Re-examining the predictive power of the yield curve with quantile regression*

Rafael B. De Rezende† Mauro S. Ferreira‡

February 12, 2019

Abstract

We use quantile regression to re-examine the term spread - future GDP growth relationship in the US. Structural breaks at upper percentiles of conditional GDP growth and changes in the relationship towards longer horizons suggest the Fed started to respond tougher and in greater advance towards inflationary pressures arising from excessive growth after the mid-1980s. In addition, the term spread has been consistently correlated to negative and low conditional GDP growth, which explains its success in predicting recessions. We also use quantile regression to forecast GDP growth and recessions probabilities in an out-of-sample scheme. In several occasions, the term spread delivered more accurate forecasts against the benchmark and professional forecasters. Quantile models were responsible for most of this success. The predictive power of the yield curve remains.

Keywords: Quantile regression; term spread; GDP growth; recessions; monetary policy shifts; inflation; forecasting.

JEL Classifications: E32; E37; E43; E44

*We are grateful to Magnus Dahlquist, Refet Gürkaynak, Jonathan Wright and Enrique Sentana as well as seminar participants at the Stockholm School of Economics, the Latin American Meeting of the Econometric Society 2014, the European Meeting of the Econometric Society 2012, the North American Summer Meeting of the Econometric Society 2012, the Meeting of the Brazilian Econometric Society 2013 and the Brazilian Time Series and Econometrics School 2013 for comments and suggestions. Rafael B. De Rezende kindly thanks the Swedish Bank Research Foundation (BFI) for financial support.

†Monetary Analysis, Bank of England, Threadneedle St, London EC2R 8AH. Email: rafael.rezende@bankofengland.gsi.gov.uk

‡Federal University of Minas Gerais, Department of Economics. Email: mferreira@cedeplar.ufmg.br

1 Introduction

The relationship between economic activity and the term structure of interest rates has been widely studied over the last decades. Several studies have concentrated on the predictive power of the difference between long and short yields (the term spread or yield curve slope) regarding future recessions (Estrella and Hardouvelis, 1991; Estrella and Mishkin, 1998; Bernard and Gerlach, 1998; Wright, 2006; among others). Others have verified whether the term spread informs about future GDP growth (Laurent, 1988; Harvey, 1988, 1989; Estrella and Hardouvelis, 1991; among others).

Despite some evidence that movements in bond risk premia may be linked to economic activity (Ludvigson and Ng, 2009; Rudebusch and Swanson, 2012; Joslin, Priebsch and Singleton, 2014), the rationale behind the predictive ability of the term spread rests mainly on the forward looking behavior of market participants that anticipate future reactions of the central bank (Ang, Piazzesi and Wei, 2006; Rudebusch, Sack and Swanson, 2007). As an example, a likely future recession implies the central bank will aggressively reduce interest rates to counteract GDP contractions and disinflationary pressures. By anticipating such scenario, current long rates become smaller than short ones, resulting in a negative spread in the present.

Guided by this intuition, it is therefore not surprising that the yield curve slope has been largely used as a leading recession indicator. Supporting this relevance, several works (Estrella and Hardouvelis, 1991; Estrella and Mishkin, 1998; Bernard and Gerlach, 1998; Wright, 2006; Rudebusch and Williams, 2009) have verified that the term spread has been consistently related to odds of recessions over subsequent quarters, with every recession being preceded by an yield curve inversion since the mid-1960s. In addition, the yield spread has participated as an important input of various leading indicators, including the one from the Conference Board and also from the OECD.

Given this performance, it is puzzling that the term spread ability to predict economic activity beyond recessions is not a consensus. For instance, several works have verified a weakening in the ability of the term spread to predict GDP growth around the mid-1980s. Stock and Watson (2003) find that the term spread is the most successful predictor of GDP growth during 1971-1984, but the predictive power disappears from 1985. Estrella, Rodrigues and Schich (2003) find a structural break in GDP growth regressions in 1983. Giacomini and Rossi (2006) also find breakdowns in the relationship during the Burns-Miller and Volker monetary policy regimes. Other works, on the other hand, have reassured the ability of the spread to anticipate economic growth. For instance, Aguiar-Conraria, Martins and Soares (2012) verified a weakening in the spread - GDP growth relationship in 1985, but the predictability ressurges in the 1990s. Chinn and Kucko (2010) finds that the ability of the term spread to predict 1 year ahead in the US has decreased after 1998, but they observed significance in the 2 years ahead predictions.

We consider these controversies puzzling. Is it possible to explain the apparent higher stability of the term spread - recession relationship? Moreover, what are the fundamental reasons behind the weakening of GDP growth regressions verified around the mid-1980s? Our aim is to shed more light to this debate. More specifically, we examine the predictive ability of the term spread throughout the whole conditional distribution of GDP growth using quantile regression methods, which allowed us to verify several new results.

First, results indicate the existence of asymmetries between our variables of interest. More specifically, regardless of the sample period, the term spread is found to be more consistently correlated to negative and lower conditional GDP growth than to intermediate and high growth, corroborating previous studies documenting the success of the spread to forecast recessions. Moreover, we verify changes in the predictive relationship towards longer horizons after the mid-1980s and identify structural breaks happening mostly in the second quarter of 1984. The main novelty, however, is that these breaks were concentrated at higher conditional percentiles of GDP growth. This is consistent with a view that the Fed started to respond tougher and in greater advance to inflationary pressures resulted from excessive growth around mid-1980s.

Accordingly, a large part of a literature focused on studying changes in the conduct of monetary policy after the Volcker's presidency describes the Fed being tougher to inflation. The findings by Boivin, Hiley, and Mishkin (2010) describing that a more intense and faster reaction of output gap to movements in short-rates was the pattern during 1962-1979, but a less intense, slower and more persistent reaction was observed from 1984 to 2008 encounter support in our results. In addition, Kim and Nelson (2006) suggests that this change may be due to the Fed being more responsive to inflation during this period, which is also supported by our findings. Similar conclusions were also reached by Clarida, Galí and Gertler (2000), Galí, Lopez-Salido, and Valles (2003), Boivin and Giannoni (2006) and Galí and Gambetti (2009).

Our results also validate the use of the entire sample to estimate probability of future recessions, given the absence of breaks in low percentile equations. This explains why binary models focused on estimating probabilities of future recessions conditioned on current spread have been more stable over time, as documented by Estrella, Rodrigues, and Schich (2003) and emphasized by Rudebusch and Williams (2009).

Motivated by the existence of conditional asymmetries, we innovated and also forecasted recessions probabilities from empirical distributions for GDP growth estimated using quantile regressions. This approach performed better than standard probit models. More important, it presented superior forecasts than other benchmarks, including professional forecasters.

Since conditional asymmetries can potentially bias OLS estimators and affect point forecasts, we also re-evaluated the performance of the spread in forecasting GDP growth out-of-sample using the conditional median, as it is a more robust estimator. The forecasts were carried using final revised and real time data considering a full sample - from 1955 to 2011 - and a commonly used subsample after the break, from 1985 to 2011. Anticipating, the median models delivered superior forecasts than a naive autoregressive model and the Survey of the Professional Forecasters (SPF), changing the conclusions one would normally reach if relying on OLS models. This result was particularly stronger for models that incorporated nonlinearities through a quadratic specification.

To sum up, besides the new findings suggesting that the changes in monetary policy affected higher conditional percentiles of the term spread - GDP growth relationship, our results also indicate that the spread has become more related, after 1984, to longer run economic activity, losing its ability to signal about GDP growth in the short run. This is consistent with a more forward looking monetary policy after the mid-1980s. Two other novelties of our work were the use of empirical densities obtained from quantile regressions to forecast probabilities of recessions and the use of median estimators to infer about future output growth. In general, quantile methods delivered more accurate forecasts than competitors in both exercises. Overall, results indicate that the predicitive power of the yield curve remains.

The remaining of the paper is organized as follows. In the next section we revisit the term spread - GDP growth relationship using quantile regressions; the third section presents the out-of-sample forecasting framework; the fourth section describes the forecasting results; and the last section concludes.

2 The term spread - GDP growth relationship

In this section we revisit the term spread - GDP growth relationship. Our main variable of interest is the quarterly (annualized) real GDP growth between quarters t and $t-1$, which we denote by y_t and is computed according to

$$y_t = 400 \times [log(GDP_t) - log(GDP_{t-1})] \tag{1}$$

Following the literature, real GDP growth can be predicted by estimating the following equation by OLS

$$y_t = \alpha_0 + \alpha' X_{t-h} + u_t \tag{2}$$

where u is a random shock, α is a $k \times 1$ vector of coefficients and X_{t-h} is a $k \times 1$ vector of covariates, which include the term spread, but may also include other variables such as lagged GDP growth that is commonly used to deal with autocorrelation in y_t.[1]

In our specifications we also allow for a quadratic term spread variable. Its inclusion is motivated

[1] Stock and Watson (2003), Hamilton and Kim (2002) and Ang, Piazzesi and Wei (2006) are among some authors that also use lagged GDP growth in their forecasting equations.

by previous findings, such as those of Galbraith and Tkacz (2000), Venetis, Paya and Peel (2003), and Duarte, Venetis and Paya (2005) revealing that the intensity of the positive relation between GDP growth and the term spread diminishes if the former crosses a certain threshold. The quadratic spread variable is used to approximate such nonlinearity, as suggested by Koenker (2005) in a general context of quantile regression. When modeling y_t, we thus allow for four different specifications: (i) $X_{t-h} = (Spread_{t-h})$, implying in $\alpha = \alpha_1$, (ii) $X_{t-h} = (Spread_{t-h}, Spread^2_{t-h})$, so that $\alpha = (\alpha_1, \alpha_2)$, (iii) $X_{t-h} = (Spread_{t-h}, y_{t-h})$ with $\alpha = (\alpha_1, \alpha_3)$, and (iv) a specification with all covariates included, $X_{t-h} = (Spread_{t-h}, Spread^2_{t-h}, y_{t-h})$.[2] To be consistent with threshold models, one should expect $\alpha_1 > 0$ and $\alpha_2 < 0$, resulting in a concave relation between the spread and future GDP growth.

While OLS regression models are often used to approximate the conditional mean of y_t, quantile regression models are suitable to approximate its conditional quantiles. Following the linear specification in (2) and denoting $Q_u(\tau)$ as the τ-th quantile of u, the τ-th quantile of y_t, conditioned on X_{t-h}, can be obtained by

$$Q_{y_t}(\tau | X_{t-h}) = \alpha_0(\tau) + \alpha(\tau)' X_{t-h} \tag{3}$$

where, $\alpha_0(\tau) = \alpha_0 + Q_u(\tau)$ and $\alpha(\tau)$ can be estimated according to Koenker and Basset (1978). In particular, the conditional median of y_t is approximated after setting $\tau = 0.5$.

2.1 Mean and median regressions

We collect yield data from the FRED database. As in the extant literature the term spread is computed as the difference between the 10-year Treasury Bond and the 3-month Treasury Bill interest rates. The quarterly frequency is obtained by averaging the monthly spread over each quarter. For the in-sample analyses performed in this section we use real GDP data collected from the Fed of Philadelphia. More specifically, we use the vintage that became available on the third quarter of 2011 and we decided to work with two samples normally used in this literature: a full sample ranging from 1955Q1 to 2011Q2, and a subsample ranging from 1985Q1 to 2011Q2. This subsample coincides with the Great Moderation period and is justified by several works reporting a breakdown in the predictive ability of the term spread after the mid-1980s (Estrella, Rodrigues, and Schich, 2003; Stock and Watson, 2003; Giacomini and Rossi, 2006).

In Table 1 we report predictive results for OLS and QR(0.5) regressions. In order to save space, we only show results for specifications (iii) and (iv), which incorporate y_{t-h} as a covariate. As commonly found in the literature the spread keeps a positive relation with future GDP growth. Different patterns, however, arise depending on the sample. For example, significance is observed

[2] We also tested specifications with the Federal Funds rate or the 3-month Treasury Bill rate as regressors. The estimated coefficients on both variables showed very small values and no significance was achieved for any forecast horizon. Moreover, the regressions showed poor performance in the out-of-sample forecasting exercise.

from $h = 1$ to $h = 6$ in the full sample (1955-2011), and from $h = 4$ to $h = 8$ in the subsample (1985-2011). In addition, magnitude of the coefficients also indicates that changes have occurred over time. Coefficients linking the spread to future GDP growth are larger for horizons up to four quarters ahead in the full sample, while the largest coefficients are found for longer horizons when using post-1985 data. These results, which are robust to estimation method and whether we include or not the quadratic spread variable, indicate that the spread started signaling about future economic activity in greater advance after the mid-1980s.

Results also indicate the existence of a nonlinear relationship between our variables of interest. Similar to Galbraith and Tkacz (2000), Venetis, Paya and Peel (2003), and Duarte, Venetis and Paya (2005), we verify a concave relationship with $\alpha_1 > 0$ and $\alpha_2 < 0$, indicating that the positive relation loses strength after a certain level of the spread is crossed. Most quadratic models have superior fitting than their linear counterparts as indicated by smaller Akaike Information Criterion (AIC) statistics.

2.2 A more complete analysis: the conditional distribution of GDP growth

Analyzing the conditional distribution of GDP growth across percentiles provides a more complete picture of its relation with the term spread. In each box of Figure 1 we plot quantile regression estimates using the two samples considered previously. The first row brings coefficient estimates of the spread in the standard linear model, while the second and third rows present estimates in a quadratic model. In both cases we allowed for the presence of lagged growth among the regressors. Figure 2 shows the model's goodness of fit by presenting the adjusted $R1(\tau)$ statistics of Koenker and Machado (1999) computed across percentiles.[3]

Results in Figure 1 reinforce previous findings. The term spread keeps a strong relation with future output growth and this is shown over almost the entire conditional distribution of the latter variable. In addition, for horizons up to $h = 4$, we verify that the spread is significantly related to larger ranges of the conditional distribution of GDP growth when using the full sample. Contrary, for the period 1985-2011, statistical significance is mostly observed across percentiles from $h = 4$ to $h = 8$. These patterns occur in the linear and in the quadratic models, but are stronger once we deal with nonlinearity. Results in Figure 2 reinforce these conclusions. In addition, they indicate the ability of the spread to predict a large portion of the conditional distribution of economic activity with greater advance when using post-1985 data. Observe that, for this subsample, highest values for $\overline{R1}(\tau)$ were obtained exactly for $h = 8$.

Figures 1 and 2 also reveal the existence of a strong asymmetric relationship between our

[3]The $\overline{R1}(\tau)$ statistic of Koenker and Machado (1999) is similar in nature to the R^2 statistic, as it informs (for each percentile) the goodness of a fit of a model compared to a regression on the intercept $\alpha_0(\tau)$. The closer $\overline{R1}(\tau)$ is to 1, the better the adjustment.

variables of interest. While the spread is more correlated to negative and low GDP growth, it informs little about high conditional output growth. This result is present in all specifications and explains why, regardless the sample, the spread has successfully signalled about recession probabilities.

Since our exercises consider a subsample starting at 1985, it seems natural to formally verify if this is a reasonable splitting. This was done using the test statistics SQ and DQ proposed by Qu (2008) and Oka and Qu (2011). Table 2 shows results for the statistic SQ, which tests the null hypothesis of no break in a specific quantile equation. Table 3 shows results for the statistic DQ, which tests the null of no break in multiple quantiles. SQ(1) and DQ(1) are used to test for the absence of only one break, while SQ(2|1) and DQ(2|1) test for the absence of a second break given the first one was found. In both procedures we allowed for a maximum of two breaks.[4] When rejecting H_0, both tests also estimate the break date.

Table 2 reveals that most of the detected breaks happened at higher percentiles. The smallest percentile in which breaks were detected was $\tau = 0.35$ when forecasting 4 quarters ahead with the linear model. For the linear model with $h = 8$ and the quadratic with $h = 4$, breaks were found only in regressions for $\tau \geq 0.65$. In the case of the quadratic model and $h = 8$, breaks were verified for $\tau \geq 0.5$. In line with the existing literature, most of the estimated break dates occurred in the mid-1980s. Out of the 19 breaks found, 9 were in 1984Q2 and 1 in 1984Q1. The test that searches for breaks in multiple equations rejected the absence of one break in the linear specification for $h = 4, 8$, and also in the quadratic for $h = 4$. These breaks were all detected to have occurred in 1984Q2.

Results suggest the spread started to signal about future economic activity in a different manner after 1984. Besides indicating a change in the predictive ability of the spread towards longer horizons, this shift happened for higher conditional GDP growth, which is consistent with a market view that the Fed started to respond tougher and in greater advance to inflationary pressures arising from excessive growth. Additionally, the fact that breaks were not observed at lower percentile equations suggest that the conduct of monetary policy was not expected to be modified in the presence of future adverse cycles.

Accordingly, a large part of the literature focused on changes in the conduct of monetary policy after the Volcker's presidency describes the Fed being tougher to inflation. The findings by Boivin, Hiley, and Mishkin (2010) describing that a more intense and faster reaction of output gap to movements in short-rates was the pattern during 1962-1979, but a less intense, slower and more persistent reaction was observed from 1984 to 2008 encounter support in our results. In addition, Kim and Nelson (2006) suggests that this change may be due to the Fed being more responsive to inflation during this period, which is also supported by our findings. Similar conclusions were also

[4]Following the previous analyses, these tests were conducted using regressions that allowed for the presence of lagged GDP growth among covariates.

reached by Clarida, Galí and Gertler (2000), Galí, Lopez-Salido, and Valles (2003), Boivin and Giannoni (2006) and Galí and Gambetti (2009).

Our results also validate the use of the entire sample to estimate probability of future recessions, given the absence of breaks in low percentile equations. This explains why binary models focused on estimating probabilities of future recessions conditioned on current spread have been more stable over time, as found by Estrella, Rodrigues, and Schich (2003) and emphasized by Rudebusch and Williams (2009). It seems, however, that similar exercise to estimate probability of booms, or even intermediate growth, would be more problematic since breaks were identified at intermediate and higher percentiles.

3 Out-of-sample forecasting

Based on our previous findings, we also verify the forecasting ability of the term spread in an out-of-sample framework. Having identified the existence of conditional asymmetries in the GDP growth - term spread relationship, it seems natural to rely on empirical probability distributions to compute probabilities of future recessions. Contrasting with most of the works in this literature that relies on standard probit regressions (assuming conditional normality of the GDP growth), we estimate these probabilities from quantile regressions. A potential advantage of the latter is its flexibility, as it may accommodate unusual conditional distributions. Additionally, since conditional asymmetries can potentially bias OLS estimators and affect point forecasts, we also re-evaluate GDP growth forecasting regressions using the conditional median, as it is a more robust estimator.

Given the evidence of breaks only in mid and high conditional percentiles, we forecast recessions using the full sample, but conduct output growth predictions using post-1985 data as well.

We conduct forecasting analyses using real-time and final revised real GDP data collected from the Fed of Philadelphia. For the real-time data we decided to work with the *advance* data, which is the Bureau of Economic Analysis' (BEA) first estimate for the previous quarter. While advance data are subject to greater measurement error, their use makes more sense as we are also willing to address comparisons to forecasts reported by professional forecasters in the Survey of Professional Forecasters (SPF henceforward), who know the advance data when submitting their projections.[5]

The forecasts presented in this work will be contrasted to the final revised data consisting of the vintage that became available on the third quarter of 2011. At this quarter the BEA released an extensive revision that incorporated important methodological changes in the way it estimates national accounts, so we consider this vintage a good proxy for the final revised data.

[5]This happens because the Fed of Philadelphia sends its survey questionnaires right after the advance estimates become public. Since these questionnaires must be returned within the following two or three weeks, which is before the release of the next revision, professional forecasters' information set, when submitting their projections, include the advance data.

3.1 Real GDP growth

Real GDP growth point forecasts can be readily computed from prediction models (2) and (3). More specifically, given the OLS estimates $\hat{\alpha}_0$ and $\hat{\alpha}$, mean forecasts of y_t in $t+h$, conditioned on X_t, can be computed as

$$E\left(y_{t+h}|X_t\right) = \hat{\alpha}_0 + \hat{\alpha}' X_t \tag{4}$$

Similarly, median forecasts of y_t in $t+h$ can be obtained using

$$\hat{Q}_{y_{t+h}}\left(0.5|X_t\right) = \hat{\alpha}_0\left(0.5\right) + \hat{\alpha}\left(0.5\right)' X_t \tag{5}$$

where $\hat{\alpha}_0(\tau)$ and $\hat{\alpha}(\tau)$ are estimated according to Koenker and Basset (1978). In this paper, OLS and QR(0.5) forecasts are confronted against each other, against mean and median forecasts reported in the SPF, and also against a direct autoregressive model, $\hat{y}_{t+h} = \hat{\alpha}_0 + \hat{\alpha}_1 y_t$.

3.2 Recessions

When it comes to forecasting recessions, we first have to define what a recession is in our study. We follow Rudebusch and Williams (2009) and use the following rule linking real GDP changes to recessions: the economy is in recession at quarter t, implying $R_t = 1$, if a negative quarterly real GDP growth ($y_t < 0$) is observed; $R_t = 0$, otherwise. The reason for not using the NBER recession series is that NBER recessions are observed ex-post, meaning that it is not suitable for our real-time exercise. Nevertheless, our rule produces 22 recessionary quarters that match the 34 NBER recession quarters in the period 1955Q1-2011Q2, with only 11 false alarms.[6]

Given each forecast $\hat{Q}_{y_{t+h}}(\tau|X_t)$, $\forall\tau \in [\underline{\tau}, \overline{\tau}]$, the probability of a recession in $t+h$ computed using quantile regressions can then be obtained as follows

$$\hat{P}_t^{QR}(R_{t+h} = 1) = \hat{P}_t^{QR}(y_{t+h} \leq 0) = sup\left\{\tau \in [\underline{\tau}, \overline{\tau}] : \hat{Q}_{y_{t+h}}(\tau|X_t) \leq 0\right\}$$

When $\hat{Q}_{y_{t+h}}(\tau|X_t) > 0$, $\forall\tau \in [\underline{\tau}, \overline{\tau}]$, the model indicates $\hat{P}_t^{QR}(R_{t+h} = 1) = \hat{P}_t^{QR}(y_{t+h} \leq 0) = 0$.[7] In order to estimate these probabilities, we set $\tau \in [\underline{\tau} = 0.0025, \overline{\tau} = 0.9975]$ and allow for our four sets of covariates. In addition, we avoid crossing of $\hat{Q}_{y_{t+h}}(\tau|X_t)$ across percentiles through the "rearrangement" procedure of Chernozhukov, Fernandez-Val and Galichon (2010).

[6]As stated by the NBER (2003): *"The NBER considers real GDP to be the single measure that comes closest to capturing what it means by 'aggregate economic activity'. The [NBER] therefore places considerable weight on real GDP and other output measures".*

[7]Here we define any event in which $y_t \leqslant 0$ as a recession. The use of the operator $\leq$ instead of $<$ does not really make any difference in our results.

As it is the standard procedure in this literature, we also forecast recession probabilities from Probit models,

$$\hat{P}_t^{PR}\left(R_{t+h}=1\right)=\Phi\left(\hat{\theta}_0+\hat{\theta}'X_t\right) \tag{6}$$

where Φ is the standard normal cumulative distribution function.

QR forecasts are compared to Probit forecasts, SPF recession forecasts and to a simple probit model, $\hat{P}_t^{PR}\left(R_{t+h}=1\right)=\Phi\left(\hat{\theta}_0+\hat{\theta}_1 y_t\right)$.

3.3 Assessing the accuracy of forecasts

Real GDP growth point forecasts are compared using the Root Mean Squared Forecast Error (RMSFE) measure. When comparing recessions probabilities we follow Diebold and Rudebusch (1989) and employ two different measures: the Root Quadratic Probability Score (RQPS) and the Log Probability Score (LPS). Letting T be the number of out-of-sample forecasts, and i a particular model, these two last score measures are respectively computed as

$$RQPS^i=\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[\hat{P}_t^i\left(R_{t+h}=1\right)-R_{t+h}\right]^2}$$

$$LPS^i=-\frac{1}{T}\sum_{t=1}^{T}\left[\left(1-R_{t+h}\right)ln\left(1-\hat{P}_t^i\left(R_{t+h}=1\right)\right)+R_{t+h}ln\left(\hat{P}_t^i\left(R_{t+h}=1\right)\right)\right]$$

A more rigorous comparison between each model, however, can be assessed by relying on the Harvey, Leybourne and Newbold (1997) test (HLN henceforward), which allows verifying if the difference between the average forecast errors of two competing models is statistically significant. HLN test is based on a modification of the Diebold and Mariano (DM, 1995) statistic that corrects for its tendency to be over-sized in finite samples.[8] Through Monte Carlo experiments, the authors show that the modified statistic performs considerably better than the original one, providing important size corrections.

Although the HLN test was originally designed for non-nested models, Clark and McCracken (2012), when analysing the use of alternative HAC estimators in nested models, find that comparing the HLN test statistic to standard normal critical values delivers a test that has size fairly close to the nominal level in both population and finite-samples. In this study, we thus follow Harvey, Leybourne and Newbold (1997) approach.[9] As suggested by these authors, we compare the HLN statistic to critical values of a Student's t-distribution with $T-1$ degrees of freedom (t_{T-1})[10].

[8] The adjustment consists of estimating the variance of the mean loss differentials using the rectangular lag window with a lag truncation parameter equal to $h-1$ and then multiplying the DM statistic by $\sqrt{\left(T+1-2h+T^{-1}h\left(h-1\right)\right)/T}$.

[9] This test is also used by Faust and Wright (2013) when forecasting inflation.

[10] Our assessments indicated that the t_{T-1} distribution delivered results that are a bit more conservative than a stardard

In order to pursue such analysis, we rely on $D^{i,j}_{t+h}$, which measures the (loss) forecasting error differential between models i and j at $t+h$. Sample loss differentials are computed according to RMSFE, RQPS and LPS using the following equations:

$$\hat{D}^{i,j}_{t+h}(MSFE) = \left(\hat{y}^{i}_{t+h} - y_{t+h}\right)^2 - \left(\hat{y}^{j}_{t+h} - y_{t+h}\right)^2$$

$$\hat{D}^{i,j}_{t+h}(QPS) = \left(\hat{P}^{i}_{t}(R_{t+h}=1) - R_{t+h}\right)^2 - \left(\hat{P}^{j}_{t}(R_{t+h}=1) - R_{t+h}\right)^2$$

$$\begin{aligned}\hat{D}^{i,j}_{t+h}(LPS) = -\left(1-R_{t+h}\right)\left[ln\left(1-\hat{P}^{i}_{t}(R_{t+h}=1)\right) - ln\left(1-\hat{P}^{j}_{t}(R_{t+h}=1)\right)\right] - \\ -R_{t+h}\left[ln\left(\hat{P}^{i}_{t}(R_{t+h}=1)\right) - ln\left(\hat{P}^{j}_{t}(R_{t+h}=1)\right)\right]\end{aligned}$$

Models i and j perform equally if the mean loss differential is zero; $H_0 : E\left(D^{i,j}_{t}\right) = 0,\ \forall t$. As we are interested in verifying whether model i delivers more accurate out of sample forecasts than model j, we state the alternative as $H_1 : E\left(D^{i,j}_{t}\right) < 0,\ \forall t$.[11]

It is important to mention that the HLN test is not suitable for real-time forecasting. Clark and McCracken (2009) proposed an alternative test of equal predictive accuracy for real-time data, the construction of which requires further assumptions on the nature of the data revisions and evidence that these assumptions are met in the real-time data. Even if their assumptions were satisfied, we would still not be able to use their test, since it is not designed for recursive forecasts, which is the strategy we adopt. As an alternative we apply the accuracy tests on ex-post (our *final*) revised data, but also report tests' results for the real-time forecasts.

4 Forecasting results

Tables 4, 5, and 6 summarize our findings. The numbers reported in tables 4 and 5 correspond to RMSFE ratios against a direct AR benchmark. In table 6 we report RQPS and LPS ratios against a simple probit model. Values smaller (larger) than 1 indicate numerically superior (inferior) forecasts over benchmarks. Symbols ○ and • indicate statistical significance against the benchmark and against SPF forecasts, respectively. We also verify the superiority of quadratic specification against similar linear model, which is indicated by †, while asterisks refer to superiority of median forecasts against those carried by the OLS estimator.

normal also when comparing nested models.

[11] The HLN test is not available for a RMSFE (RQPS) loss function, so we report the results from the MSFE (QPS) version of the test.

4.1 GDP Growth

Full sample

Table 4 shows results using the full sample. The forecasting exercise is carried according to the recursive method, meaning that each forecast model is re-estimated after incorporating the newest information. For real-time data, we used the most recent vintage available in each recursion. The first estimation window ranges from 1955Q1 to 1985Q4 when using final revised data, and 1955Q1 to 1986Q1 for real-time data.

Table 4 reveals that the *spread* models started to deliver smaller RMSFEs than either the benchmark autoregressive model or the SPF from $h = 4$, with stronger results found for $h = 5$. Statistical superiority against the benchmark was observed in the following situations: QR4 model for $h = 4$; QR2 model for $h = 5$; and OLS4 for $h = 6$. Only quadratic specifications were able to improve upon naïve AR forecasts, which suggests that some of the previous findings about the inability of the *spread* to beat simple autoregressive models may have been a result of mispecification. This perception is reinforced by the statistical superiority of quadratic specifications against their linear counterparts, which was largely observed from $h = 1$ to $h = 5$.

Another result is the higher statistical accuracy against SPF when $h = 5$, which is observed only when relying on median estimates: models QR1 and QR2. Notice, however, that recursively testing for superiority of the spread models against SPF reveals that higher accuracy was also observed for other horizons during several periods. This result is shown by Figure 3, where we plot, in each graph, RMSFEs computed recursively for all forecasting strategies. Asterisks indicate superiority against the SPF at 10% level.[12] For instance, notice that the QR2 model was more accurate than the SPF from 2003 until the beginning of the recession in 2008 when forecasting 3 quarters ahead. When forecasting 5 quarters ahead, three quadratic specifications (OLS2, OLS4, QR2) and one linear (QR1) were statistically more accurate than SPF from 1997. The end of the statistical superiority of the OLS models coincides with the end of the great recession (in 2010), whereas quantile models continue to deliver superior results.

Two main conclusions emerge. First, the use of a more robust estimator (median) improves forecasting performance. The inability of OLS spread regressions to beat benchmarks should have already raised concerns about how results could have been changed if robust methods, like conditional median, were adopted. Second, ignoring the nonlinearity captured by the quadratic term deteriorates the forecasting performances of models. All together, had quadratic models with median estimator been previously considered, conclusions regarding the forecasting power of the spread would have been different.

[12] The first set of recursive forecasts used for testing ranges from 1986Q1 to 1990Q1.

Post-1985 sample

Motivated by our findings described in Section 2 we also examine the forecasting performance of models using the sub-sample 1985Q1-2011Q2.[13] This period, referred to as "Great Moderation", has been characterized by an increased stability of various macroeconomic variables.Several authors have verified a decline in the predictive power of the spread during this period (Stock and Watson, 2003; Estrella, Rodrigues, and Schich, 2003; D'Agostino, Giannone and Surico, 2006).[14]

Results are shown in Table 5, from which we verify smaller RMSFE ratios against the benchmark for $h \geq 4$. Significance, however, was only achieved for $h = 8$, with QR1 being the only linear model that was able to beat the benchmark in statistical terms. Notice also that, at this same horizon, all quadratic specifications (OLS2, OLS4, QR2, and QR4) delivered superior forecasts than their linear counterparts. Another result is the inability of the spread models to statistically improve upon SPF, even though smaller RMSFEs were observed for $h = 4$ and $h = 5$.

Despite of observing, for most horizons and specifications, higher accuracy of conditional median over conditional mean, statistical superiority was not systematically verified in the post-1985 sample.[15] This contrasts to the superior performance of QR over OLS models in the full sample, which may have occurred due to the fact that the median estimator was less affected by the presence of the detected breaks around 1984. Indeed, one should expect conditional means to be more severely impacted by such breaks.

From our results, the documented breakdown in the predictive performance of the term spread around the mid-1980s (Stock and Watson, 2003; Estrella, Rodrigues and Schich, 2003; D'Agostino, Giannone and Surico, 2006 among others), needs qualification. As we have previously noticed when analyzing the significance of conditional quantile coefficients, a breakdown indeed happened in forecasting equations. But, on the other hand, and in line with the conduct of a more forward looking monetary policy, the yield spread gained power to forecast economic activity longer in the future, from 6 to 8 quarters ahead. What has been disputed as a breakdown has actually configured a change on how market participants priced the Treasuries in the presence of a more forward looking Fed.

[13]The first estimation window ranges from 1985Q1 to 1991Q4.

[14]Notice that we consider here the post-2007 period part of the "great moderation". As Clark (2009) points out, the higher volatility of several macro variables during the recent crisis was mostly driven by temporary large shocks to oil prices and financial markets, which did not mark the end of the great moderation period. In addition, despite having the Fed switching its main instrument of monetary policy from short-term policy rate to large-scale asset purchases since the crisis, the use of such unconventional monetary policy has not changed its degree of transparency and credibility, which are associated to the macroeconomic stability observed since the mid-1980s.

[15]Statistical superiority of QR over similar OLS specification was verified only for QR2 at $h = 8$. However, in 28 out of the 56 specification/horizon combinations (68%) median forecasts resulted in smaller RMFSE than OLS.

4.2 Probability of recession

We now evaluate estimates of recessions probabilities. Since we have documented different patterns across conditional quantiles, we found reasonable to compute probabilities using empirical probability distributions obtained from quantile forecasts. We also report estimates from probit models, which is the standard procedure in this literature.

To minimize potential instabilities in extreme quantile regressions that may occur due to paucity and sparsity of data at tails (He, 1997; Wang, Li and He, 2012) the out of sample probability exercise was only carried using the full sample. Since breaks were mainly detected in upper percentiles, the use of the whole sample for estimation may not be problematic given that we are focusing on lower percentiles. As in the GDP growth forecasting exercise, the first window to estimate parameters is 1955Q1-1985Q4 for final revised data, and 1955Q1-1986Q1 for real-time data, with models being re-estimated recursively upon the arrival of new information.

Table 6 reports RQPS and LPS of each model relative to probabilities inferred by a simple probit with y_{t-h} as the unique regressor. Some results are noteworthy. First, spread models were statistically more accurate than the benchmark at all horizons from $h = 3$, whether using RQPS or LPS, and whether relying on final revised or real time GDP data. In addition, even though statistical superiority was not observed, quantile models still delivered smaller RQPS and LPS against probits for almost all horizons and regression specifications, indicating advantage in relying on the flexibility of quantile models for forecasting recessions.

We can also see that the term spread delivers statistically more accurate forecasts than professional forecasters for $h = 4, 5$. This is true for probit and quantile models. When these tests were carried recursively we also verified statistical superiority over SPF for $h = 3$ from 1990 to 2009. This result can be observed in Figure 4, where we plot recursive RQPS and LPS statistics for quantile regression models, standard probit AR and the SPF. Asterisks indicate statistical improvement upon SPF. Figure 4 also shows that spread models have been delivering statistical superior probability estimates than SPF since 1990 when $h = 5$. In the case of $h = 2$, superiority was found for a long period of time although results were not statistically significant.

In line with Rudebusch and Williams (2009), who conducted similar analyses using a simple probit model, our results indicate important information content of spread models for forecasting recessions. As shown, this is also true when forecasting GDP growth, reinforcing the continuing predictive power of the yield spread.

A natural question that arises is about the performance of our quantile models towards the great recession of 2008/2009. Figure 5 provides a good visualization of this exercise by showing GDP growth quantile forecasts in real time for the period 2003Q1-2011Q2. Given the found similarity between different specifications, we only show results generated when $X_t = (spread_t, spread_t^2)$.

We observe higher probability of negative output growth before and during the recession. For

$h = 2$ the model misses the recession a little by anticipating it, while for $h = 4, 6, 8$ it certainly hits the NBER recession with high accuracy, indicating that the spread based QR model would have been able to warn about risks of a recession in 2008/2009 with at least two years in advance.

Another interesting feature observed in Figure 5 is that while forecasts of high GDP growth barely modify, the same is not true for lower conditional output growth. This result is basically capturing a conditional asymmetric behavior in the term spread - GDP growth relationship. This asymmetry reveals a type of heterokedasticity that is not normally identified in standard time series models.

5 Conclusions

The use quantile regression allowed us to improve the understanding of the relationship between interest rate term spread and future output growth in the US.

We evidenced changes in the predictive relationship that are consistent with shifts in the conduction of monetary policy after the presidency of Volcker, regarded as more forward looking and tougher to inflation. Structural breaks were detected mostly in 1984, at intermediate and higher conditional percentiles of GDP growth, which is consistent with the view that monetary policy shifted in the direction of being less tolerant to inflationary pressures resulted from excessive output growth. In addition, we verified that the yield spread has become more informative about economic activity further in the future, but less so in the short run.

We were also able to understand an apparent puzzle according to which the term spread has not changed its ability to anticipate recessions despite the weakening in term spread - GDP growth regressions verified after the mid-1980s. According to our results, recession probability models based on the yield spread have not been affected by shifts in monetary policy as they model the lower tail of the conditional GDP growth distribution, while changes following Volcker have mostly affected intermediate and higher percentiles.

These findings have similar implications for the out of sample forecasting of GDP growth. Through a sample ranging from 1955 to 2011, the term spread was statistically superior than an autoregressive benchmark and professional forecasters at mid forecast horizons. When using post-1985 data, statistical superiority was observed for longer horizons, which is consistent with a more forward looking Fed. We also documented the ability of the spread to signal about probabilities of future recessions when compared to different competitors. In both exercises, quantile models were responsible for most of the success.

All together, our findings suggest that the predictive power of the yield curve remains.

References

[1] Aguiar-Conraria, L., Martins, M. M., and Soares, M. J. (2012), “The yield curve and the macro-economy across time and frequencies”, *Journal of Economics Dynamics and Control*, 36(12), pp. 1950-1970.

[2] Ang, A., Piazzesi, M., and Wei, M. (2006), “What does the yield curve tell us about GDP growth?”, *Journal of Econometrics* , 131(1), pp. 359-403.

[3] Bernanke, B. (2004). “The Great Moderation”. Speech at the Meeting of the Eastern Economic Association, Washington DC, February 20.

[4] Bernard, H. and Gerlach, S. (1998). “Does the Term Structure Predict Recessions? The International Evidence.” International Journal of Finance and Economics, 3(3), pp. 195-215.

[5] Boivin, J., & Giannoni, M. (2006), “DSGE models in a data-rich environment”. National Bureau of Economic Research, WP 12772.

[6] Boivin, J., Kiley, M. T., and Mishkin, F. S. (2010), “How has the monetary transmission mechanism evolved over time?”. National Bureau of Economic Research, WP 15879.

[7] Cheznozhukov, V., Fernandez-Val, I. and Galichon, A. (2010). “Quantile and Probability Curves Without Crossing”. Econometrica, 78(3), pp. 1093-1125.

[8] Chinn, M. and Kucko, K. (2010). “The Predictive Power of the Yield Curve across Countries and Time”. National Bureau of Economic Research, WP 16398.

[9] Christensen, R., Lopez, J. and Rudebusch, G. (2011). “Extracting deflation probability forecasts from treasury yields”, FRBSF Working Paper 2011-10.

[10] Clarida, R., Galí, J. and Gertler, M. (2000). "Monetary Policy Rules And Macroeconomic Stability: Evidence And Some Theory", The Quarterly Journal of Economics, 115(1), pp. 147-180.

[11] Clark, T. (2009). “Is the Great Moderation Over? An Empirical Analysis”. Economic Review, Federal Reserve Bank of Kansas City, pp. 5-42

[12] Clark, T. and McCracken (2012). “Advances in Forecast Evaluation”. Working Paper Series 2011-025B, Federal Reserve Bank of St. Louis

[13] D’Agostino, A., Gianonne, D. and Surico, P. (2006). “(Un)Predictability and Macroeconomic Stability”. Working Paper Series 605, European Central Bank.

[14] Diebold, F. X. and Mariano, R. S. (1995). “Comparing Forecast Accuracy”. Journal of Business and Economics Statistics, (13), pp. 253-265.

[15] Diebold, F. X., and Rudebusch, G. D. (1989). “Scoring the Leading Indicators”. Journal of Business, (62), pp. 369-391.

[16] Duarte, A., Venetis, I. A. and Paya, I. (2005) “Predicting Real Growth and the Probability of Recession in the Euro Area Using the Yield Spread.” International Journal of Forecasting, 21(2), pp. 262-77.

[17] Estrella, A. and Hardouvelis, G. A. (1991). “The Term Structure as a Predictor of Real Economic Activity.” Journal of Finance, 46(2), pp. 555-76.

[18] Estrella, A. and Mishkin, F. S. (1998). “Predicting U.S. Recessions: Financial Variables as Leading Indicators.” Review of Economics and Statistics, 80(1), pp. 45-61.

[19] Estrella, A., Rodrigues, A. P. and Schich, S. (2003). “How Stable Is the Predictive Power of the Yield Curve? Evidence from Germany and the United States.” Review of Economics and Statistics, 85(3), pp. 629-44.

[20] Faust J. and Wright, J. (2013). “Forecasting Inflation” Handbook of economic forecasting, 2 (part A)

[21] Gaglianone, W. and Lima, L. R. (2012). “Constructing Density Forecasts from Quantile Regressions”. Journal of Money, Credit and Banking, 44(8), pp. 1589-1607.

[22] Galbraith, J. W. and Tkacz, G. (2000). “Testing for Asymmetry in the Link Between the Yield Spread and Output in the G-7 Countries.” Journal of International Money and Finance, 19(5), pp. 657-72.

[23] Galí, J., and Gambetti, L. (2009), “On the sources of the great moderation”, *American Economic Journal: Macroeconomics*, 1(1), pp. 26-57.

[24] Galí, J., Lopez-Salido, J. D., and Valles, J. (2003). “Technology shocks and monetary policy: assessing the Fed’s performance”, *Journal of Monetary Economics*, 50(4), pp. 723-743.

[25] Giacomini, R. and Rossi, B. (2006). “How Stable Is the Forecasting Performance of the Yield Curve for Output Growth?” Oxford Bulletin of Economics and Statistics, 68(Suppl. 1), pp. 783-95.

[26] Hamilton, J. and Kim, D. H. (2002). "A Reexamination of the Predictability of Economic Activity Using the Yield Spread". Journal of Money, Credit and Banking, 34(2), pp. 340-360.

[27] Harvey, C. R. (1988). “The Real Term Structure and Consumption Growth,” Journal of Financial Economics, 22(2), pp. 305-33.

[28] Harvey, C. R. (1989). “Forecasts of Economic Growth From the Bond and Stock Markets.” Financial Analysts Journal, 45(5), pp. 38-45.

[29] Harvey, D. I., Leybourne, S. J., and Newbold, P. (1997). “Testing the equality of prediction mean squared errors”. International Journal of Forecasting (13), 281–291.

[30] Haubrich, J.G., Dombrosky, A.M. (1996), “Predicting Real Growth Using the Yield Curve", *Economic Review,* 32, 26-34.

[31] Joslin, S., Priebsch, M. and Singleton, K. (2012). “Risk Premiums in Dynamic Term Structure Models with Unspanned Macro Risks”. Working Paper, Graduate School of Business, Stanford University.

[32] Kim, C-L., and Nelson, C.R. (2006), “Estimation of a Forward-Looking Monetary Policy Rule: A Time-Varying Parameter Model using Ex-Post Data”, *Journal of Monetary Economics*, 53(8), 1949-1966.

[33] Kitsul, Y. and Wright, J. H. (2012). “The economics of options-implied inflation probability density functions”. Journal of Financial Economics.

[34] Koenker, R. (2005). “Quantile Regression”. Econometric Society Monographs No. 38, Cambridge University Press.

[35] Koenker, R. and Bassett, G. (1978). “Regression Quantiles,” Econometrica, 46(1), pp. 33–50.

[36] Koenker, R. and Hallock, K. F. (2001). “Quantile Regression”. Journal of Economic Perspectives, 15(4), pp. 143–156.

[37] Koenker, R. and Machado, J.A.F. (1999). “Goodness of fit and related inference processes for quantile regression”. Journal of the American Statistical Association (94), pp. 1296–1310.

[38] Koenker, R. and Xiao, Z. (2004) “Unit root quantile autoregression inference”. Journal of the American Statistical Association, 99(467), pp. 775-787.

[39] Koenker, R. and Xiao, Z. (2006) “Quantile Autoregression”. Journal of the American Statistical Association, 101(475), pp. 980-990.

[40] Laurent, R. D. (1988). “An Interest Rate-Based Indicator of Monetary Policy.” Federal Reserve Bank of Chicago Economic Perspectives, 12(1), pp. 3-14.

[41] Ludvigson, S.C. and Ng, S. (2009). “Macro factors in bond risk premia". Review of Financial Studies, v. 22, p. 5027-5067.

[42] National Bureau of Economic Research (2003). “The NBER’s Business-Cycle Dating Procedure”, unpublished manuscript.

[43] Oka, T. and Qu, Z. (2011). “Estimating Structural Changes in Regression Quantiles”. Journal of Econometrics, 162, pp. 248-267.

[44] Qu, Z. (2008). “Testing for Structural Change in Regression Quantiles”. Journal of Econometrics, 148, pp. 170-184.

[45] Rudebusch, G., Sack, B. and Swanson, E. (2007). "Macroeconomic implications of changes in the term premium," Review, Federal Reserve Bank of St. Louis, pp. 241-270.

[46] Rudebusch, G. and Swanson, E. (2012). “The Bond Premium in a DSGE Model with Long-Run Real and Nominal Risks”. American Economic Journal: Macroeconomics, v. 4, pp. 105-143.

[47] Rudebusch, G. and Williams, J. (2009). “Forecasting Recessions: The Puzzle of the Enduring Power of the Yield Curve”. Journal of Business and Economic Statistics, 27(4), pp. 492-503.

[48] Rudebusch, G. and Wu, T. (2007). “Accounting for a Shift in Term Structure Behavior with No-Arbitrage and Macro-Finance Models”. Journal of Money, Credit and Banking, 39(2-3), pp. 395-422.

[49] Stock, J. H. and Watson, M. W. (2003). “Forecasting Output and Inflation: The Role of Asset Prices.” Journal of Economic Literature, 41(3), pp. 788-829.

[50] Venetis, I. A., Paya, I. and Peel, D. A. (2003). “Re-Examination of the Predictability of Economic Activity Using the Yield Spread: A Nonlinear Approach.” International Review of Economics and Finance, 12(2), pp. 187-207.

[51] Wheelock, D. and Wohar, M. (2009). “Can the Term Spread Predict Output Growth and Recessions? A Survey of the Literature”. Federal Reserve Bank of St. Louis Review, 91(5, Part 1), pp. 419-40.

[52] Wright, J. H. (2006). “The Yield Curve and Predicting Recessions.” Finance and Economics Discussion Series No. 2006-07, Federal Reserve Board of Governors, www.federalreserve.gov/pubs/feds/2006/200607/200607pap.pdf.

Table 1: OLS and QR(0.5) regressions results

Notes: Standard errors (in parentheses) for OLS coefficients are computed by the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator of Andrews and Monahan (1992). Standard errors for QR are computed by paired bootstrap using 2000 replications. $\star$ denotes statistical significance at (***)1%, (**)5% and (*)10%, respectively.

Sample		OLS				QR(0.5)			
		$spread_{t-h}$	$spread^2_{t-h}$	y_{t-h}	AIC	$spread_{t-h}$	$spread^2_{t-h}$	y_{t-h}	AIC
1955 − 2011	$h=1$	0.476**		0.335***	1196.9	0.199		0.295***	1183.3
		(0.196)		(0.075)		(0.156)		(0.103)	
		1.204**	−0.311***	0.325***	1191.3	1.758**	−0.484**	0.231***	1175.6
		(0.333)	(0.104)	(0.080)		(0.722)	(0.210)	(0.089)	
	$h=2$	0.719***		0.194**	1202.7	0.537**		0.178**	1183.5
		(0.217)		(0.083)		(0.212)		(0.071)	
		1.825***	−0.476***	0.181**	1187.6	1.984***	−0.528***	0.161***	1162.1
		(0.279)	(0.092)	(0.076)		(0.415)	(0.117)	(0.055)	
	$h=4$	0.688***		−0.007	1205.8	0.382**		−0.010	1186.0
		(0.255)		(0.079)		(0.198)		(0.071)	
		1.350**	−0.286	−0.015	1202.2	2.101***	−0.514***	−0.020	1171.8
		(0.621)	(0.201)	(0.078)		(0.601)	(0.166)	(0.174)	
	$h=6$	0.419**		−0.034	1203.0	0.255		−0.019	1179.5
		(0.199)		(0.074)		(0.161)		(0.063)	
		0.646**	−0.100	−0.037	1204.4	0.832*	−0.223	−0.041	1179.6
		(0.304)	(0.106)	(0.075)		(0.440)	(0.149)	(0.060)	
	$h=8$	0.171		−0.076	1194.9	0.060		−0.028	1169.7
		(0.276)		(0.081)		(0.209)		(0.079)	
		0.446	−0.124	−0.080	1195.9	0.730	−0.236	−0.072	1169.1
		(0.651)	(0.202)	(0.080)		(0.664)	(0.196)	(0.073)	
1985 − 2011	$h=1$	0.061		0.482***	479.8	0.020		0.307***	469.6
		(0.135)		(0.152)		(0.177)		(0.116)	
		0.366	−0.091	0.484***	481.6	−0.414	0.128	0.303**	471.4
		(0.580)	(0.161)	(0.142)		(0.946)	(0.251)	(0.120)	
	$h=2$	0.173		0.417***	486.9	0.056		0.300**	469.7
		(0.158)		(0.097)		(0.206)		(0.116)	
		0.481	−0.092	0.420***	488.7	1.011	−0.255	0.300***	470.3
		(0.623)	(0.184)	(0.092)		(0.819)	(0.213)	(0.106)	
	$h=4$	0.441*		0.146	500.6	0.293		0.038	480.0
		(0.262)		(0.114)		(0.234)		(0.090)	
		1.807***	−0.409**	0.157	498.3	1.570**	−0.398**	0.057	474.3
		(0.629)	(0.191)	(0.113)		(0.076)	(0.199)	(0.104)	
	$h=6$	0.713**		0.051	495.6	0.453**		0.023	476.6
		(0.354)		(0.067)		(0.210)		(0.066)	
		2.285**	−0.479*	0.055	491.7	2.180*	−0.499*	−0.039	473.9
		(1.077)	(0.269)	(0.064)		(1.139)	(0.298)	(0.059)	
	$h=8$	0.725*		0.043	495.7	0.423*		0.071	477.2
		(0.370)		(0.077)		(0.220)		(0.075)	
		3.535**	−0.867**	0.042	477.4	1.999***	−0.511**	0.023	460.2
		(1.328)	(0.347)	(0.068)		(0.738)	(0.198)	(0.062)	

Table 2: Breaks in regression quantiles - SQ test

Notes: This table shows estimated statistics for the break test of Qu (2008). SQ(1) refers to the test statistic for the presence of 1 break in the specified quantile regression. SQ(2|1) tests for the existence of a second break given the first one was found. $\star\star$ denotes significance at the 5% level. The sample period is 1955Q1-2011Q2.

		Quantiles	0.1	0.2	0.35	0.5	0.65	0.8	0.9
Linear	$h=4$	SQ(1)	1.112	1.408	1.771**	1.610**	1.962**	1.995**	2.083**
		SQ(2|1)	–	–	1.788**	1.653**	1.793**	1.058	0.855
		Break Date			62:4,08:2	84:2,08.3	84:2,08.3	84:2	84:2
	$h=8$	SQ(1)	1.077	1.157	1.345	1.463	1.816**	1.688**	2.015**
		SQ(2|1)	–	–	–	–	1.520**	1.479	0.986
		Break Date	–	–	–	–	66:1,00:2	84:2	84:2
Quadratic	$h=4$	SQ(1)	1.079	1.160	1.173	1.262	1.821**	1.995**	2.197**
		SQ(2|1)	–	–	–	–	1.143	1.096	1.060
		Break Date	–	–	–	–	84:2	84:2	84:1
	$h=8$	SQ(1)	1.158	1.349	1.376	1.706**	1.723**	1.542	1.878**
		SQ(2|1)	–	–	–	1.200	1.994**	–	1.035
		Break Date	–	–	–	07:2	66:1,00:2	–	84:2

Table 3: Breaks in regression quantiles - DQ test

Notes: This table shows estimated statistics for the break test of Oka and Qu (2011), allowing for a maximum of two breaks. DQ(1) refers to the test statistic of 1 break in multiple regression quantiles. DQ(2|1) tests for the existence of a second break given the first was detected. $\star\star$ denotes significance at the 5% level. The sample period is 1955Q1-2011Q2.

Specification	Linear Model		Quadratic Model	
	$h=4$	$h=8$	$h=4$	$h=8$
DQ(1)	1.035**	0.918**	0.952**	0.920
DQ(2|1)	0.940	0.905	0.914	–
Break Date	84:2	84:2	84:2	–

Table 4: GDP growth forecasting

Notes: This table shows results in the form of RMSFE ratios relative to an AR model. OLS (QR) 1, OLS (QR) 2, OLS (QR) 3 and OLS (QR) 4 denote specifications (i) $X_t = Spread_t$, (ii) $X_t = (Spread_t, Spread_t^2)$, (iii) $X_t = (Spread_t, y_t)$ and (iv) $X_t = (Spread_t, Spread_t^2, y_t)$, respectively. Blue/underscored values indicate outperformance of QR models against OLS with same specification. Large numbers indicate the most accurate forecast model. $\star$ denotes significance [(***)1%, (**)5%, (*) 10%] of QR against OLS. $\circ$ denotes significance [($^{\circ\circ\circ}$)1%, ($^{\circ\circ}$)5%, ($^{\circ}$)10%] against AR. $\dagger$ denotes significance [(†††)1%, (††)5%, (†)10%] of quadratic specification against linear specification. $\bullet$ denotes significance [($^{\bullet\bullet\bullet}$)1%, ($^{\bullet\bullet}$)5%, ($^{\bullet}$)10%] against SPF.

		OLS 1	OLS 2	OLS 3	OLS 4	QR 1	QR 2	QR 3	QR 4	SPFMEAN	SPFMED
$h=1$	REALTIME	1.469	1.316†††	1.208	**1.092**†††	1.358*	1.281*††	1.134**	1.128	0.654	0.653
	FINREV	1.663	1.500†††	1.281	**1.156**†††	1.466***	1.408***†	1.251	1.243	–	–
$h=2$	REALTIME	1.414	1.204†††	1.311	1.127††	1.222***	1.150*†	1.137***	**1.076**	0.812	0.825
	FINREV	1.606	1.367†††	1.445	1.231††	1.357***	1.266***†	1.250***	**1.133*****†	–	–
$h=3$	REALTIME	1.219	1.090††	1.232	1.105††	1.079***	**1.025****	1.125**	1.088	0.886	0.931
	FINREV	1.287	1.150††	1.301	1.168††	1.118***	**1.037*****	1.160***	1.132	–	–
$h=4$	REALTIME	1.085	1.001†	1.117	1.030†	0.986**	**0.932****†	1.017**	0.955$^{\star\star\star\dagger\circ}$	0.955	0.971
	FINREV	1.158	1.073†	1.207	1.121†	1.016***	**0.935*****†	1.125**	1.030**	–	–
$h=5$	REALTIME	0.985	0.950	1.033	0.988	0.930$^{\star\star\bullet\bullet}$	**0.921**$^{\star\circ\bullet}$	0.993	0.976	0.997	0.991
	FINREV	1.022	0.978	1.103	1.050†	0.944***	**0.907**$^{\star\star\star\dagger\circ}$	1.022**	0.976***†	–	–
$h=6$	REALTIME	0.949	**0.944**	0.959	0.953°	0.962	0.977	0.972	0.974	–	–
	FINREV	0.971	0.967	0.992	0.986	**0.966**	**0.966**	0.981	0.973	–	–
$h=8$	REALTIME	**0.990**	1.011	0.995	1.007	0.996	1.028	1.028	1.043	–	–
	FINREV	0.977	0.984	1.003	1.004	0.969	**0.950***	1.022	1.013†	–	–

Table 5: GDP growth forecasting - post-1985 data

Notes: As in Table 4.

		OLS 1	OLS 2	OLS 3	OLS 4	QR 1	QR 2	QR 3	QR 4	SPFMEAN	SPFMED
$h=1$	REALTIME	1.180	1.182	1.021	1.014	1.149	1.154	1.030	1.051	0.612	0.607
	FINREV	1.328	1.363	1.037	1.056	1.347	1.363	1.160	1.191	–	–
$h=2$	REALTIME	1.139	1.156	1.052	1.055	1.085	1.103	1.026	1.039	0.794	0.799
	FINREV	1.252	1.296	1.064	1.085	1.247	1.282	1.040	1.075	–	–
$h=3$	REALTIME	1.068	1.088	1.073	1.089	1.027	1.056	1.057	1.071	0.865	0.915
	FINREV	1.015	1.026	1.101	1.103	1.028	1.008	1.082	1.001	–	–
$h=4$	REALTIME	0.996	1.010	1.033	1.028	0.932	0.974	1.016	1.011	0.899	0.902
	FINREV	1.020	1.004	1.029	1.008	1.015	0.989	1.020	0.988	–	–
$h=5$	REALTIME	1.000	1.146	0.997	1.210	0.969	1.164	0.977	1.324	0.976	0.983
	FINREV	0.963	1.033	0.999	1.108	0.953	1.090	0.977	1.117	–	–
$h=6$	REALTIME	0.963	1.070	1.013	1.103	0.928	1.097	0.968	1.106	–	–
	FINREV	0.955	1.027	1.036	1.042	0.950	1.084	1.012	1.114	–	–
$h=8$	REALTIME	0.913	$0.817^{\dagger\dagger\circ}$	0.964	$0.891^{\dagger\circ}$	0.893°	$0.774^{\star\star\dagger\dagger\circ\circ}$	0.960	$0.895^{\dagger\dagger\dagger\circ}$	–	–
	FINREV	0.921	$0.799^{\dagger\circ}$	0.983	$0.869^{\dagger}$	0.951	0.818	1.012	$0.879^{\dagger\dagger}$	–	–

Table 6: Recessions forecasting

Notes: This table shows results in the form of RQPS/LPS ratios relative to a simple probit model. PRB (QR) 1, PRB (QR) 2, PRB (QR) 3 and PRB (QR) 4 denote specifications (i) $X_t = Spread_t$, (ii) $X_t = (Spread_t, Spread_t^2)$, (iii) $X_t = (Spread_t, y_t)$ and (iv) $X_t = (Spread_t, Spread_t^2, y_t)$, respectively. Blue/underscored values indicate outperformance of QR models against Probit with same specification. Large numbers indicate the most accurate forecast model. $\star$ denotes significance [(***)1%, (**)5%, (*)10%] of QR against Probit with same specification. $\circ$ denotes significance [($^{\circ\circ\circ}$)1%, ($^{\circ\circ}$)5%, ($^{\circ}$)10%] against simple probit. $\dagger$ denotes significance [(†††)1%, (††)5%, (†)10%] of quadratic specification against linear specification. $\bullet$ denotes significance[($^{\bullet\bullet\bullet}$)1%, ($^{\bullet\bullet}$)5%, ($^{\bullet}$)10%] against SPF.

			PRB 1	PRB 2	PRB 3	PRB 4	QR 1	QR 2	QR 3	QR 4	SPF
$h=1$	RealTime	RQPS	1.421	1.458	1.344	1.382	1.442	1.439	**1.314**	1.345	0.758
		LPS	1.543	1.589	1.406	1.446	1.523	1.540	**1.353**	1.390	0.863
	FinRev	RQPS	1.154	1.196	1.055	1.102	1.165	1.172	1.037	**1.029**	–
		LPS	1.119	1.157	0.966	0.995	1.105	1.121	0.962	**0.956**	–
$h=2$	RealTime	RQPS	0.886	0.887	0.875	0.875	0.894	0.882	0.886	**0.873**	0.668
		LPS	0.847	0.839	0.799	0.791	0.804	0.795	0.783	**0.777**	0.644
	FinRev	RQPS	1.068	1.070	1.013	1.011	1.060	1.054	**0.962**	0.982	–
		LPS	1.071	1.069	0.945	0.943	1.008	0.998	**0.890**	0.906	–
$h=3$	RealTime	RQPS	0.775°	0.775°	0.774°	0.774°	0.776°	0.769°	0.783	**0.765°**	0.732
		LPS	0.746	0.748	0.742	0.744	**0.713°**	0.724°	0.716°	0.716°	0.715
	FinRev	RQPS	0.882°	0.889°	0.886$^{\circ\circ}$	0.893°	0.884°	**0.879°**	0.886$^{\circ\circ}$	0.878$^{\circ\circ}$	–
		LPS	0.860	0.867	0.857	0.863	0.836$^{\circ\circ}$	0.834$^{\circ\circ}$	0.835$^{\circ\circ}$	**0.832$^{\circ\circ}$**	–
$h=4$	RealTime	RQPS	**0.678$^{\circ\circ\bullet\bullet}$**	0.681$^{\circ\circ\bullet\bullet}$	0.684$^{\circ\circ}$	0.686$^{\circ\circ}$	0.701°	0.685$^{\circ\circ\bullet}$	0.698°	0.684$^{\circ\circ\bullet}$	0.771
		LPS	0.596$^{\circ\circ\bullet}$	0.600$^{\circ\circ}$	0.603$^{\circ\circ}$	0.606$^{\circ\circ}$	0.596$^{\circ\circ\bullet\bullet}$	0.593$^{\circ\circ\bullet}$	0.593$^{\circ\circ\bullet\bullet}$	**0.591$^{\circ\circ\bullet}$**	0.692
	FinRev	RQPS	0.832$^{\circ\circ\circ}$	0.836$^{\circ\circ\circ}$	0.855$^{\circ\circ\circ}$	0.859$^{\circ\circ\circ}$	**0.831$^{\circ\circ\circ}$**	0.835$^{\circ\circ\circ}$	0.848$^{\circ\circ\circ}$	0.845$^{\circ\circ\circ}$	–
		LPS	0.787$^{\circ\circ\circ}$	0.789$^{\circ\circ\circ}$	0.813$^{\circ\circ}$	0.815$^{\circ\circ}$	**0.770$^{\circ\circ\circ}$**	0.766$^{\circ\circ\circ}$	0.781$^{\circ\circ\circ}$	0.783$^{\circ\circ\circ}$	–
$h=5$	RealTime	RQPS	**0.670$^{\circ\circ\bullet\bullet\bullet}$**	0.679$^{\circ\circ\bullet\bullet\bullet}$	0.671$^{\circ\circ\bullet\bullet\bullet}$	0.681$^{\circ\circ\bullet\bullet\bullet}$	0.690$^{\circ\circ\bullet\bullet\bullet}$	0.679$^{\circ\circ\bullet\bullet\bullet}$	0.689$^{\circ\circ\bullet\bullet\bullet}$	0.684$^{\circ\circ\bullet\bullet\bullet}$	0.801
		LPS	**0.533$^{\circ\circ\bullet\bullet\bullet}$**	0.548$^{\circ\circ\bullet\bullet\bullet}$	0.535$^{\circ\circ\bullet\bullet\bullet}$	0.551$^{\circ\circ\bullet\bullet\bullet}$	0.545$^{\circ\circ\bullet\bullet\bullet}$	0.542$^{\circ\circ\bullet\bullet\bullet}$	0.546$^{\circ\circ\bullet\bullet\bullet}$	0.552$^{\circ\circ\bullet\bullet\bullet}$	0.664
	FinRev	RQPS	**0.811$^{\circ\circ\circ}$**	0.822$^{\circ\circ\circ}$	0.818$^{\circ\circ\circ}$	0.828$^{\circ\circ\circ}$	0.820$^{\circ\circ\circ}$	0.834$^{\circ\circ\circ}$	0.815$^{\circ\circ\circ}$	0.827$^{\circ\circ\circ}$	–
		LPS	**0.749$^{\circ\circ\circ}$**	0.764$^{\circ\circ\circ}$	0.757$^{\circ\circ\circ}$	0.774$^{\circ\circ\circ}$	0.755$^{\circ\circ\circ}$	0.763$^{\circ\circ\circ}$	0.752$^{\circ\circ\circ}$	0.760$^{\circ\circ\circ}$	–
$h=6$	RealTime	RQPS	0.694$^{\circ\circ}$	0.688$^{\circ\circ}$	0.698$^{\circ\circ}$	**0.692$^{\circ\circ}$**	0.695$^{\circ\circ}$	0.698$^{\circ\circ}$	0.693$^{\circ\circ}$	0.696$^{\circ\circ}$	–
		LPS	0.621$^{\circ\circ}$	**0.608$^{\circ\circ}$**	0.623$^{\circ\circ}$	0.611$^{\circ\circ}$	0.610$^{\circ\circ}$	0.610$^{\circ\circ}$	0.611$^{\circ\circ}$	0.609$^{\circ\circ}$	–
	FinRev	RQPS	0.832$^{\circ\circ\circ}$	0.828$^{\circ\circ\circ}$	0.834$^{\circ\circ\circ}$	0.830$^{\circ\circ\circ}$	**0.823$^{\circ\circ\circ}$**	0.830$^{\circ\circ\circ}$	0.835$^{\circ\circ\circ}$	0.838$^{\circ\circ\circ}$	–
		LPS	0.788$^{\circ\circ\circ}$	0.774$^{\circ\circ\circ}$	0.789$^{\circ\circ\circ}$	0.776$^{\circ\circ\circ}$	0.780$^{\circ\circ\circ}$	**0.773$^{\circ\circ\circ}$**	0.793$^{\circ\circ\circ}$	0.782$^{\circ\circ\circ}$	–
$h=8$	RealTime	RQPS	0.762$^{\circ\circ}$	0.759$^{\circ\circ}$	0.740$^{\circ\circ}$	0.737$^{\circ\circ}$	0.761$^{\circ\circ}$	0.784$^{\circ\circ}$	**0.733$^{\circ\circ}$**	0.744$^{\circ\circ}$	–
		LPS	0.695$^{\circ\circ}$	0.686$^{\circ\circ}$	0.668$^{\circ\circ}$	0.659$^{\circ\circ}$	0.686$^{\circ\circ}$	0.706°	**0.655$^{\circ\circ}$**	0.662$^{\circ\circ}$	–
	FinRev	RQPS	0.973	0.965††	0.902$^{\circ\circ}$	**0.893$^{\circ\circ}$**	0.905$^{\circ\circ}$	0.913$^{\circ\circ}$	0.904$^{\circ\circ}$	0.900$^{\circ\circ}$	–
		LPS	0.960	0.932†††	0.870$^{\circ\circ\circ}$	**0.851$^{\dagger\circ\circ\circ}$**	0.872$^{\circ\circ\circ}$	0.871$^{\circ\circ\circ}$	0.864$^{\circ\circ}$	0.857$^{\dagger\circ\circ\circ}$	–

Figure 1: Quantile regressions estimates

Notes: Row 1 shows estimated coefficients for *spread* when $X_{t-h} = (Spread_{t-h}, y_{t-h})$. Rows 2 and 3 show estimated coefficients for *spread* and *spread*^{2} when $X_{t-h} = (Spread_{t-h}, Spread^2_{t-h}, y_{t-h})$. Standard errors were obtained by paired bootstrap with 2000 replications. Asterisks indicate statistical significance at 10%.

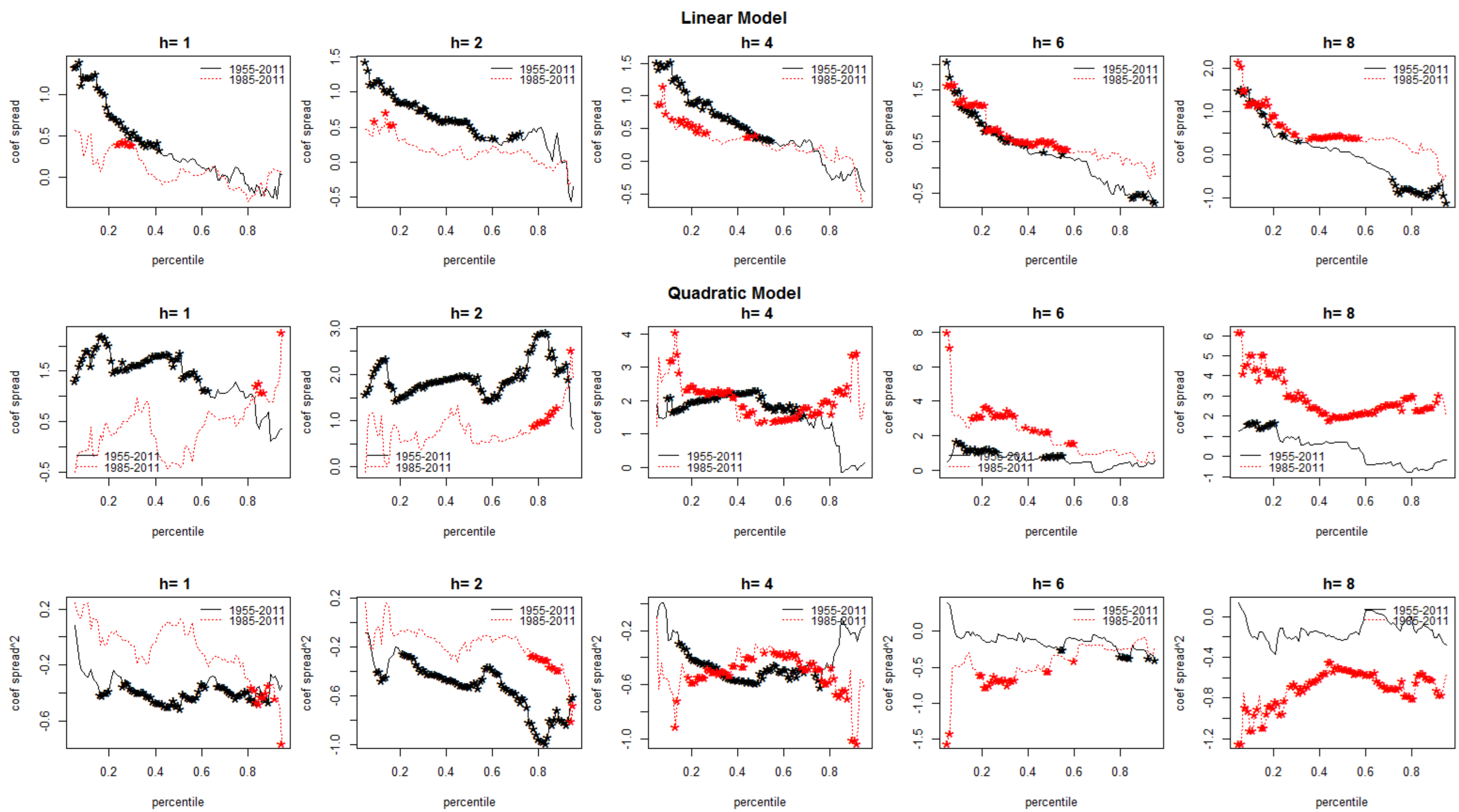

Figure 2: Adjusted $R1(\tau)$ statistics

Notes: This figure shows the $\overline{R1}(\tau)$ statistics of Koenker and Machado (1999) computed for the following specifications: (i) $X_{t-h} = Spread_{t-h}$, (ii) $X_{t-h} = (Spread_{t-h}, Spread^2_{t-h})$, (iii) $X_{t-h} = (Spread_{t-h}, y_{t-h})$ and (iv) $X_{t-h} = (Spread_{t-h}, Spread^2_{t-h}, y_{t-h})$.

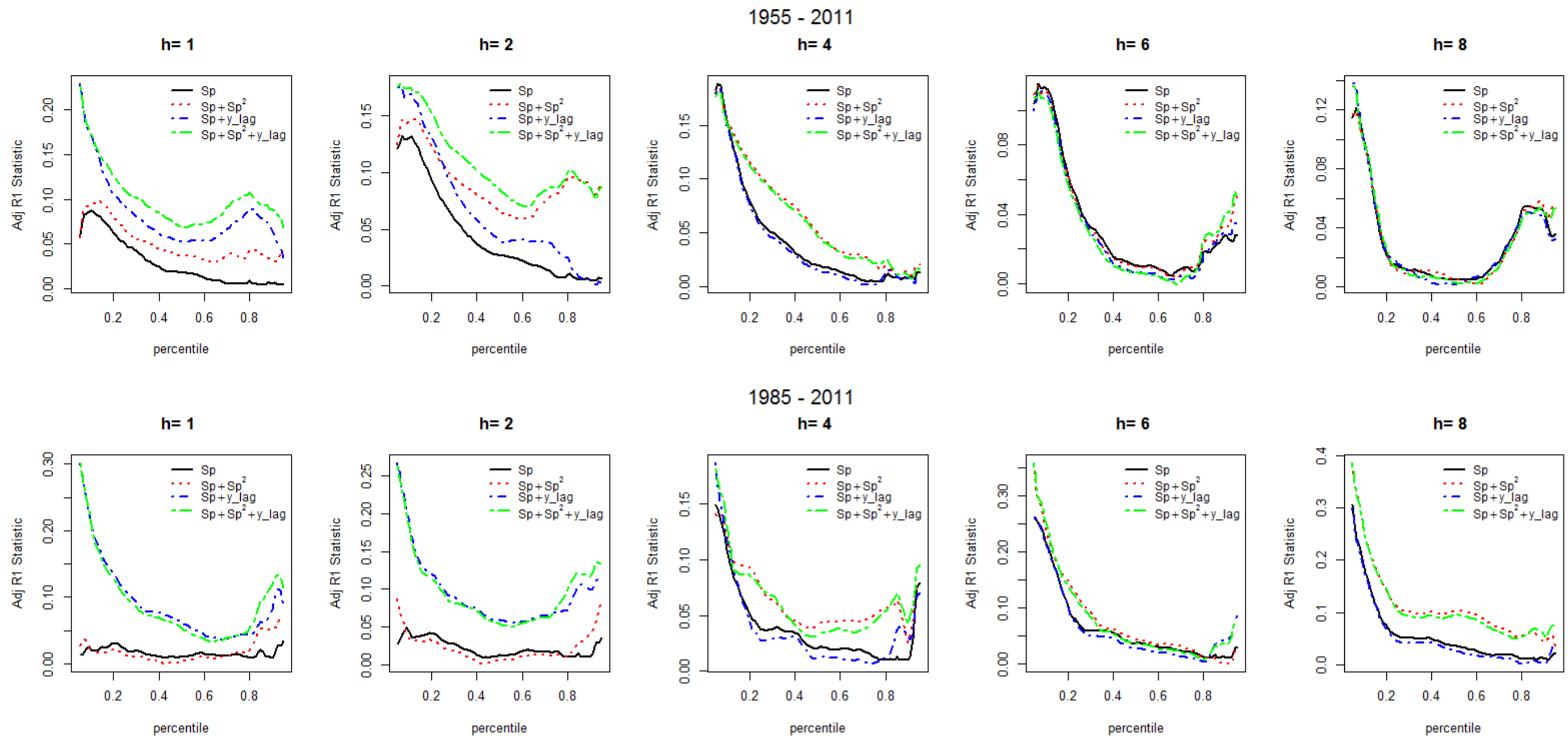

Figure 3: Recursive GDP growth forecasting in real-time

Notes: Panels (a), (b), (c) and (d) show recursive GDP growth forecast errors for 1, 2, 3 and 5 quarters ahead, respectively. Mean (Median) indicates SPF mean (median) and OLS (QR(0.5)) recursive forecast errors. The sample period is 1955Q1-2011Q2. Shaded areas show NBER recessions and asterisks denote statistical significance against SPF at the 10% level.

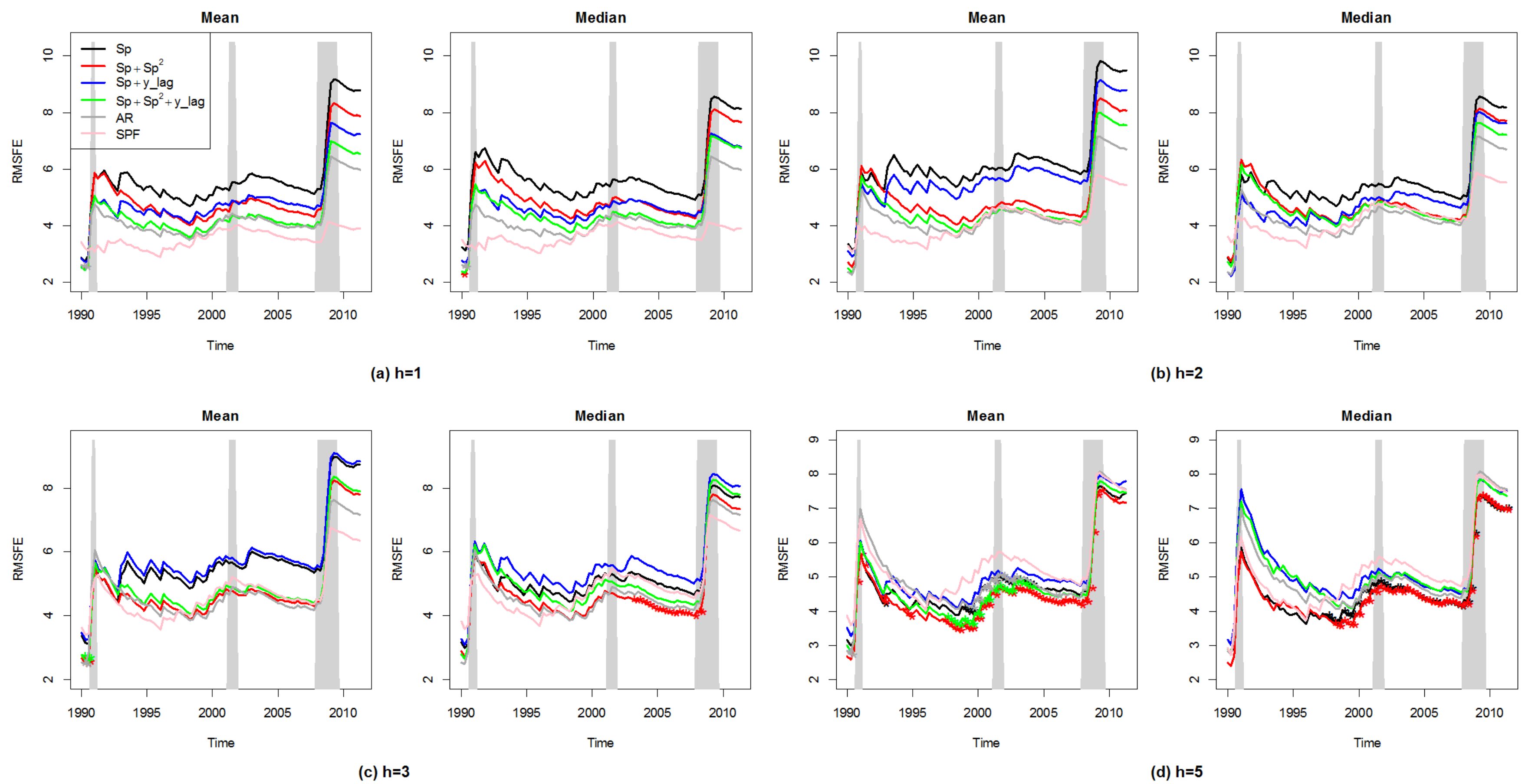

Figure 4: Recursive recession forecasting in real-time

Notes: Panels (a), (b), (c) and (d) show recursive forecast errors of QR recessions probabilities for 1, 2, 3 and 5 quarters ahead, respectively. The sample period is 1955Q1-2011Q2. Shaded areas show NBER recessions and asterisks denote statistical significance against SPF at the 10% level.

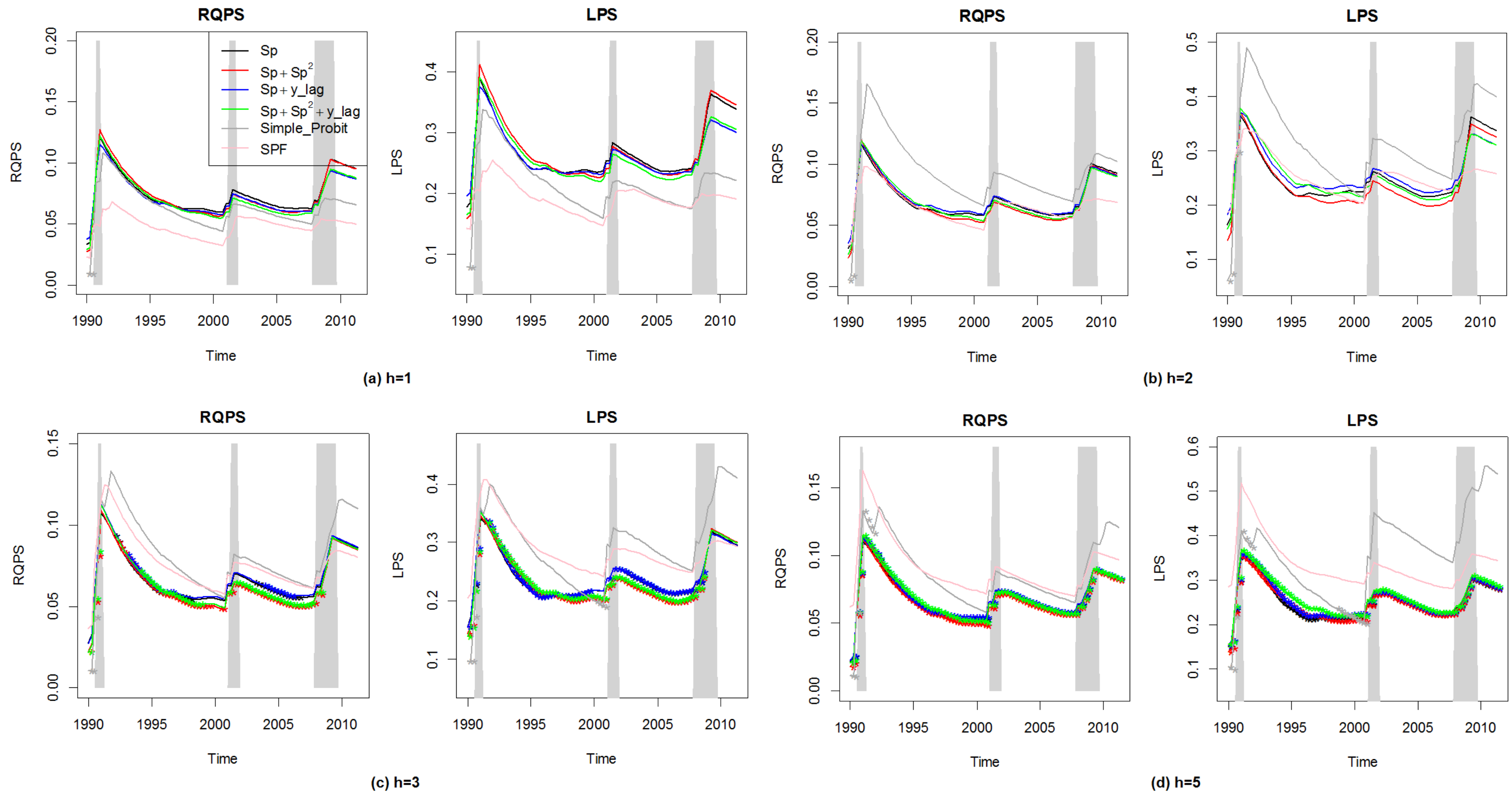

Figure 5: Forecasting the great recession in real-time

Notes: This figure shows quantile forecasts for GDP growth in real-time. We show results for $\tau = 0.02, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.98$, $h = 2, 4, 6, 8$ and specification $X_t = (Spread_t, Spread_t^2)$. The red line refers to median forecasts. The sample period is 1955Q1-2011Q2.