*	0.82*	0.83*	0.84*	0.88*	0.79*	0.92‡	0.71*	0.78*	0.77*
CRPS-C	0.26	0.94‡	0.95*	0.97*	0.97*	0.92*	0.92*	0.88*	0.88*	0.90‡	0.91‡	0.92*	0.94†	0.97	0.90‡	0.93†	0.93
CRPS-L	0.39	0.97	0.96‡	0.98*	0.98*	0.97	0.96†	0.89‡	0.90‡	0.91	0.91	0.95	0.96	0.92†	0.90†	0.93	0.94
CRPS-R	0.42	0.95*	0.95*	0.98‡	0.98*	0.91*	0.91*	0.86*	0.87*	0.87*	0.89*	0.91*	0.92‡	1.00	0.87*	0.89‡	0.90‡
h=4																	
0.05 QS	0.20	1.11	1.10	1.02	0.99	1.10	0.89	0.78	0.96	1.05	1.02	1.13	1.02	0.70†	0.97	1.05	0.99
0.10 QS	0.31	1.05	1.07	1.03	1.00	1.05	0.96	0.92	0.98	1.01	1.05	1.07	0.91	0.88	0.98	1.05	0.99
0.20 QS	0.47	0.93	0.94	0.94†	0.95†	0.95	0.93	0.99	0.98	1.10	1.11	0.92	0.85	0.89	0.99	1.03	0.97
0.30 QS	0.57	0.90	0.90	0.95‡	0.97*	0.89	0.95	0.95	1.00	1.10	1.12	0.88	0.88	0.95	1.00	1.03	0.97
0.40 QS	0.63	0.90‡	0.91†	0.96‡	0.97‡	0.87†	0.90	0.88	0.99	1.06	1.08	0.87†	0.87†	0.98	0.98	0.99	0.94
0.50 QS	0.66	0.92*	0.91*	0.95‡	0.96*	0.88‡	0.93	0.87	0.96	1.04	1.05	0.88*	0.88‡	0.99	0.94	0.94	0.93
0.60 QS	0.63	0.94*	0.92*	0.96*	0.97*	0.89*	0.94	0.87	0.96	1.00	0.98	0.91*	0.88‡	0.97	0.94	0.93	0.92
0.70 QS	0.57	0.95†	0.93‡	0.96‡	0.97‡	0.90*	0.94	0.92	0.95	0.95	0.95	0.93*	0.89‡	0.95†	0.93	0.92	0.92
0.80 QS	0.45	0.97‡	0.97‡	0.99	0.98†	0.95	0.93	0.94	0.95	0.96	0.95	0.97‡	0.91‡	0.99	0.93	0.95	0.95
0.90 QS	0.30	1.04	1.02	1.02	1.01	1.01	0.88‡	0.87	0.88†	0.92	0.92	1.04	0.93‡	1.05	0.86*	0.90‡	0.91†
0.95 QS	0.18	1.08	1.07	1.04	1.02	1.04	0.95	0.85	0.85‡	0.93	0.91	1.09	0.93‡	1.07	0.82*	0.81*	0.83*
CRPS-C	0.20	0.93†	0.93†	0.96‡	0.97*	0.90	0.93	0.91	0.97	1.03	1.03	0.91†	0.88†	0.96	0.96	0.97	0.94
CRPS-L	0.31	0.94	0.95	0.96‡	0.97‡	0.93	0.94	0.92	0.98	1.06	1.07	0.93	0.88	0.93	0.98	1.00	0.96
CRPS-R	0.31	0.96†	0.95†	0.98†	0.98‡	0.92	0.92	0.90	0.95	0.98	0.98	0.94*	0.89‡	0.99	0.93	0.94	0.93

Note: Comparison of tail forecast accuracy of QR-ridge, QR-avg, PQR, and BQR specifications relative to baseline. The baseline is the QR specification for the small GDP variable set. For specifications other than QR, values below 1 indicate improvement over baseline. For each of the small GDP, large GDP, and very large GDP applications, the best-performing specification for a given forecast horizon and metric is indicated in blue. Evaluation window from 1985:Q1 through 2019:Q4. *, ‡, and † denote statistical significance at, respectively, the 1 percent, 5 percent, or 10 percent level, based on one-sided Diebold-Mariano-West tests.

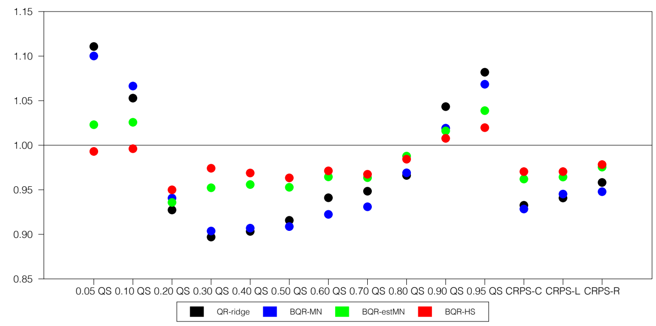
Table 4: Unemployment and inflation applications

Metric	h=1							h=4						
	QR	QR ridge	QR avg	PQR	BQR MN	BQR estMN	BQR HS	QR	QR ridge	QR avg	PQR	BQR MN	BQR estMN	BQR HS
							Unemployment							
0.05 QS	0.02	0.85‡	0.89†	0.81*	0.84‡	0.89‡	0.88†	0.06	0.79†	0.83†	0.80‡	0.83*	0.92*	0.94‡
0.10 QS	0.04	0.86*	0.88*	0.85*	0.89*	0.92*	0.89*	0.11	0.80‡	0.83‡	0.75*	0.87*	0.94‡	0.93
0.20 QS	0.06	0.93	0.95	0.94	0.94†	0.97	0.95	0.18	0.82‡	0.86‡	0.88	0.90‡	0.95‡	0.95†
0.30 QS	0.07	0.94	0.96	0.93	0.95	0.97†	0.96	0.24	0.87†	0.89†	0.90†	0.92*	0.98†	0.97
0.40 QS	0.08	0.94	0.97	0.90‡	0.94†	0.96†	0.96	0.28	0.86†	0.91	0.88‡	0.91‡	0.97*	0.97
0.50 QS	0.09	0.97	0.98	0.92	0.93†	0.95‡	0.96	0.31	0.87	0.95	0.87†	0.91‡	0.97†	0.99
0.60 QS	0.09	0.97	0.98	0.91	0.92	0.95†	0.95	0.33	0.89	0.95	0.85†	0.90†	0.97†	0.99
0.70 QS	0.08	1.00	1.00	0.92	0.93	0.97	0.97	0.33	0.93	0.98	0.82	0.91	0.97	0.98
0.80 QS	0.08	0.97	1.00	0.92	0.89‡	0.94*	0.95†	0.31	0.91	1.02	0.70*	0.93	0.97‡	0.96
0.90 QS	0.05	1.04	1.07	0.91	0.93	0.95‡	0.97	0.26	0.82	0.93	0.67‡	0.78†	0.90‡	0.93†
0.95 QS	0.03	1.13	0.96	0.98	0.90	0.91*	0.90‡	0.20	0.80	0.92	0.52‡	0.75	0.90‡	0.92
CRPS-C	0.03	0.96	0.98	0.92	0.93†	0.96‡	0.96	0.10	0.88	0.94	0.83‡	0.91‡	0.96*	0.97
CRPS-L	0.04	0.94	0.95	0.91‡	0.93‡	0.96‡	0.95	0.13	0.85†	0.90†	0.85*	0.90*	0.96*	0.96‡
CRPS-R	0.05	0.99	1.00	0.91	0.92†	0.95*	0.96	0.19	0.88	0.96	0.77‡	0.88†	0.95‡	0.96
							Inflation							
0.05 QS	0.09	0.91	0.90†	0.93	0.89‡	0.92‡	0.93†	0.07	1.18	0.86	1.42	0.91	1.02	1.03
0.10 QS	0.14	0.92	0.92‡	0.99	0.91‡	0.95‡	0.94†	0.14	0.88	0.71	1.07	0.75	0.88	0.92
0.20 QS	0.21	0.92‡	0.95‡	0.97	0.92*	0.95*	0.96‡	0.20	0.77	0.84	0.94	0.80†	0.94†	0.97
0.30 QS	0.25	0.94*	0.96*	0.98	0.95*	0.97*	0.98‡	0.25	0.68†	0.79	0.82	0.81‡	0.93‡	0.94
0.40 QS	0.27	0.97†	0.98†	0.96†	0.97*	0.98‡	0.99	0.27	0.71†	0.72‡	0.81†	0.84‡	0.96*	0.91
0.50 QS	0.29	0.96‡	0.97‡	0.94*	0.97*	0.97*	0.97*	0.29	0.69‡	0.72*	0.77†	0.84‡	0.96†	0.92
0.60 QS	0.29	0.95‡	0.97†	0.93*	0.96‡	0.97‡	0.97†	0.30	0.66*	0.68*	0.76†	0.83*	0.95*	0.91‡
0.70 QS	0.26	0.94*	0.97†	0.94‡	0.96‡	0.97‡	0.96‡	0.28	0.67*	0.67*	0.78‡	0.83*	0.95*	0.92‡
0.80 QS	0.22	0.90*	0.93*	0.90*	0.94*	0.95*	0.95*	0.24	0.71†	0.76‡	0.87	0.86‡	0.98	0.96
0.90 QS	0.15	0.87‡	0.89†	0.87‡	0.90‡	0.93‡	0.94†	0.19	0.77	0.79‡	0.81‡	0.84‡	0.96‡	0.95†
0.95 QS	0.12	0.73*	0.77*	0.76*	0.77*	0.86*	0.85*	0.15	0.74	0.84	0.76	0.83†	0.90‡	0.89‡
CRPS-C	0.09	0.94*	0.96*	0.94*	0.95*	0.97*	0.97*	0.09	0.70‡	0.73‡	0.82	0.83‡	0.95*	0.93
CRPS-L	0.14	0.94*	0.95*	0.96†	0.94*	0.96*	0.96*	0.14	0.74	0.75†	0.88	0.81	0.93†	0.93
CRPS-R	0.14	0.92*	0.95*	0.92*	0.95*	0.96*	0.96*	0.16	0.71‡	0.73*	0.81†	0.84‡	0.96*	0.93†

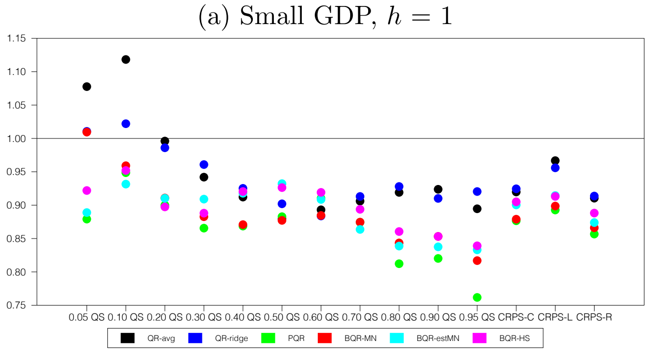
Note: Comparison of tail forecast accuracy of QR-ridge, QR-avg, PQR, and BQR specifications relative to baseline. The baseline is the QR specification for each application. For specifications other than QR, values below 1 indicate improvement over baseline. For each application, the best-performing specification for a given forecast horizon and metric is indicated in blue. Evaluation window from 1985:Q1 through 2019:Q4. *, ‡, and † denote statistical significance at, respectively, the 1 percent, 5 percent, or 10 percent level, based on one-sided Diebold-Mariano-West tests.



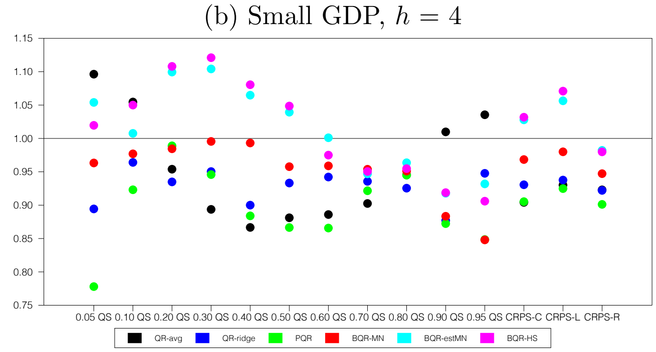
(a) Small GDP, $h = 1$



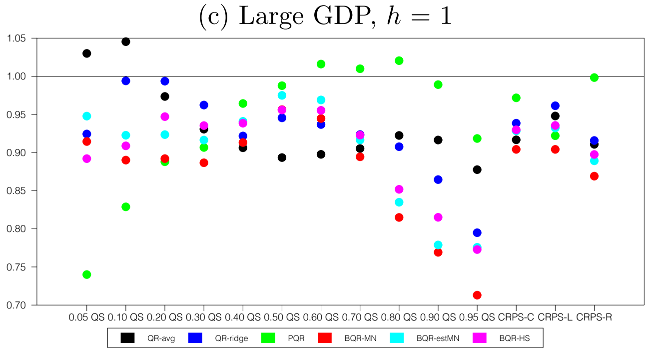
(b) Small GDP, $h = 4$



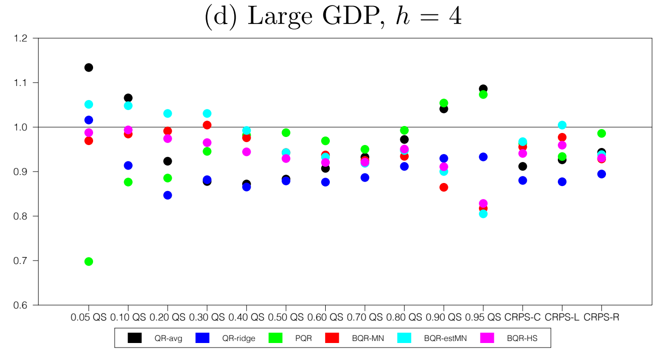
(c) Large GDP, $h = 1$



(d) Large GDP, $h = 4$

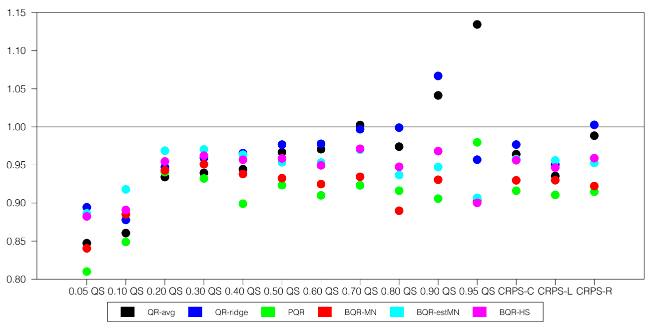


(e) Very large GDP, $h = 1$

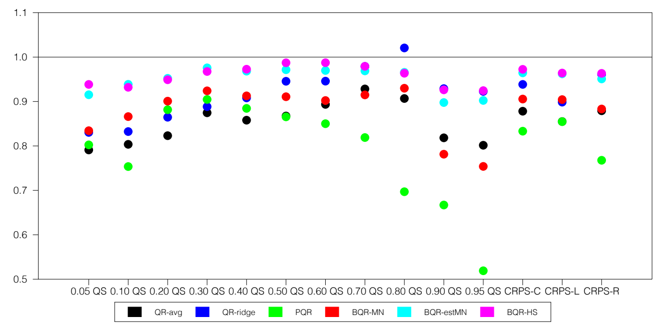


(f) Very large GDP, $h = 4$

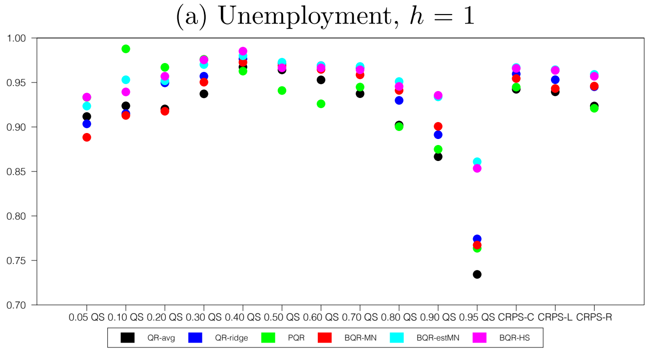
Figure 1: GDP applications: Comparison of tail forecast accuracy of QR-avg, QR-ridge, PQR, and BQR specifications relative to baseline. The baseline is the QR specification for the small GDP variable set. Values below 1 indicate improvement over the baseline. Evaluation window from 1985:Q1 through 2019:Q4.



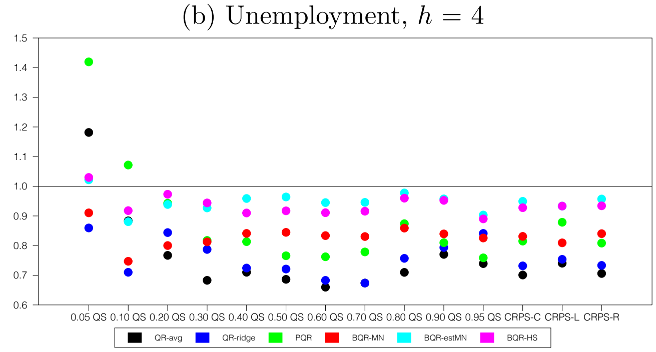
(a) Unemployment, $h = 1$



(b) Unemployment, $h = 4$

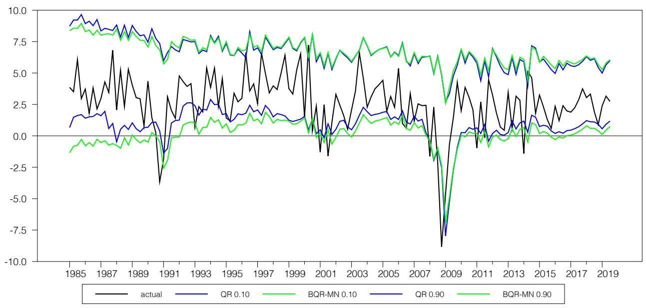


(c) Inflation, $h = 1$

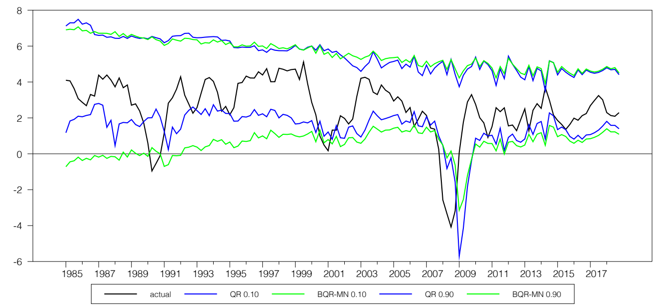


(d) Inflation, $h = 4$

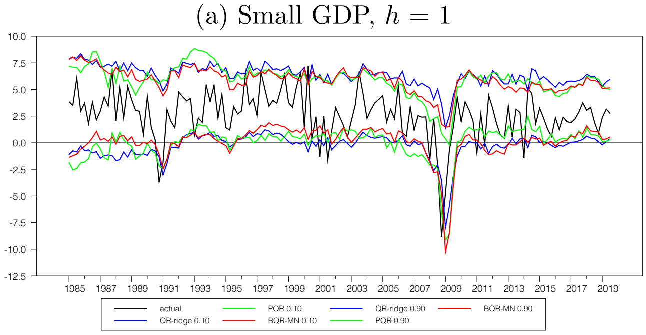
Figure 2: Unemployment and inflation applications: Comparison of tail forecast accuracy of QR-avg, QR-ridge, PQR, and BQR specifications relative to baseline. The baseline is the QR specification for each application. Values below 1 indicate improvement over the baseline. Evaluation window from 1985:Q1 through 2019:Q4.



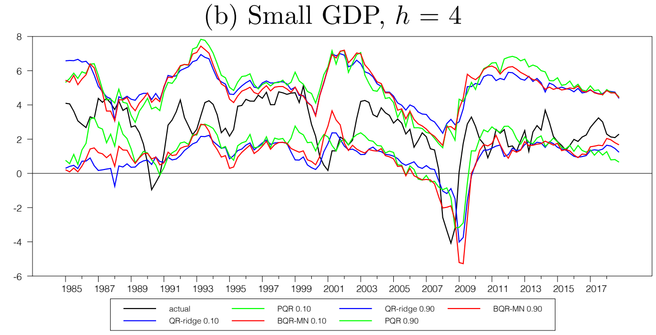
(a) Small GDP, $h = 1$



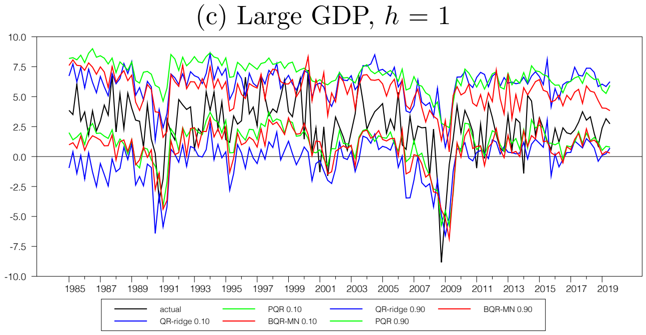
(b) Small GDP, $h = 4$



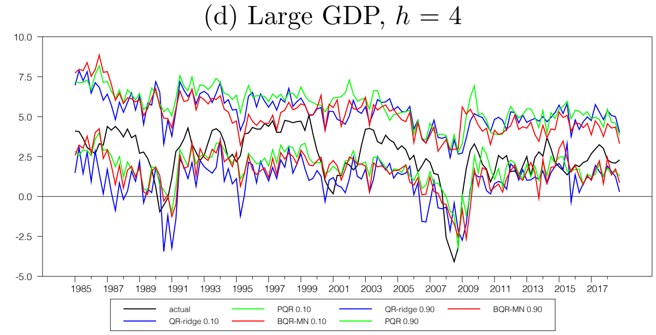
(c) Large GDP, $h = 1$



(d) Large GDP, $h = 4$

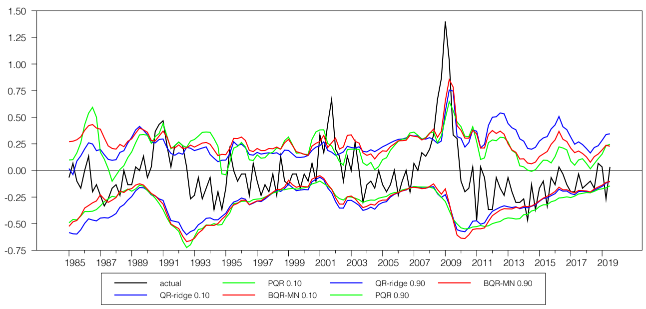


(e) Very large GDP, $h = 1$

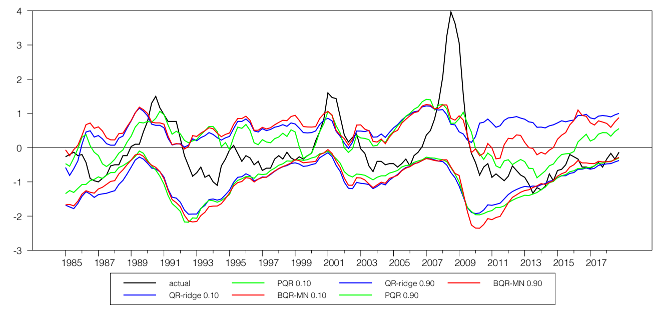


(f) Very large GDP, $h = 4$

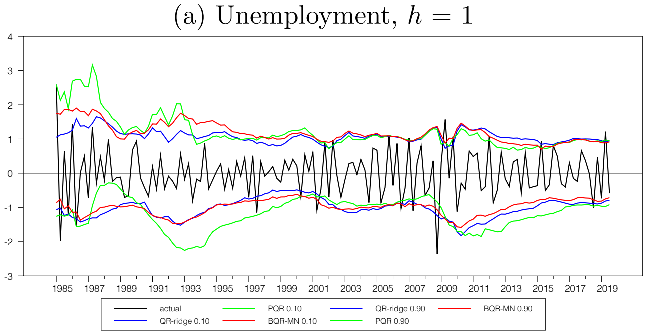
Figure 3: GDP applications: Time series of selected forecasts for quantiles of $\tau = 0.10, 0.90$.



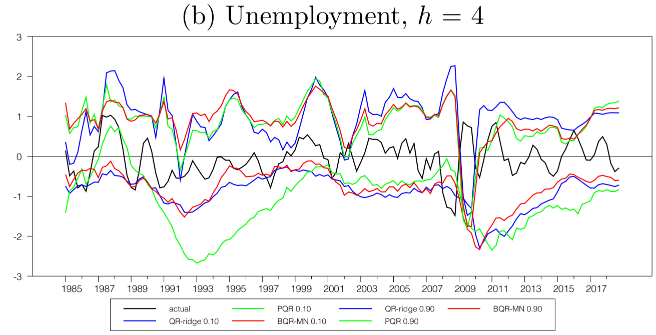
(a) Unemployment, $h = 1$



(b) Unemployment, $h = 4$

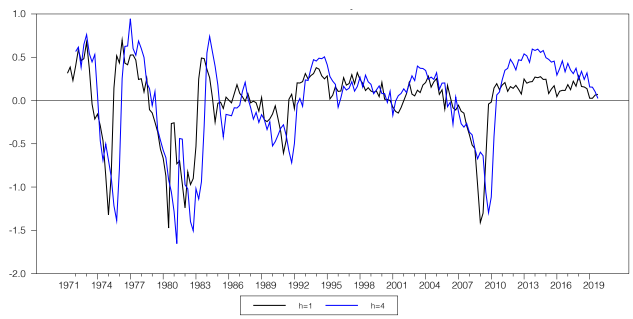


(c) Inflation, $h = 1$

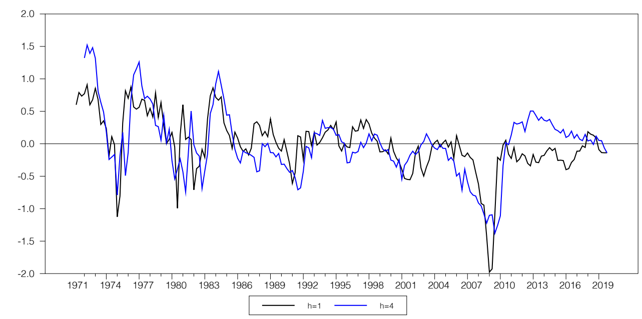


(d) Inflation, $h = 4$

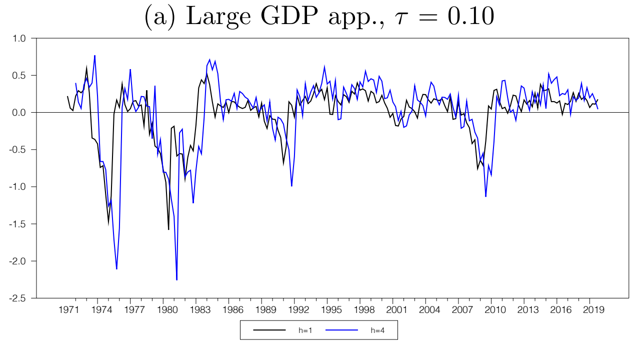
Figure 4: Unemployment and inflation applications: Time series of selected forecasts for quantiles of $\tau = 0.10, 0.90$.



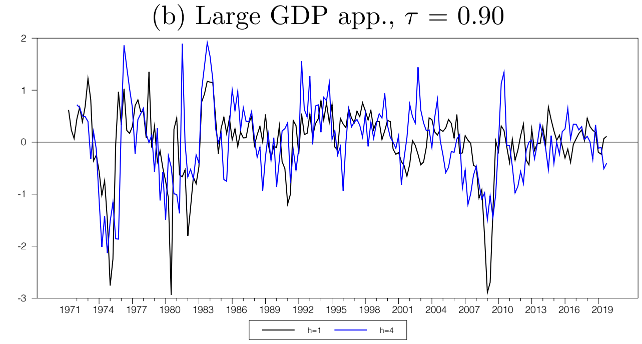
(a) Large GDP app., $\tau = 0.10$



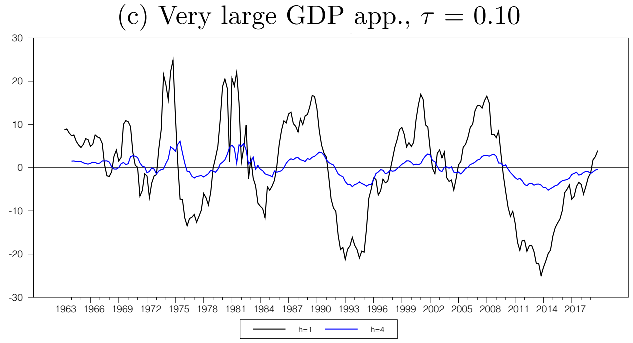
(b) Large GDP app., $\tau = 0.90$



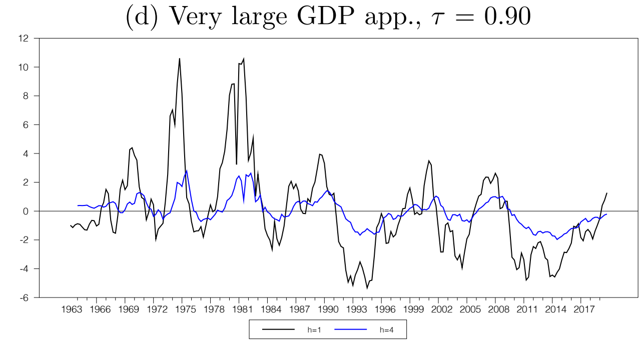
(c) Very large GDP app., $\tau = 0.10$



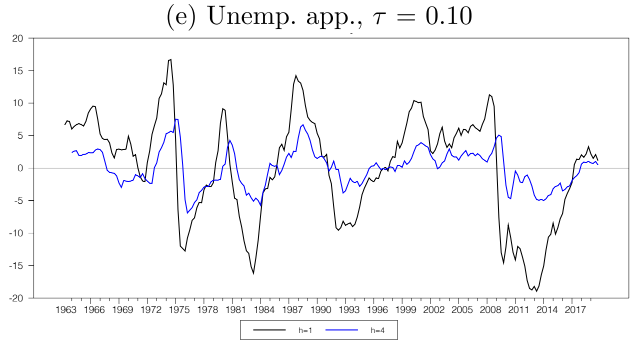
(d) Very large GDP app., $\tau = 0.90$



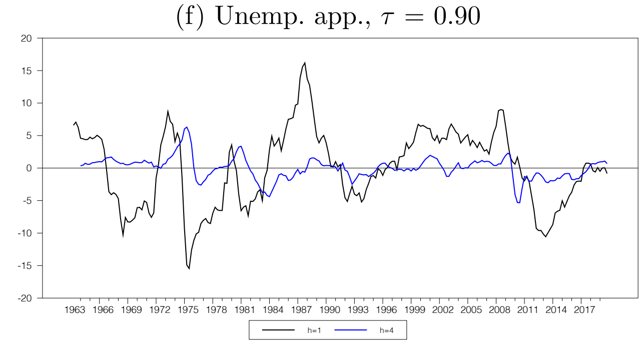
(e) Unemp. app., $\tau = 0.10$



(f) Unemp. app., $\tau = 0.90$



(g) Inflation app., $\tau = 0.10$



(h) Inflation app., $\tau = 0.90$

Figure 5: PQR factors estimated with data through 2019:Q4, selected quantiles.