Essays in Economic Forecasting

André B. M. Souza

TESI DOCTORAL UPF / 2021

THESIS SUPERVISOR
Professor Christian Brownlees
Department d'Economia i Empresa

Universitat Pompeu Fabra Barcelona

Para a minha família.
Til min familie.

Acknowledgements

First and foremost, I am extremely grateful to my advisor Christian Brownlees for his invaluable guidance and support throughout my doctoral studies. His feedback, patience, and consideration were paramount in my development.

I am also very grateful to Geert Mesters, Barbara Rossi, Kirill Evdokimov, Katerina Petrova, Eulalia Nualart, Piotr Zwiernik and Majid Al-Sadoon for their time and thoughtful comments. This thesis greatly benefited from discussions with Francesco Ravazzolo and Marcelo Medeiros, to whom I am very thankful. I would also like to thank Albert Banal-Estañol, Javier Gomez-Biscarri and Jo Seldeslachts for their consideration and support, and Marta Araque and Laura Agustí for their crucial assistance throughout the PhD.

I am grateful to my colleagues and friends Adam, Bjarni, Jordi, Katharina, Lukas, Philipp, and Yiru for their advice and comments, but also for the shared coffees, lunches and barbecues. I am also very grateful to my partner, Synne, for her support and understanding throughout the years.

I gratefully acknowledge financial support from the Spanish Ministry of Science and Technology (FPI Grant BES-2017-080615).

Finally, none of this would have been possible without the unconditional support provided by my parents. *A vocês a minha mais profunda gratidão.*

Abstract

This dissertation consists of two independent chapters on economic and financial forecasting. The first chapter introduces a nonlinear forecasting framework that combines forecasts of the sign and absolute value of a time series into conditional mean forecasts. In contrast to linear models, the proposed framework allows different predictors to separately impact the sign and absolute value of the target series. An empirical application using the FRED-MD dataset shows that forecasts from the proposed model substantially outperform linear forecasts for series that exhibit persistent volatility dynamics, such as output and interest rates. The second chapter, coauthored with Christian Brownlees, provides an extensive comparison of methods to forecast downside risks to GDP growth for a panel of 24 OECD economies. We consider forecasts constructed from standard quantile regressions as well as from conditional volatility models. Our evidence suggests that standard volatility models such as the GARCH(1,1) are at least as accurate as quantile regressions.

Resum

Aquesta dissertació consta de dos capítols independents sobre previsió econòmica i financera. El primer capítol introdueix un modelo de predicció no lineal que combina les previsions del signe i del valor absolut d'una sèrie temporal en previsions mitjanes condicionals. A diferència dels models lineals, el modelo proposat permet que diferents variables afectin per separat el signe i el valor absolut de la sèrie d'interés. Una aplicació empírica que utilitza el conjunt de dades FRED-MD mostra que les previsions basadas en el modelo proposat superen substancialment les previsions lineals per a sèries que presenten dinàmiques de volatilitat persistents, com la producció industrial i els tipus d'interès. El segon capítol, coautorado con Christian Brownlees, proporciona una àmplia comparació de mètodes per predir els riscos negatius per al creixement del PIB per a un grup de 24 economies de l'OCDE. Considerem les previsions construdes a partir de regressions quàntils estàndard, així com a partir de models de volatilitat condicional. La nostra evidència suggereix que els models de volatilitat, com el GARCH (1,1), són almenys tan precisos com les regressions quantils.

Preface

Whether employed by policymakers as a basis for policy discussions or by financial market participants as a risk management tool, forecasting is a central topic in economics. This thesis consists of two self-contained chapters on forecasting economic time series. The first chapter focuses on producing *point* forecasts for the conditional mean of a random variable. The second chapter studies the construction of *interval* forecasts for the growth rates of gross domestic product (GDP) for 24 OECD economies. Interval forecasts provide a measure of uncertainty about future economic conditions and are widely used by central banks and international institutions alike.

The first chapter of this thesis studies the construction of conditional mean forecasts. Several economic time series exhibit strong evidence of volatility persistence and yet weak, if any, evidence of linear conditional mean predictability. I introduce a non-linear forecasting framework that exploits conditional volatility dynamics by combining forecasts of the sign and absolute value of a time series into conditional mean forecasts. Among other results, I show that the conditional mean may be accurately approximated by the product of the mean squared error optimal sign and absolute value forecasts. In contrast to linear models, the proposed framework allows different predictors to separately impact the sign and absolute value of the target series, thus being able to capture potentially non-linear dynamics. In an application to the FRED-MD dataset, a dataset containing several macroeconomic series for the United States, I compare the performance of forecasts based on the proposed framework with that of a number of standard forecasting models such as Principal Components regression, Ridge and LASSO. I find that forecasts based on the proposed framework substantially outperform linear forecasts for series with persistent volatility dynamics, such as output and interest rates.

In the second chapter of the thesis, "Backtesting Global Growth-at-Risk" which is coauthored with Christian Brownlees, we compare several methods to construct quantile forecasts of GDP growth. More commonly known as Growth-at-Risk (GaR), such forecasts have recently become a standard tool in policymakers' toolbox. We conduct an extensive out-of-sample backtesting exercise of GaR predictions for 24 OECD countries. We consider forecasts constructed from quantile regressions, the standard tool employed by policymakers and central banks, and GARCH models. The quantile regression forecasts are based on several measures of downside risks to GDP, including the national financial conditions index. Our results show that quantile regression and GARCH forecasts have a similar performance. In fact, if anything, our evidence suggests that standard volatility models such as the GARCH(1,1) are more accurate in predicting downside risks to GDP than the widely used quantile regressions.

Sumari

Índex de figures

Índex de taules

Capítol 1

Composite Absolute Value and Sign Forecasts

1.1 Introduction

Forecasting macroeconomic and financial time series is a challenging task. A ubiquitous finding in the economic forecasting literature is that linear models for the conditional mean improve only marginally, if at all, over simple benchmarks (Stock and Watson, 2007; Goyal and Welch, 2008). In contrast, standard volatility models have shown considerable success in capturing volatility dynamics (Engle, 1982; Andersen et al., 2006; Brownlees et al., 2011) and there is evidence of directional predictability (Leung et al., 2000; Diebold et al., 2007; Nyberg, 2011). Several explanations are available for the unconvincing performance of linear conditional mean forecasts. The predictable component of the target series may be small relative to the unpredictable error term, in which case even a correctly specified model will display only mild gains over simple benchmarks. Alternatively, this finding may be taken as evidence of model misspecification, implying that the class of models considered is not rich enough to exploit all available information.

In this work, I introduce composite absolute value and sign (CAVS) forecasts, a nonlinear forecasting framework that exploits predictability in signs and absolute values to generate conditional mean forecasts.[1] Based on the fact that any random variable Y_t can be written as $|Y_t|\text{sign}(Y_t)$, CAVS forecasts are defined as a function of mean squared error (MSE) optimal sign and absolute value forecasts. In contrast to linear models, in which conditional mean predictors must impact both the sign and absolute value, CAVS allows for different predictors to

[1] I assume throughout that the sign of the target series is not constant, as would be the case for series expressed in growth rates.

separately affect each of the components of the target series. In contrast to general nonlinear models, CAVS forecasts are simple to interpret and are designed to exploit two specific features that figure prominently in the macroeconomic and financial forecasting literature: volatility and sign predictability.

I introduce a framework to formalize CAVS forecasts and study its properties. The proposed framework is employed to establish three results. First, the conditional mean can be written as the product of MSE optimal forecasts of signs and absolute values as well as a covariance term that can be explicitly modeled. If the underlying data-generating process (DGP) is additive with symmetric shocks, the covariance term will be small relative to the variance of the shocks. Second, I provide an upper bound to the MSE of CAVS forecasts that scales with the losses in sign and absolute value forecasting. This result highlights that CAVS-based forecasts are particularly suited for series that exhibit persistent volatility dynamics, and hence absolute value predictability. Third, I study a nonlinear DGP in which variables may affect signs and absolute values differently. I show that, for this DGP, the MSE of the best linear predictor increases quadratically with the degree of nonlinearity. A simulation study highlights that departures from linearity generate substantial MSE gains for CAVS models.

The proposed framework is applied to the FRED-MD dataset (McCracken and Ng, 2016), which consists of 128 monthly financial and macroeconomic time series. I construct and evaluate 1, 3, 6, and 12 months ahead forecasts for the conditional mean, the absolute value and the sign of each series in the dataset. It is well-known that when the number of predictors is large, dimension reduction techniques may improve forecast accuracy (Stock and Watson, 2012; Ng, 2013; Kim and Swanson, 2014). I consider principal components regression (PCR), ridge and LASSO as the baseline linear models. CAVS forecasts are constructed as the product of sign and absolute value forecasts, and a number of specifications may be entertained. I consider CAVS-PCR, CAVS-Ridge, and CAVS-LASSO as the baseline CAVS models. Each baseline model is constructed using the same method to forecast both the sign and the absolute value of the target variable. In addition to the baseline CAVS models, I consider absolute value forecasts implied by a GARCH(1,1), sign forecasts obtained by a variety of machine learning algorithms, as well as combinations of sign and absolute value forecasts constructed using different methods to forecast each component. Detailed results are presented for the series considered in Kim and Swanson (2014). These include the unemployment rate, personal income less transfer payments, the 10-year treasury rate, the consumer price index, the producer price index, nonfarm payroll employment, industrial production, M2 money stock, and the S&P 500 index. In addition to the variables considered in Kim and Swanson (2014), I present detailed results for the federal funds rate, which may be of particular interest to private sector forecasters.

A number of findings emerge from the empirical application. First, CAVS-based directional and conditional mean forecasts outperform their linear counterparts for the majority of the selected series across all horizons considered. In particular, CAVS-based forecasts are substantially more accurate than linear forecasts for the federal funds rate, industrial production, nonfarm payroll employment, and the S&P 500 across all horizons. Among the linear models considered, ridge and LASSO display similar performance and outperform PCR, particularly for short forecast horizons. All baseline CAVS specifications considered perform similarly, with CAVS-Ridge modestly outperforming CAVS-LASSO and CAVS-PCR.

Second, I compare the performances of CAVS-Ridge and PCR for all components of the FRED-MD dataset. CAVS-Ridge outperforms PCR for the majority of the FRED-MD components, and is particularly successful for series that exhibit persistent conditional volatility dynamics, such as interest rates, stocks, and output series. As the forecast horizon increases, the CAVS-Ridge performance gains become widespread, outperforming PCR for about 75% of the FRED-MD components. Notably, these findings remain qualitatively the same for any choice of CAVS specification and linear benchmark considered.

Finally, I explore a number of alternative CAVS specifications. In addition to combinations between signs and absolute values obtained by PCR, Ridge, and LASSO, I consider absolute value forecasts based on a GARCH(1,1) and sign forecasts obtained from random forests (Breiman, 2001), AdaBoost (Freund and Schapire, 1995), k-nearest neighbors (kNN; see Devroye et al., 1996) and neural networks (NN; see White, 2006). All specifications perform similarly, with GARCH(1,1)-Ridge displaying modest gains over the baseline CAVS specifications. Among the machine learning algorithms, the most accurate forecasts are achieved by methods based on regression trees. In particular, random forests produce more accurate forecasts when compared to the baseline CAVS specifications for 20% of the selected series. Additionally, I consider forecast combinations and model selection strategies. In line with Timmerman (2006) and Diebold and Shin (2019), I find that forecast averaging performs better than model selection. In addition, I find that forecast averages that include CAVS forecasts outperform those that do not. Overall, the results support the view that exploiting nonlinearities in macroeconomics series improves forecast accuracy (see also Marcellino, 2002; Clements et al., 2004; Terasvirta, 2006).

It is important to emphasize that this is not the first paper to consider the decomposition $Y_t = |Y_t|\text{sign}(Y_t)$. When applied to stock returns, Diebold and Christoffersen (2006) argue that both of the right-hand side components are predictable, yet their product is not. Anatolyev and Gerko (2005) apply this decomposition to develop a market timing test, whereas Rydberg and Shephard (2003) employ it to model the dynamics of trade-by-trade price movements in a market microstructure setting. Closely related to this work, Anatolyev and Gospodinov (2010) and

Anatolyev and Gospodinov (2019) model excess returns by combining a multiplicative error model for the absolute values, a dynamic binary model for signs, and a copula for their interaction. In fact, the CAVS framework is a special case of the decomposition model proposed in Anatolyev and Gospodinov (2010) where the independence copula is employed. As noted in Anatolyev and Gospodinov (2010), employing the independence copula avoids the most effort-consuming ingredient of the decomposition model – the evaluation of the conditional expected cross-product of signs and absolute values.

This paper is related to the macroeconomic forecasting literature, which includes work by Stock and Watson (2002, 2012), Kim and Swanson (2014), and Cheng and Hansen (2015), among others. Additionally, this work also relates to the literature on nonlinear forecasting methods in economics. This includes work by Clements et al. (2004), Terasvirta (2006), and White (2006), among others.

The remainder of this chapter is structured as follows. Section 1.2 introduces the CAVS framework, Section 1.3 contains a simulation study, and Section 1.4 presents the empirical application. Concluding remarks follow in Section 1.5.

1.2 CAVS Forecasts

1.2.1 Notation and Definition

Let $\{Y_t\}$ denote a zero mean scalar time series of growth rates[2] and $\{X_t\}$ with $X_t \in \mathbb{R}^n$ a time series of $t-1$ measurable predictors. In this work, I consider a forecaster whose objective is to construct a forecast of Y_t given X_t that minimizes the MSE. It is well known (see Brockwell and Davis, 1991) that the MSE optimal forecast of Y_t given X_t is the conditional expectation

$$\mu(X_t) = \mathbb{E}\left[Y_t \middle| X_t\right] = \arg\min_{m\in\mathcal{M}} \mathbb{E}\left[\left(Y_t - m(X_t)\right)^2\right] ,$$

where $\mathcal{M}$ is the collection of measurable functions m of X_t having finite variance.

The methodology introduced in this paper builds on a decomposition of the conditional mean into the product of the conditional expectations of the absolute value and sign of the target variable. It is straightforward to verify that the identity $Y_t = |Y_t|\text{sign}(Y_t)$ implies that the conditional mean may be expressed as

$$\mu(X_t) = \mathbb{E}\left[|Y_t| \middle| X_t\right] \mathbb{E}\left[\text{sign}(Y_t) \middle| X_t\right] + \mathbb{C}\text{ov}\left(|Y_t|, \text{sign}(Y_t) \middle| X_t\right) . \qquad (1.1)$$

The representation in (1.1) highlights that the optimal forecast for Y_t given X_t can be written as the product of MSE optimal absolute value and sign forecasts and

[2] I assume that $\mathbb{E}[Y_t^2] < \infty$ throughout.

a conditional covariance term. This decomposition motivates the introduction of CAVS forecasts, a nonlinear forecasting framework that exploits componentwise predictability in the absolute value and sign to approximate the conditional mean.

Formally, denoting by $\mathcal{C}_A$ and $\mathcal{C}_S$ collections of functions for the absolute value and the sign of Y_t, CAVS forecasts are defined as

$$\mu_{\text{CAVS}}(X_t) = m_A^*(X_t)m_S^*(X_t) + c(X_t) \ , \tag{1.2}$$

where

$$m_A^*(X_t) = \arg\min_{m\in\mathcal{C}_A} \mathbb{E}\Big[\Big(|Y_t| - m(X_t)\Big)^2\Big] \ , \tag{1.3}$$

$$m_S^*(X_t) = \arg\min_{m\in\mathcal{C}_S} \mathbb{E}\Big[\Big(\text{sign}(Y_t) - m(X_t)\Big)^2\Big] \ , \tag{1.4}$$

$$c(X_t) = \mathbb{C}\text{ov}\Big(|Y_t|, \text{sign}(Y_t)\Big|X_t\Big) \ . \tag{1.5}$$

It is important to emphasize that CAVS forecasts do not generally coincide with the conditional mean. In particular, if $\mathcal{C}_A$ and $\mathcal{C}_S$ do not include the optimal forecasts $\mu_A(X_t) = \mathbb{E}\big[|Y_t|\big|X_t\big]$ and $\mu_S(X_t) = \mathbb{E}\big[\text{sign}(Y_t)\big|X_t\big]$ respectively, the CAVS forecast will differ from the conditional mean.

It is widely documented in the economic forecasting literature that linear models for the conditional mean may exhibit poor out-of-sample performance (Stock and Watson, 2007; Rossi, 2013). Nonlinear models are typically able to approximate arbitrary functions (White, 1990). This flexibility comes at the expense of increased computational complexity, heightened risks of overfitting, and difficulties of interpretation (White, 2006), and the evidence regarding the performance of nonlinear models in macroeconomic and financial time series is inconclusive (Clements et al., 2004; Terasvirta, 2006). In contrast to general nonlinear models, the CAVS framework is designed to exploit two specific features that figure prominently in the macroeconomic and financial forecasting literature: volatility and sign predictability (see Diebold and Christoffersen, 2006).

The interpretation of the CAVS forecast is straightforward: it is the expected magnitude of the next realization weighted by the probability that it will be positive (or negative). In fact, the CAVS forecast is the MSE optimal forecast given knowledge of either the sign or absolute value of the next realization. Computationally, the construction of CAVS forecasts requires the estimation of models for the absolute value and sign of the target series. The results from the empirical application show that CAVS forecasts based on generalized linear models — which are simple to compute — add substantial flexibility when compared to linear models for the conditional mean.

To make the CAVS definition operational, the forecaster must specify appropriate functions for the absolute value, the sign, and the covariance term $c(X_t)$.

The absolute value of the target series may be modeled as generalized linear functions of the predictors. Alternatively, building on the empirical success of GARCH models (Engle, 1982; Brownlees et al., 2011), absolute value forecasts may be constructed based on conditional volatility forecasts.

A number of possibilities are available to construct sign forecasts. First, one may employ standard generalized linear models, such as logit and probit, to forecast the signs. Alternatively, the machine learning literature has put forward a number of methods to forecast binary random variables and their associated probabilities. In the empirical application, generalized linear models as well as nonparametric methods are used to construct probability forecasts. Note that the sign forecast used in the construction of CAVS forecasts is based on the MSE (of signs) and hence is a probabilistic, rather than binary, forecast. The MSE is a natural choice in this setting as it elicits the (componentwise) conditional expectation. Alternative combinations of componentwise loss functions may be explored, particularly for forecasting under different loss functions.

Finally, I remark that the conditional covariance of signs and absolute values is often negligible for forecasting purposes. Empirically, Anatolyev and Gospodinov (2010) find evidence of weak dependence between the conditional signs and absolute values of excess returns. In fact, Proposition 3 shows that, for additive models with symmetric shocks, $c(X_t)$ is small relative to the irreducible MSE and may therefore be ignored for forecasting purposes.

1.2.2 Theoretical Results

This section provides a number of theoretical results for CAVS forecasts. Throughout this section, I express Y_t as the sum of the conditional mean and an unpredictable error term:

$$Y_t = \mu(X_t) + \sigma_u u_t \ , \ u_t \overset{i.i.d}{\sim} \mathcal{D}(0,1) \ , \tag{1.6}$$

where $\mathbb{E}[Y_t] = 0$, $\mathbb{E}[Y_t^2] < \infty$ and $\mathcal{D}$ is a distribution with mean zero and unit variance. I assume throughout that $\mathbb{E}[|Y_t||X_t] \leq c < \infty$ for all X_t and some $c \in \mathbb{R}$. All proofs are presented in the Appendix.

Bounding the loss of CAVS forecasts. Forecasts of economic series are typically based on misspecified approximations to the conditional mean (see White, 2006). The following proposition provides an upper bound for the MSE of CAVS forecasts that depends the approximating properties of $\mathcal{C}_A$ and $\mathcal{C}_S$.

Proposition 1. *The MSE of the CAVS forecast given in* (1.2) *is such that*

$$\mathbb{E}\Big[\Big(Y_t - \mu_{CAVS}(X_t)\Big)^2\Big] \leq \sigma_u^2 + a_1 \mathbb{E}\Big[\Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2\Big] \ ,$$

where $a_1 > 0$ *is a constant.*

Proposition 1 shows that the accuracy of CAVS forecasts depends on the accuracy of the approximations to $\mu_A(X_t)$ and $\mu_S(X_t)$. If $\mu_A(X_t) \in \mathcal{C}_A$ and $\mu_S(X_t) \in \mathcal{C}_S$, the CAVS forecast is equivalent to the conditional mean. A large body of research has documented the good forecasting performance of standard volatility models (see Andersen et al., 2006; Brownlees et al., 2011). This observation suggests that, for series with persistent volatility dynamics, $\mu_A(X_t)$ may be accurately approximated. In addition, Diebold and Christoffersen (2006) show that conditional volatility dynamics implies directional predictability, and Leung et al. (2000) document evidence of sign predictability in excess of that captured by linear models for the conditional mean. Taken together, these observations suggest that componentwise approximations to the conditional mean may be able to leverage volatility and sign predictability into accurate conditional mean forecasts.

Proposition 1 assumes that the forecaster knows $c(X_t)$, the conditional covariance of signs and absolute values. In practice, this is typically not the case. The forecaster may explicitly model this term. Alternatively, the next proposition describes the MSE of CAVS forecasts when $c(X_t)$ is unknown and set to 0 in (1.2).

Proposition 2. *The MSE of the forecast constructed as* $m^*_A(X_t)m^*_S(X_t)$ *is such that*

$$\mathbb{E}\Big[\Big(Y_t - m^*_A(X_t)m^*_S(X_t)\Big)^2\Big] \leq \sigma_u^2 + a_2\mathbb{E}\Big[\big(\mu_S(X_t) - m^*_S(X_t)\big)^2 + \big(\mu_A(X_t) - m^*_A(X_t)\big)^2\Big] + 2\mathbb{E}\Big[c(X_t)^2\Big] ,$$

where $a_2 > 0$ *is a constant.*

Proposition 2 shows that the excess risk, in MSE terms, incurred due to setting $c(X_t)$ to 0 in the construction of the CAVS forecast is additive and proportional to $\mathbb{E}[c(X_t)^2]$. Next, I show that $\mathbb{E}[c(X_t)^2]$ is small relative to σ_u^2, the irreducible uncertainty.

Proposition 3. *Considering the representation given in* (1.6) *and assuming* $\mathcal{D}$ *is a t distribution with* $\nu > 2$ *degrees of freedom, then*

$$\mathbb{E}\Big[\Big(Y_t - \mu_A(X_t)\mu_S(X_t)\Big)^2\Big] \leq (1+\gamma)\mathbb{E}\Big[\Big(Y_t - \mu(X_t)\Big)^2\Big] ,$$

with $\gamma \approx 0.04187$.

Proposition 3 provides an upper bound to the contribution of the conditional covariance term — the approximation error obtained from a componentwise approximation of $\mu(X_t)$ — to the overall MSE. Combining this result with Proposition 2 yields the following corollary.

Corollary 1. *Consider the representation given in* (1.6) *and assume* $\mathcal{D}$ *is a* t *distribution with* $\nu > 2$ *degrees of freedom. Then, the MSE of the forecast constructed as* $m_A^*(X_t)m_S^*(X_t)$ *is such that*

$$\mathbb{E}\Big[\Big(Y_t - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] \leq 1.05\sigma_u^2 + a_2\mathbb{E}\Big[\big(\mu_S(X_t) - m_S^*(X_t)\big)^2\Big] + a_2\mathbb{E}\Big[\big(\mu_A(X_t) - m_A^*(X_t)\big)^2\Big] \, ,$$

where a_2 *is the same as in Proposition 2.*

Corollary 1 shows that the MSE of the CAVS forecast constructed by neglecting $c(X_t)$ is driven by the losses in sign and absolute value forecasting. Taken together, Propositions 1 – 3 highlight that CAVS forecasts may provide suitable approximations to the conditional mean, especially when the componentwise forecasts are accurate approximations of their targets. Next, I present a DGP in which accurate approximations are available for each of the components, yet linear models for the conditional mean are not able to exploit all available information.

Linear forecasts in a nonlinear setting. Consider the following DGP:

$$Y_t = \big|X_t'(\beta + \delta e_1)\big|\text{sign}(X_t'\beta) + u_t, \qquad u_t \overset{iid}{\sim} \mathcal{D}(0, \sigma_u^2) \, , \tag{1.7}$$

where $X = (x_{1,t}, x_{2\,t})'$, $e_1 = (1, 0)$, and $\delta \in \mathbb{R}$. This choice of functional form allows x_1 to affect the sign and absolute value of the target variable differently. Note that linear models are obtained by setting $\delta = 0$. In contrast, as $|\delta|$ increases, the resulting model exhibits weak linear conditional mean predictability with strong sign and absolute value predictability. In addition, $\kappa = \frac{\beta_1 + \delta}{\beta_1}$ controls the size of δ relative to β_1, thus parameterizing departures from linearity in this model. Both the absolute value and the sign of Y_t may be accurately approximated by generalized linear models, yet linear models for the conditional mean will have poor performance as $|\kappa - 1|$ increases.

Proposition 4. *Consider the model given in* (1.7) *with* $\beta_1 = \beta_2 = \beta$ *and* $X_t \sim \mathcal{D}_X(0, \sigma^2 I)$. *The MSE of the best linear predictor is given by*

$$\begin{aligned}\mathbb{E}\Big[\Big(Y_t - X_t'\beta^*\Big)^2\Big] &= \sigma_u^2 + \beta^2\kappa^2\Big(\sigma^2 - \frac{\mathbb{E}[x_1^2 S(\kappa)]^2}{\sigma^2} - \frac{\mathbb{E}[x_1 x_2 S(\kappa)]^2}{\sigma^2}\Big) \\ &\quad - 2\beta^2\kappa\mathbb{E}[x_1 x_2 S(\kappa)]\Big(\frac{\mathbb{E}[x_1^2 S(\kappa)]}{\sigma^2} + \frac{\mathbb{E}[x_2^2 S(\kappa)]}{\sigma^2}\Big) \\ &\quad + \beta^2\Big(\sigma^2 - \frac{\mathbb{E}[x_2^2 S(\kappa)]^2}{\sigma^2} - \frac{\mathbb{E}[x_1 x_2 S(\kappa)]^2}{\sigma^2}\Big) \, ,\end{aligned}$$

where $\beta^* = \arg\min_\beta \mathbb{E}\Big[\Big(Y_t - X_t'\beta\Big)^2\Big]$ *and* $S(\kappa) = \text{sign}(x_{1\,t}^2\kappa + x_{1\,t}x_{2\,t}(1+\kappa) + x_{2\,t}^2)$.

Proposition 4 provides an exact formula for the MSE of the best linear predictor in a model where variables impact the sign and absolute value of the target differently. Although no closed formula solution is available for the expectations involved, they can be evaluated by numerical methods. In addition, the order of the MSE of the best linear predictor can be computed for large κ.

Corollary 2. *There exists a κ_0 such that for all $\kappa > \kappa_0$,*

$$\mathbb{E}\left[\left(Y_t - X_t'\beta^*\right)^2\right] \geq a_3\kappa^2 ,$$

where $a_3 > 0$ is a constant that depends on $\mathbb{E}[x_{1\,t}^2 S(\kappa)]$, $\mathbb{E}[x_{2\,t}^2 S(\kappa)]$ and $\mathbb{E}[x_{1\,t}x_{2\,t}S(\kappa)]$.

Corollary 2 shows that, for large enough κ, the MSE of the best linear predictor grows with κ^2. The next section contains detailed simulation results for this model.

1.3 Simulation Study

In this section, I carry out a simulation study to numerically evaluate the performance of CAVS forecasts. I consider two simulation settings. In addition to the model given in (1.7), I simulate from the following model:

$$Y_t = \left|X_t'\beta\right|\text{sign}(X_t'(\beta + \delta e_1)) + u_t, \qquad u_t \overset{iid}{\sim} N(0, 1) . \tag{1.8}$$

Both models are nonlinear in x_1 yet approximately linear in x_2, and nest a linear model when $\kappa = 1$. The nonlinearity in model (1.7) arises due to increases in the coefficient attached to x_1 in the absolute value component. This implies that the variance of the target variable increases with $\kappa = \frac{\beta_1+\delta}{\beta_1}$. In contrast, the nonlinearity in model (1.8) arises on the sign component, and hence does not influence the variance of the target variable. In both models, departures from linearity are obtained by varying κ. Each predictor is an i.i.d draw from a standard Gaussian distribution. Three forecasting strategies are compared. First, I consider the linear forecast obtained by ordinary least squares regression of Y_t on X_t. Second, I consider the CAVS forecast in (1.2) constructed by setting $c(X_t)$ to 0. Finally, I consider a CAVS forecast where $c(X_t) = 0$ and sign forecasts are obtained from a probit regression of $\mathbb{1}\{Y_t > 0\}$ on X_t.

FIGURE 1.1 ABOUT HERE

Figure 1.1 illustrates the MSE ratios of each forecasting strategy relative to the best linear predictor for the two simulation settings considered. Values below 1 indicate that the MSE of a given strategy is smaller than that of the linear forecast.

Solid lines represent the CAVS forecasts constructed by setting $c(X_t) = 0$, whereas dashed lines represent the CAVS forecasts constructed with $c(X_t) = 0$ and sign forecasts based on a probit regression. The results from both simulation settings are similar. For large values of κ or for $\kappa \approx 0$, CAVS forecasts substantially outperform the linear forecast. In particular, for large κ, CAVS forecasts display the strongest performance gains on the DGP in which nonlinearities arise in the absolute value of the target series. In contrast, for $\kappa \approx 0$, CAVS forecasts display the largest gains on the DGP in which nonlinearities arise in the sign of the target series. For $\kappa \approx 1$, the benefits of exploiting the existing nonlinearities are outperformed by the cost of neglecting the conditional covariance term by setting $c(X_t)$ to 0, and linear models that ignore the nonlinearities will perform better than a CAVS forecast that does not model the conditional covariance term. Overall, the simulations highlight that deviations from linearity — such as different predictors impacting different components of the target series — generate sizable gains for CAVS forecasts.

1.4 Empirical Application

I employ the framework introduced in this work to forecast the components of the FRED-MD dataset (McCracken and Ng, 2016), which consists of 128 monthly financial and macroeconomic series. As in McCracken and Ng (2016), series are arranged in eight groups: (1) output and income; (2) labor market; (3) housing; (4) consumption, orders and inventories; (5) money and credit; (6) interest and exchange rates; (7) prices; and (8) stock market. Detailed results are presented for a subset of the selected series considered in Kim and Swanson (2014). Additionally, I also report results for the Fed Funds rate, which may be particularly of interest to private sector forecasters.

All series are transformed as suggested in McCracken and Ng (2016). In particular, some series are not expressed in growth rates.[3] Series that do not change signs over the whole sample are excluded from the forecast comparison, but are kept as predictors to forecast the remaining variables. Additionally, the suggested transformations assume price series are integrated of order two, implying that the transformed series are twice differenced. This may hamper sign predictability. Details of the data and their transformations can be found in McCracken and Ng (2016). Five series with more than 10 missing observations are dropped from the dataset.[4] The remaining missing values are replaced by the unconditional mean of

[3]For example, series in the housing group are the logs of housing starts, and hence are always positive.

[4]These are: New orders for consumer goods (ACOGNO), New orders for nondefense capital goods (ANDENOx), Trade weighted U.S. Dollar Index: Major currencies (TWEXMMTH),

the series computed over the whole sample. After all transformations, the panel consists of data for 123 series from March 1959 to January 2020, corresponding to 731 months.

There is a large literature on macroeconomic forecasting. Stock and Watson (2002, 2007, 2012) introduce diffusion index models and document the good performance of PCR for macroeconomic forecasting. Kim and Swanson (2014) compare the forecasting performance of a variety of dimension reduction techniques, and find that PCR is improved when combined with other shrinkage-based techniques. Cheng and Hansen (2015) consider forecast averaging methods, and find that model averaging improves over PCR at longer horizons and performs on par with other shrinkage methods at shorter forecast horizons. Most of these studies focus on linear models and emphasize the issues related to the dimensionality of the data.

In contrast, Stock and Watson (1999), White and Swanson (1997), Marcellino (2002), and Bai and Ng (2008) examine nonlinear forecasts of economic variables. There is mixed evidence regarding the performance of nonlinear models. Stock and Watson (1999) find limited evidence of univariate nonlinear models improving upon linear forecasts, and that the best nonlinear models are typically tightly parameterized. White and Swanson (1997) find that flexible nonlinear models are particularly suitable for forecasting at longer horizons and document evidence of nonlinearities in nine macroeconomic series. Marcellino (2002) documents substantial evidence of exploitable nonlinearities in a number of macroeconomic series in the European Monetary Union, and Bai and Ng (2008) find that allowing for nonlinearities in the construction of the factors may improve the accuracy of PCR forecasts. See Clements et al. (2004), White (2006), and Terasvirta (2006) for extended discussions on forecasting economic variables with nonlinear models.

1.4.1 Forecasting Methodology

I carry out a pseudo out-of-sample forecasting exercise on the FRED-MD dataset. I construct and evaluate conditional mean, absolute value and sign forecasts for $h = 1, 3, 6$ and 12 months ahead for 113 series.[5] Following Stock and Watson (2012) and Boot and Nibbering (2019), all models include 4 lags of the dependent variable on the predictive regression, and dimension reduction techniques are applied to the remaining predictors. Forecasts are produced recursively starting from January 1985 until the end of the sample.

consumer sentiment index (UMCSENTx), and the VXO volatility index (VXOCLSx).

[5]These are the 123 series considered with the exception of those that do not change sign in the evaluation period.

Linear Forecasts

Principal components regression For each series in the panel and at each out-of-sample period T, PCR forecasts are constructed using

$$\widehat{Y}^{\text{PCR}}_{T+h}(r) = \sum_{p=1}^{4} \widehat{\rho}_p Y_{T-p+1} + \sum_{i=1}^{r} \widehat{\lambda}_i \widehat{F}_{iT} \ ,$$

where $\widehat{F}_{iT}$ is the i-th principal component of $\{X_t\}_{t=1}^{T}$, and X_t denotes a vector of predictors. The predictive regression is estimated by ordinary least squares, and forecasts are constructed for $r = 1, \ldots, 100$. Following Boot and Nibbering (2019), for each out-of-sample period T, I choose,

$$r^* = \arg \min_{1,\ldots,100} \sum_{t=T-60}^{T} \left(Y_t - \widehat{Y}_t(r) \right)^2 ,$$

and take the PCR forecast to be $\widehat{Y}^{\text{PCR}}_{T+h}(r^*)$.

Penalized regression For each series in the panel and at each out-of-sample period T, penalized forecasts are constructed using the model:

$$\widehat{Y}_{T+h}(\lambda_l, l) = \sum_{p=1}^{4} \widehat{\rho}_p Y_{T-p+1} + X_T' \widehat{\beta} \ ,$$

where

$$(\widehat{\rho}_1, \ldots, \widehat{\rho}_4, \widehat{\beta}) = \arg \min \frac{1}{T} \sum_{t=4}^{T} \left(Y_{t+h} - \sum_{p=1}^{4} \widehat{\rho}_p Y_{t-p+1} + X_t' \widehat{\beta} \right)^2 + \lambda g_l(\beta)$$

and where λ is a tuning parameter, $g_1(\beta) = 2 \sum_{i=1}^{n} |\beta_i|$ is the LASSO penalty, and $g_2(\beta) = \sum_{i=1}^{n} \beta_i^2$ is the ridge penalty. Following Boot and Nibbering (2019), LASSO forecasts are constructed for $\log(\lambda) \in \{-30, -29.7, \ldots, 0\}$, whereas ridge forecasts are constructed for $\log(\lambda) \in \{-15, -14.7, \ldots, 15\}$. I then select, for each out-of-sample period T and choice of l,

$$\lambda_l^* = \arg \min_{\lambda} \sum_{t=T-60}^{T} \left(Y_t - \widehat{Y}_t(\lambda, l) \right)^2 ,$$

and the LASSO forecasts are given by $\widehat{Y}^{\text{LASSO}}_{T+h}(\lambda_1^*, 1)$, and ridge forecasts by $\widehat{Y}^{\text{Ridge}}_{T+h}(\lambda_2^*, 2)$.

CAVS Forecasts

Absolute value forecasts Absolute value forecasts are obtained using the same baseline methods employed to construct linear conditional mean forecasts, with minor adjustments. In particular, the target variable is $|Y_{T+h}|$ rather than Y_{T+h}, and a larger set of predictors $W_t = (X_t', |X_t'|)$ is considered.[6] In addition to PCR, ridge, and LASSO, I consider absolute value forecasts implied by an AR(4)-GARCH(1,1) model:

$$Y_{T+1} = \sum_{p=1}^{4} \widehat{\rho}_p Y_{T-p+1} + \sigma_{T+1|T} u_{T+1}, \qquad u_{T+1} \overset{i.i.d}{\sim} \mathcal{D}(0,1)$$

$$\sigma^2_{T+1|T} = \omega + \alpha \Big(Y_T - \sum_{p=1}^{4} \widehat{\rho}_p Y_{T-p} \Big)^2 + \beta \sigma^2_T \ ,$$

where $\mathcal{D}$ is a distribution with zero mean and unit variance. Forecasts are constructed as:

$$|Y_{T+1}|^{\mathrm{G}} = \frac{1}{B} \sum_{b=1}^{B} \Big| \sum_{p=1}^{4} \widehat{\rho}_p Y_{T-p+1} + \sigma_{T+1} \widehat{u}^b_{T+1} \Big| \ ,$$

where $\widehat{u}^b$ is a draw from the GARCH filtered residuals (Barone-Adesi et al., 2008). Forecasts for $h > 1$ are obtained by iterating the model forwards (Brownlees and Souza, 2020).

Sign forecasts Similarly to absolute value forecasts, sign forecasts are constructed employing the baseline methods to forecast the signs of the target variable, with a few adjustments. First, the target variable is $Z_{T+h} = \mathbb{1}\{Y_{T+h} > 0\}$. For Sign-PCR, a logit regression based on the same factors ($\widehat{F}_T$) created to forecast Y_{T+h} is estimated. Sign-Ridge and Sign-LASSO are estimated by penalized maximum likelihood.[7] In addition to PCR, ridge and LASSO, I consider a number of machine learning algorithms to construct sign forecasts. In particular, I consider random forests (Breiman, 2001), AdaBoost (Freund and Schapire, 1995), k-nearest neighbors (Devroye et al., 1996) and neural networks (White, 2006). All tuning parameters are selected on the basis of past predictive performance.

[6] The absolute value is taken coordinate wise.

[7] The grid for λ is constructed based on the whole sample and is given by $\epsilon, \ldots, \lambda_{\max}$, where $\epsilon \approx 0$ and λ_{max} is the λ value such that all coefficients in the model are zero (see Friedman et al., 2010).

Baseline CAVS forecasts CAVS forecasts are constructed as the product of sign and absolute value forecasts, and a number of specifications may be entertained. I consider a set of baseline CAVS specifications where the same method is used to forecast signs and absolute values, and denote them as CAVS-PCR, CAVS-Ridge, and CAVS-LASSO.

Forecast combinations

In addition to baseline linear and CAVS models, I explore the performance of forecast combinations. First, I consider model selection based on past predictive performance for both CAVS and linear forecasts. The linear forecast is based on the model (PCR, ridge, or LASSO) that minimizes the MSE over the last 60 months. In contrast, model selection in the CAVS framework amounts to choosing the best forecasting strategy for each component. Component-specific models are chosen by their predictive performance in forecasting the appropriate target variable, where performance is measured by the MSE. Second, I consider equally weighted forecast combinations based exclusively on linear or CAVS forecasts. I consider 12 CAVS specifications to average over: the permutations of PCR, Ridge and LASSO, in addition to GARCH-based absolute value forecasts. In contrast, there are 3 linear specifications to average over. Finally, I consider a hybrid forecast given by the average of the two previously constructed equal weighted forecasts.

Forecast evaluation As is standard in the forecasting literature, I evaluate conditional mean forecasts by their pseudo out-of-sample MSE, defined as

$$MSE_{im} = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{Y}_{it}(m) - Y_{it}\right)^2 ,$$

where $i = 1, \ldots, n$ denotes the target series, T the number of pseudo out-of-sample observations, and $\widehat{Y}_{it}(m)$ is the forecast for Y_{it} based on method m. Diebold-Mariano tests (Diebold and Mariano, 1995) of superior predictive ability are carried out to assess whether strategies improve forecast accuracy relative to the PCR benchmark. In particular, denoting by $\epsilon_{t+h}(m)$ model m's prediction error, the null hypothesis of the DM test considered is $H_0 : \mathbb{E}[\epsilon^2_{t+h}(m)] < \mathbb{E}[\epsilon^2_{t+h}(\text{PCR})]$. The test statistic is constructed as the sample analog of $\mathbb{E}[\epsilon^2_{t+h}(m)] - \mathbb{E}[\epsilon^2_{t+h}(\text{PCR})]$, scaled by a heteroskedasticity and autocorrelation robust estimator of its standard deviation.

Directional forecasts are evaluated by the proportion of incorrect sign forecasts, defined as

$$DL_{im} = 1 - \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{\text{sign}(Y_{it}) = \text{sign}(\widehat{Y}_{it}(m))\} .$$

Assuming $H_{it}(m) = \mathbb{1}\{\text{sign}(Y_{it}) = \text{sign}(\widehat{Y}_{it}(m))\} \sim Ber(p_m)$, interest lies in verifying whether $p_m = p_{PCR}$ for the remaining forecasting strategies m. In other words, I test whether model m has the same probability of correctly classifying the sign of the next realization relative to the PCR benchmark. Following Christoffersen (1998), a likelihood ratio test is conducted to test the null hypothesis of $p_m = p_{PCR}$.

In addition to comparing forecasts across models, forecasts are compared against a benchmark constructed from the unconditional distribution of the target variable. If a model outperforms the unconditional benchmark, we say there is evidence of predictability, that is, the conditioning set improves forecast accuracy. Conditional mean predictability is assessed by DM tests of each strategy relative to the recursively estimated unconditional mean of each variable.

1.4.2 Empirical Results

Results for Selected Series

TABLE 1.1 ABOUT HERE

Table 1.1 reports the ratio of the MSE of each forecasting strategy relative to PCR for each selected series and forecast horizon. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. DM tests of superior predictive ability are carried out and stars denote significance levels. CAVS-based forecasts are the MSE best for the majority of series across all forecast horizons considered, substantially outperforming linear forecasts for the Federal funds, industrial production, nonfarm payroll employment, and the S&P 500 uniformly across forecast horizons. In particular, for the Federal funds, the best CAVS specification displays MSE reductions of about 15, 18, 6, and 3% relative to the best linear model at $h = 1, 3, 6,$ and 12 months ahead, respectively. CAVS-Ridge is the best performing specification, selected as the MSE best across all series and forecast horizon combinations 25% of the time, followed by CAVS-LASSO, PCR, and CAVS-PCR, which are the MSE best 20, 17, and 15% of the time, respectively. For $h = 1$ month ahead forecasting, Ridge is the best performing linear model for 5 out of the 10 selected series, followed by LASSO and PCR, the best performing linear models for 4 and 1 series, respectively. CAVS-Ridge is the best performing CAVS specification for 5 out of the 10 selected series, followed by CAVS-PCR and CAVS-LASSO, the best performing CAVS specifications for 3 and 2 series, respectively. This ranking remains largely unchanged for $h = 3$ and $h = 6$ months ahead forecasts. For $h = 12$ months ahead, Ridge is the best performing linear model for 7 out of 10 series, followed by PCR, the best performing model for 3 series. LASSO is not the best linear model for any series. Among CAVS specifications, CAVS-Ridge dominates CAVS-PCR and is the best CAVS specification for 7 out of 10 series. CAVS-LASSO is the best performing CAVS specification for the remaining 3 series.

Overall, the good performance of the baseline CAVS forecasts highlights that exploiting directional and volatility predictability yields more accurate macroeconomic forecasts for the selected series. In particular, CAVS-Ridge forecasts are, on average, 6.5% more accurate than PCR forecasts, displaying the largest gains among all the baseline models considered.

Results for all FRED-MD components

Next, I provide a comparison of CAVS-Ridge and PCR for all FRED-MD components.

FIGURE 1.2 ABOUT HERE

Figure 1.2 reports the ratios of the MSE of CAVS-Ridge relative to PCR for all FRED-MD components, sorted by groups. Values below 1 indicate that CAVS-Ridge forecasts outperforms PCR. Colors indicate significance at the 10% level (gray), 5% level (dark-gray), or 1% level (black), based on a DM test of superior predictive ability. A number of findings emerge from inspection of Figure 1.2. First, CAVS-Ridge outperforms PCR for the majority of series and across all horizons, with particularly strong performance for series that exhibit persistent volatility dynamics, such as interest rates, output, and stocks series. Second, as the forecast horizon increases, CAVS-Ridge performance gains become widespread, outperforming PCR for 53, 80, 67, and 75% of series for $h = 1, 3, 6$, and 12 months ahead, respectively. Finally, CAVS-Ridge — and CAVS forecasts in general — display poor performance for series that are not expressed in growth rates. In particular, none of the series that are in the top 90% quantile of MSE ratios — *i.e.*, series for which CAVS provides MSE increases greater than 12.6% — are expressed in growth rates. This is the case for all price series, most money, and a few labor series, all of which are twice differenced. Additionally, some series in the interest and exchange rate groups are expressed as spreads. Because spreads rarely change signs, CAVS forecasts generally display poor performance for these series — in stark contrast to its good performance for their growth rates counterparts. For example, CAVS-Ridge provides 33% more accurate forecasts of the first differences of the 3-month treasury bill when compared to PCR. By contrast, PCR provides 5% more accurate forecasts of the spread constructed as the 3-month treasury bill minus the federal funds rate relative to CAVS-Ridge. Overall, the results highlight that CAVS-Ridge compares favorably to PCR, particularly for series that display persistent volatility dynamics.

Componentwise Forecast Accuracy

Directional forecasts Directional forecasts of macroeconomic series are objects of interest in their own right (see Pesaran and Timmerman, 1992; Sinclair et al., 2010, among others). I compare the performance of directional forecasts implied by CAVS and linear forecasts for the selected series.

TABLE 1.2 ABOUT HERE

Table 1.2 reports the directional loss ratio of each forecasting strategy relative to the sign of the linear PCR forecast for each selected series and forecast horizon. Note that signs of CAVS forecasts directly target the sign of the series considered by construction. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. Likelihood ratio tests of superior predictive ability relative to the linear PCR benchmark, as described in Section 1.4.1, are carried out and stars denote significance levels. CAVS-based directional forecasts outperform their linear counterparts for the majority of series and across all horizons considered. In particular, CAVS-based directional forecasts are more accurate for the four series for which CAVS display the largest gains relative to linear forecasts. For most of the remaining series, CAVS-based directional forecasts perform on par with their linear counterparts. In particular, CAVS directional forecasts are outperformed by linear forecasts for the CPI and M2 series. This finding highlights that twice differencing may hamper sign predictability, particularly at short forecast horizons. Moreover, CAVS-based directional forecasts provide substantial improvements relative to their linear counterparts at intermediate forecast horizons. The performance of all CAVS specifications is similar. CAVS-Ridge is the best directional forecasting strategy for the most series across all horizons, but performance gains relative to CAVS-LASSO and CAVS-PCR are modest. Among directional forecasts based on linear models, Ridge and LASSO perform similarly at $h = 1$ month ahead forecasting. PCR is the best performing linear model for horizons greater than 1 month ahead. Overall, CAVS-based directional forecasts are more accurate than their linear counterparts, highlighting that there is sign predictability in excess of that implied by linear models for the conditional mean. In particular, the results show that directly targeting the sign yields more accurate directional forecasts than forecasts based on the sign of conditional mean forecasts.

Absolute value forecasts The performance of CAVS forecasts hinges on sign and absolute value predictability. This section reports the results for the performance of absolute value forecasts for the selected series.

TABLE 1.3 ABOUT HERE

Table 1.3 reports, for each forecasting strategy, selected series, and forecast horizon, the MSE of absolute value forecasts relative to a PCR benchmark. I report results for the absolute values of linear forecasts, as well as absolute value forecasts that directly target the absolute values of the series considered. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. DM tests of superior predictive ability are carried out and stars denote significance levels. Absolute value forecasts based on a standard AR(4)-GARCH(1,1) are selected as the MSE best forecast for the majority of series across all forecast horizons. For $h = 1$ month ahead, GARCH-based forecasts exhibit sizable performance gains relative to linear models for the absolute value. This finding highlights that standard time series models are able to accurately model conditional volatility dynamics for macroeconomic series, and that the added value of exogenous predictors is modest (see also Brownlees and Souza, 2020). As the forecast horizon increases, all models that target the absolute value perform similarly, with GARCH-based forecasts modestly outperforming their linear counterparts. Overall, Tables 1.2 and 1.3 show that there is componentwise predictability in excess of that implied by linear models.

Additional Results

Alternative CAVS specifications I explore whether alternative CAVS specifications can improve forecasting performance relative to the baseline CAVS models considered. In particular, given the good performance of the GARCH(1,1) in forecasting absolute values, I consider 4 absolute value forecasting strategies (PCR, ridge, LASSO and GARCH) and 3 sign forecasting strategies (PCR, ridge and LASSO), leading to 12 possible combinations of signs and absolute values.

TABLE 1.4 ABOUT HERE

Table 1.4 reports the MSE of CAVS-based forecasts relative to that of PCR, for each of the selected series and forecast horizons. The first and second rows denote the absolute value and sign forecasting strategy, respectively. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. DM tests of superior predictive ability relative to PCR are carried out, and stars denote significance levels. In line with Tables 1.2 and 1.3, CAVS forecasts based on the standard GARCH (the best performing absolute value model), and ridge (the best performing sign model), are the MSE best for the most series and forecast horizon combinations considered. However, performance across all CAVS models is similar, suggesting that standard CAVS specifications are sufficient to exploit the relevant nonlinearities.

Machine learning sign forecasts. I consider whether CAVS forecasts can be improved upon by employing machine learning algorithms to forecast the signs.

TABLE 1.5 ABOUT HERE

Table 1.5 displays results for selected series. For each series and forecast horizon, I report the ratios of the MSE of the CAVS specification considered relative to PCR forecasts. Best performing methods according to each criteria are highlighted in boldface, and stars represent significance according to a DM test, evaluated at the 5% significance level. The most successful machine learning algorithm is random forests. CAVS forecasts where the sign forecasts are constructed from random forests outperform baseline CAVS forecasts for 2 out of the 10 series considered. Perhaps surprisingly, no other machine learning algorithm outperforms the baseline CAVS forecasts. It is important to emphasize that I consider standard implementations of the machine learning algorithms. It is well-known that training machine learning models requires a degree of experimentation (White, 2006). This implies, in particular, that careful tuning of each model's parameters could improve accuracy.

Models against an unconditional benchmark. I compare models in terms of their ability to outperform forecasts based on the recursively estimated unconditional mean.

TABLE 1.6 ABOUT HERE

Table 1.6 reports, for each FRED-MD group and forecast horizon, the percentage of series in each group for which each strategy outperforms the unconditional mean forecasts according to a DM test at the 5% significance level. Best performing methods are highlighted in boldface. For $h = 1$ month ahead, both CAVS-based and linear forecasts outperform the unconditional mean benchmark for the majority of the series. In particular, CAVS-based forecasts outperform the unconditional mean benchmark more often than linear forecasts for labor, money, and interest rates series. In contrast, linear forecasts outperform the unconditional mean benchmark more often than CAVS-based forecasts for consumption and prices series. The performance of all models is similar in the remaining series. For $h > 1$ months ahead, CAVS-based forecasts outperform the unconditional mean more often than linear models for nearly all groups across all horizons. In particular, at $h = 6$ months ahead, CAVS-based forecasts outperform the unconditional mean for up to 40% of the Output series, in contrast to linear models, which outperform the unconditional mean for up to 25% of the series. Overall, CAVS-based forecasts outperform a unconditional mean benchmark for more series and across longer forecast horizons than linear models.

Model selection and forecast averaging. I consider the performance of forecast combinations, as described in Section 1.4.1.

TABLE 1.7 ABOUT HERE

Table 1.7 reports the MSE of each forecast combination strategy relative to that of PCR, for each selected series and forecast horizon. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. DM tests of superior predictive ability relative to PCR are carried out, and stars denote significance levels. Forecasts constructed by model averaging generally outperform those constructed by model selection for both linear and CAVS forecasts. Additionally, linear model averaging significantly outperforms PCR for nearly all series at all horizons considered. Moreover, forecast combinations that include CAVS forecasts outperform those based exclusively on linear models for the majority of series considered, and are particularly well-suited for intermediate forecast horizons. Finally, simple forecast combinations of linear and nonlinear models are a competitive strategy. Overall, the findings reported in Table 1.7 are in line with those reported in Table 1.1, suggesting that series characteristics, rather than specific modelling choices, are the determinants of whether CAVS will perform favorably compared to linear forecasts.

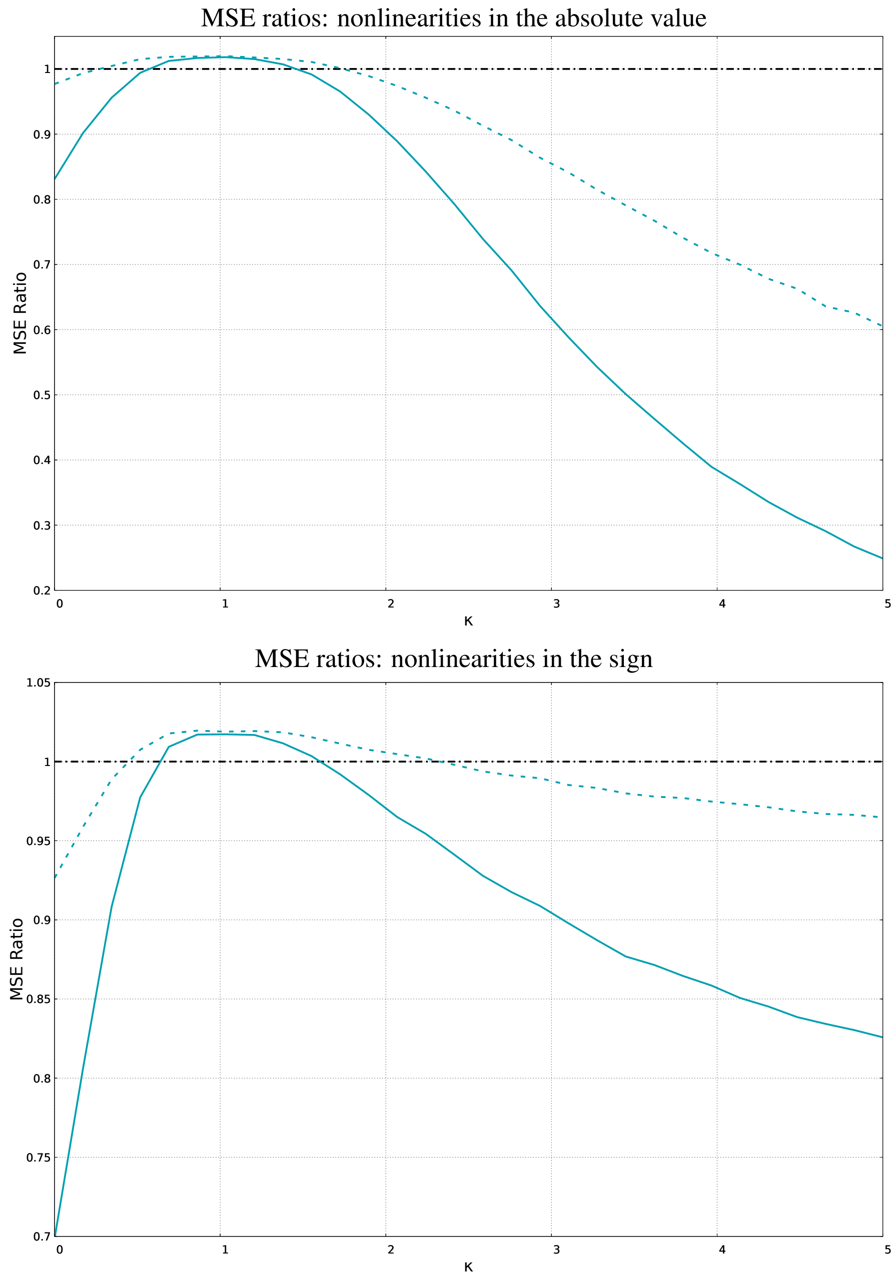
1.5 Concluding Remarks

This paper introduces CAVS forecasts, a nonlinear framework that combines forecasts of the sign and absolute value of a time series into conditional mean forecasts. In contrast to linear models, in which variables that affect the mean of the target variable must affect both its sign and absolute value, the proposed framework allows different predictors to affect either the sign, the absolute value, or both.

I provide a number of theoretical results for CAVS forecasts. First, I show that the conditional mean can be written as the product of MSE optimal forecasts of signs and absolute values and a covariance term that can be explicitly modeled. If the underlying DGP is additive with symmetric shocks, the covariance term is small relative to the variance of the shocks and therefore may be ignored for forecasting purposes. Second, I show that the performance of CAVS forecasts hinges on the ability to accurately forecast each of the components of the target series. This result highlights that CAVS-based forecasts are particularly suited for series that exhibit persistent volatility dynamics, and hence absolute value predictability. Third, I study a nonlinear DGP in which variables may affect signs and absolute values differently, and show that the MSE of the best linear predictor increases quadratically with the degree of nonlinearity.

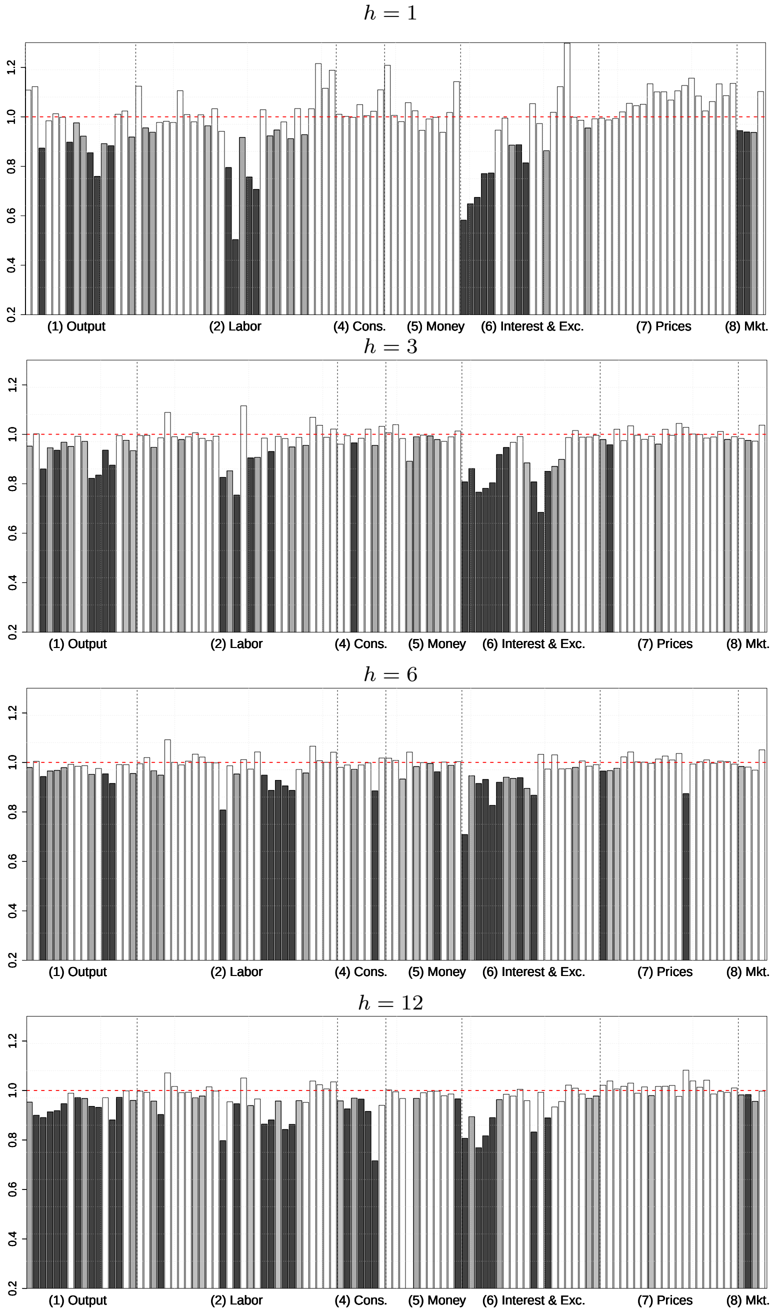
The proposed methodology is applied to forecast each of the components of the FRED-MD dataset. I find that CAVS forecasts substantially outperform linear forecasts in series that exhibit strong conditional volatility dynamics, such as Output and Interest Rate series, and the performance gains remain sizable across forecast horizons. Moreover, I find that CAVS-based directional forecasts outperform linear forecasts for the majority of the selected series considered, across all horizons. Additionally, I document that CAVS forecasts outperform the recursively estimated unconditional mean benchmark for more series and across longer horizons than linear forecasts. Finally, I find that forecast combinations that include CAVS forecasts outperform those based exclusively on linear models for the majority of series and across all horizons considered. Overall, the empirical application highlights that exploiting directional and volatility predictability improves forecast accuracy in macroeconomic series.

Figura 1.1: Simulation Study



This figure illustrates the findings of the simulation study. It depicts the MSE ratios of each strategy considered over the linear benchmark. The x-axis represents distance to linearity, which increases with $|\kappa - 1|$. Solid lines represent sign forecasts based on correctly specified models, whereas dashed lines represent sign forecasts based on a misspecified probit regression. Values below 1 indicate that the MSE of a given strategy is smaller than that of the linear forecast.

Figura 1.2: MSE ratios for all FRED-MD components



This figure reports the ratio of the MSE of the equally weighted CAVS forecast relative to that of the equally weighted linear forecast. Values below 1 indicate that CAVS forecasts outperform PCR forecasts. Colors indicate that CAVS combinations outperform linear combinations at the 10% level (gray), 5% level(dark-gray), or 1% level (black), based on a one-sided DM test.

Taula 1.1: Linear and CAVS Forecasts

		Linear		*CAVS*		
h	Series	Ridge	LASSO	PCR	Ridge	LASSO
1	Fed Funds	0.725^{***}	0.662^{***}	0.692^{**}	0.582^{***}	$\mathbf{0.573}^{***}$
	Ind. Prod.	0.930^{***}	0.982	0.891^{***}	$\mathbf{0.874}^{***}$	0.894^{***}
	Nonfarm Empl.	0.985	1.023	$\mathbf{0.922}^{*}$	0.941	0.926^{*}
	S&P 500	0.954^{***}	0.958^{**}	1.011	$\mathbf{0.944}^{***}$	0.961^{**}
	Unemp.	0.980	0.987	**0.946**	0.982	1.014
	M2 (Real)	0.972	0.968	1.000	0.981	**0.960**
	PPI: FG	0.967^{*}	$\mathbf{0.941}^{**}$	1.134	0.995	0.995
	10-Year T. Rate	$\mathbf{0.948}^{**}$	0.969	1.039	0.995	1.006
	CPI	0.902^{***}	$\mathbf{0.874}^{***}$	1.224	1.051	1.058
	RPI ex. Rec.	1.019	1.038	1.069	1.122	1.138
3	Fed Funds	0.940^{*}	0.971	$\mathbf{0.763}^{***}$	0.808^{***}	0.818^{***}
	Ind. Prod.	0.912^{***}	0.960	0.878^{***}	$\mathbf{0.860}^{***}$	0.874^{***}
	Nonfarm Empl.	0.897^{***}	0.947	0.854^{***}	$\mathbf{0.826}^{***}$	0.856^{***}
	S&P 500	0.988^{**}	0.986^{**}	1.000	0.983	$\mathbf{0.980}^{*}$
	Unemp.	1.022	1.051	1.047	1.088	1.100
	M2 (Real)	$\mathbf{0.963}^{*}$	0.966	0.987	0.983	0.992
	PPI: FG	0.997	1.000	$\mathbf{0.975}^{**}$	0.979^{**}	0.978^{**}
	10-Year T. Rate	0.970^{*}	0.972	0.966^{*}	$\mathbf{0.947}^{***}$	0.951^{**}
	CPI	0.972^{**}	$\mathbf{0.971}^{**}$	0.987	0.980	0.984
	RPI ex. Rec.	1.006	1.020	**0.990**	1.002	1.017
6	Fed Funds	0.739^{***}	0.815^{***}	$\mathbf{0.690}^{***}$	0.708^{***}	0.727^{***}
	Ind. Prod.	0.980	0.994	0.950^{**}	$\mathbf{0.943}^{***}$	0.964^{**}
	Nonfarm Empl.	0.886^{***}	0.915^{**}	0.853^{***}	0.807^{***}	$\mathbf{0.788}^{***}$
	S&P 500	0.994	0.992	0.991	0.983^{*}	$\mathbf{0.982}^{*}$
	Unemp.	1.006	1.011	1.107	1.091	1.105
	M2 (Real)	$\mathbf{0.916}^{***}$	0.954^{*}	0.953^{*}	0.933^{*}	0.932^{*}
	PPI: FG	0.984^{**}	0.984^{**}	0.985	0.965^{***}	$\mathbf{0.964}^{**}$
	10-Year T. Rate	0.973	0.970	0.973	$\mathbf{0.940}^{**}$	0.955
	CPI	0.998	**0.996**	1.013	1.001	1.006
	RPI ex. Rec.	1.002	1.003	1.002	1.004	1.015
12	Fed Funds	0.839^{***}	0.902^{**}	0.826^{***}	$\mathbf{0.807}^{***}$	0.814^{***}
	Ind. Prod.	0.921^{***}	0.947^{***}	0.923^{***}	$\mathbf{0.890}^{***}$	0.892^{***}
	Nonfarm Empl.	0.979	1.051	0.802^{***}	$\mathbf{0.797}^{***}$	0.805^{***}
	S&P 500	0.991^{**}	0.992^{*}	0.989	0.982^{**}	$\mathbf{0.982}^{**}$
	Unemp.	1.023	1.019	1.081	1.071	1.094
	M2 (Real)	$\mathbf{0.924}^{***}$	0.952^{**}	1.014	0.968	0.992
	PPI: FG	1.015	1.022	1.025	1.021	1.026
	10-Year T. Rate	$\mathbf{0.983}^{*}$	0.990	1.015	0.985	0.991
	CPI	1.001	1.008	1.027	1.015	1.015
	RPI ex. Rec.	0.969^{**}	0.973^{*}	0.909^{***}	0.900^{***}	$\mathbf{0.897}^{***}$

This table reports the MSE of each forecasting strategy (columns) relative to that of PCR, for each of the selected series and forecast horizon. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. If no method is highlighted, the PCR benchmark is the best performing method. DM tests of superior predictive ability relative to the PCR are carried out, and stars denote significance levels. ($^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$).

Taula 1.2: Directional Forecasts

		Linear			*CAVS: Sign*		
h	Series	PCR(%DL)	Ridge	LASSO	PCR	Ridge	LASSO
1	Fed Funds	40.24	0.994	0.970	0.994	0.976	**0.947**
	Ind. Prod.	31.19	0.947	0.985	**0.924**	1.023	1.008
	Nonfarm Empl.	11.43	0.958	0.938	1.000	0.938	**0.875**
	S&P 500	36.67	0.968	0.942	0.994	**0.935**	0.942
	Unemp.	57.38	1.004	1.021	**0.938**	0.967	0.992
	M2 (Real)	26.67	0.893	**0.884**	1.045	0.911	0.902
	PPI: FG	32.86	1.051	1.014	1.159	1.014	**0.971**
	10-Year T. Rate	40.95	1.023	1.023	1.012	**0.988**	1.017
	CPI	34.52	**0.945**	0.952	1.200	0.966	1.014
	RPI ex. Rec.	25.71	0.991	0.972	0.963	**0.954**	0.954
3	Fed Funds	50.48	1.100	1.261	**0.853*****	0.915*	0.872***
	Ind. Prod.	33.49	1.000	1.007	**0.914**	1.007	1.014
	Nonfarm Empl.	13.16	1.036	1.036	1.073	1.055	1.036
	S&P 500	40.19	0.976	0.970	0.946	0.917	**0.887***
	Unemp.	54.31	1.018	1.035	1.048	1.057	1.026
	M2 (Real)	29.43	1.016	0.992	**0.927**	0.976	0.976
	PPI: FG	51.91	1.018	1.037	0.977	**0.954**	0.954
	10-Year T. Rate	52.87	1.109	1.059	0.977	**0.941**	0.946
	CPI	51.20	1.000	1.009	**0.855*****	0.869***	0.855***
	RPI ex. Rec.	24.64	1.068	1.049	1.049	1.078	1.087
6	Fed Funds	52.53	1.009	1.018	**0.917***	0.931	1.028
	Ind. Prod.	35.66	0.980	0.986	**0.966**	0.966	0.973
	Nonfarm Empl.	15.90	1.030	1.076	0.955	**0.939**	1.015
	S&P 500	38.55	1.000	0.975	0.981	0.975	**0.950**
	Unemp.	56.14	1.017	1.000	1.052	1.056	1.056
	M2 (Real)	28.19	1.051	1.085	1.068	1.000	1.017
	PPI: FG	52.29	0.982	1.000	0.940	**0.931**	0.949
	10-Year T. Rate	50.84	1.043	0.962	0.976	**0.948**	0.957
	CPI	51.57	1.000	0.977	**0.897****	0.897**	0.902**
	RPI ex. Rec.	24.82	1.000	0.981	**0.971**	0.990	1.000
12	Fed Funds	50.86	1.014	1.202	1.067	1.091	1.168
	Ind. Prod.	35.70	**0.966**	0.966	0.966	0.966	0.966
	Nonfarm Empl.	20.29	0.928	0.928	0.940	**0.916**	0.928
	S&P 500	37.65	0.974	1.006	0.974	**0.961**	0.987
	Unemp.	58.19	1.071	1.042	1.042	1.034	1.038
	M2 (Real)	32.27	**0.955**	1.053	0.977	0.992	0.992
	PPI: FG	48.41	0.970	0.995	**0.914***	0.944	0.934
	10-Year T. Rate	46.45	1.074	1.021	1.095	1.105	1.084
	CPI	46.70	1.000	1.031	0.932	**0.901***	0.916
	RPI ex. Rec.	26.41	**0.981**	1.009	0.981	0.981	1.009

This table reports, for each selected series and forecast horizon, the directional loss ratio of each sign forecasting strategy relative to the sign of the linear PCR forecast. Linear sign forecasts are the signs of the linear forecasts constructed, whereas the remaining forecasts are obtained by directly targeting the sign of the series considered. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. If no method is highlighted, the PCR benchmark is the best performing method. Likelihood ratio tests of superior predictive ability relative to the linear PCR benchmark, as described in Section 1.4.1, are carried out. Stars denote significance levels. $({}^{*}p < 0.1, {}^{**}p < 0.05, {}^{***}p < 0.01)$.

Taula 1.3: Absolute Value Forecasts

		Linear		*CAVS: Absolute Value*			
h	Series	Ridge	LASSO	PCR	Ridge	LASSO	GARCH
1	Fed Funds	0.679^{***}	0.590^{***}	0.862	0.752^{***}	0.699^{***}	$\mathbf{0.568}^{***}$
	Ind. Prod.	1.056	1.076	0.778^{***}	$\mathbf{0.714}^{***}$	0.741^{***}	0.807^{***}
	Nonfarm Empl.	0.980	1.064	0.981	0.994	1.000	$\mathbf{0.917}^{*}$
	S&P 500	1.006	1.030	0.639^{***}	0.636^{***}	$\mathbf{0.633}^{***}$	0.637^{***}
	Unemp.	1.086	1.111	0.862^{***}	$\mathbf{0.849}^{***}$	0.860^{***}	0.856^{***}
	M2 (Real)	1.007	1.014	0.910^{*}	0.902^{**}	$\mathbf{0.890}^{**}$	0.925^{*}
	PPI: FG	0.955^{**}	0.927^{**}	0.805^{***}	0.794^{***}	0.803^{***}	$\mathbf{0.700}^{***}$
	10-Year T. Rate	1.006	1.002	0.833^{***}	0.809^{***}	0.797^{***}	$\mathbf{0.738}^{***}$
	CPI	0.910^{**}	0.894^{**}	0.869	0.838^{**}	0.866^{*}	$\mathbf{0.731}^{***}$
	RPI ex. Rec.	0.994	1.007	0.768^{**}	0.760^{**}	$\mathbf{0.751}^{**}$	0.854
3	Fed Funds	0.946	0.935	0.941	0.890	0.955	$\mathbf{0.810}^{**}$
	Ind. Prod.	1.005	0.952^{*}	0.775^{***}	$\mathbf{0.760}^{***}$	0.777^{***}	0.762^{***}
	Nonfarm Empl.	0.921^{*}	1.019	1.017	0.975	0.981	$\mathbf{0.908}^{*}$
	S&P 500	1.017	1.009	0.586^{***}	0.582^{***}	0.589^{***}	$\mathbf{0.578}^{***}$
	Unemp.	1.055	1.072	$\mathbf{0.704}^{***}$	0.715^{***}	0.725^{***}	0.708^{***}
	M2 (Real)	0.984	0.981	0.829^{***}	$\mathbf{0.806}^{***}$	0.812^{***}	0.857^{***}
	PPI: FG	1.001	0.986^{**}	0.537^{***}	0.512^{***}	0.515^{***}	$\mathbf{0.465}^{***}$
	10-Year T. Rate	1.092	1.091	0.666^{***}	0.627^{***}	0.626^{***}	$\mathbf{0.564}^{***}$
	CPI	1.023	1.017	0.580^{***}	0.563^{***}	0.578^{***}	$\mathbf{0.555}^{***}$
	RPI ex. Rec.	0.998	1.005	0.911^{***}	0.905^{***}	$\mathbf{0.904}^{***}$	1.010
6	Fed Funds	0.975	0.938	0.913	$\mathbf{0.909}$	0.995	0.951
	Ind. Prod.	0.966^{**}	0.954^{**}	0.780^{***}	$\mathbf{0.779}^{***}$	0.785^{***}	0.810^{***}
	Nonfarm Empl.	0.891^{***}	0.840^{***}	0.895^{*}	0.866^{**}	0.866^{**}	$\mathbf{0.835}^{***}$
	S&P 500	1.019	1.031	0.588^{***}	0.579^{***}	0.581^{***}	$\mathbf{0.571}^{***}$
	Unemp.	1.079	1.087	0.666^{***}	0.652^{***}	$\mathbf{0.648}^{***}$	0.652^{***}
	M2 (Real)	0.952^{*}	0.977	0.813^{***}	0.783^{***}	$\mathbf{0.781}^{***}$	0.805^{***}
	PPI: FG	1.024	1.029	0.540^{***}	0.536^{***}	0.536^{***}	$\mathbf{0.491}^{***}$
	10-Year T. Rate	0.950^{**}	0.967^{*}	0.622^{***}	0.613^{***}	0.649^{***}	$\mathbf{0.577}^{***}$
	CPI	1.016	1.011	0.583^{***}	0.576^{***}	$\mathbf{0.576}^{***}$	0.582^{***}
	RPI ex. Rec.	0.995^{*}	0.986^{***}	0.915^{***}	$\mathbf{0.914}^{***}$	0.921^{***}	0.954
12	Fed Funds	1.018	1.002	1.035	1.088	1.301	1.199
	Ind. Prod.	1.047	1.026	0.948	$\mathbf{0.894}^{**}$	0.906^{**}	0.952
	Nonfarm Empl.	0.788^{***}	$\mathbf{0.777}^{***}$	0.866^{**}	0.861^{**}	0.860^{**}	0.842^{***}
	S&P 500	1.018	1.013	0.609^{***}	0.597^{***}	0.602^{***}	$\mathbf{0.596}^{***}$
	Unemp.	1.079	1.051	0.672^{***}	0.651^{***}	0.649^{***}	$\mathbf{0.642}^{***}$
	M2 (Real)	0.940^{***}	0.974	0.885^{***}	0.848^{***}	0.844^{***}	$\mathbf{0.820}^{***}$
	PPI: FG	1.040	1.042	0.608^{***}	0.602^{***}	0.609^{***}	$\mathbf{0.581}^{***}$
	10-Year T. Rate	1.052	1.039	0.596^{***}	0.571^{***}	0.592^{***}	$\mathbf{0.563}^{***}$
	CPI	1.003	0.997	0.641^{***}	$\mathbf{0.622}^{***}$	0.622^{***}	0.630^{***}
	RPI ex. Rec.	1.033	1.048	0.964	0.956	$\mathbf{0.951}$	0.989

This table reports the MSE of absolute value forecasts, for each forecasting strategy relative to that of PCR, for each of the selected series and forecast horizons. Absolute values of linear models are the absolute values of the linear forecasts constructed. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. If no method is highlighted, the PCR benchmark is the best performing method. DM tests of superior predictive ability relative to the PCR are carried out and stars denote significance levels. $({}^{*}p < 0.1, {}^{**}p < 0.05, {}^{***}p < 0.01)$.

Taula 1.4: Additional CAVS Specifications

	Absolute Value:	*PCR*			*Ridge*			*LASSO*			*GARCH*		
h	Series	PCR	Ridge	LASSO	PCR	Ridge	LASSO	PCR	Ridge	LASSO	PCR	Ridge	LASSO
1	Fed Funds	0.692**	0.653***	0.687**	0.587***	0.582***	0.584***	0.569***	0.585***	0.573***	0.578***	0.579***	**0.561*****
	Ind. Prod.	0.891***	0.876***	0.898***	0.877***	0.874***	0.903***	0.867***	**0.865*****	0.894***	0.900***	0.899***	0.926**
	Nonfarm Empl.	0.922*	0.975	0.963	**0.898****	0.941	0.940	0.901**	0.931*	0.926*	0.923*	0.972	0.963
	S&P 500	1.011	0.946***	0.962**	1.008	0.944***	0.960**	1.004	0.944***	0.961**	1.014	**0.941*****	0.956**
	Unemp.	**0.946**	0.970	1.004	0.956	0.982	1.011	0.961	0.986	1.014	0.986	1.001	1.027
	M2 (Real)	1.000	0.981	0.970	0.997	0.981	0.971	0.988	0.969	**0.960**	1.018	1.014	1.002
	PPI: FG	1.134	1.001	1.002	1.142	0.995	0.996	1.140	0.995	0.995	1.112	**0.969**	0.972
	10-Year T. Rate	1.039	0.995	1.006	1.034	0.995	1.009	1.026	0.993	1.006	1.020	**0.980**	0.996
	CPI	1.224	1.108	1.133	1.193	1.051	1.068	1.191	1.038	1.058	1.149	1.044	1.055
	RPI ex. Rec.	1.069	1.115	1.128	1.074	1.122	1.136	1.077	1.126	1.138	1.128	1.205	1.217
3	Fed Funds	**0.763*****	0.786***	0.793***	0.791***	0.808***	0.814***	0.794***	0.814***	0.818***	0.774***	0.783***	0.793***
	Ind. Prod.	0.878***	0.867***	0.884***	0.870***	0.860***	0.878***	0.868***	0.857***	0.874***	0.850***	**0.839*****	0.859***
	Nonfarm Empl.	0.854***	0.864***	0.890**	0.810***	0.826***	0.856***	**0.810*****	0.826***	0.856***	0.822***	0.847***	0.880**
	S&P 500	1.000	0.984	0.981*	0.998	0.983	0.980*	0.997	0.983	**0.980***	0.998	0.983	0.980*
	Unemp.	1.047	1.075	1.089	1.064	1.088	1.099	1.065	1.090	1.100	1.066	1.089	1.101
	M2 (Real)	0.987	0.998	1.006	0.973	0.983	0.990	**0.973**	0.985	0.992	0.985	0.994	1.001
	PPI: FG	**0.975****	0.978**	0.979**	0.977**	0.979**	0.979**	0.977*	0.978**	0.978**	0.983	0.985	0.986
	10-Year T. Rate	0.966*	0.945***	0.951**	0.970*	0.947***	0.954**	0.962**	**0.943*****	0.951**	0.971*	0.951**	0.959**
	CPI	0.987	0.983	0.982	**0.978**	0.980	0.980	0.988	0.986	0.984	0.996	0.993	0.992
	RPI ex. Rec.	0.990	1.000	1.006	**0.989**	1.002	1.009	0.990	1.006	1.017	1.044	1.056	1.068
6	Fed Funds	**0.690*****	0.707***	0.730***	0.694***	0.708***	0.726***	0.700***	0.711***	0.727***	0.704***	0.721***	0.736***
	Ind. Prod.	0.950**	0.948***	0.977	0.944***	0.943***	0.971*	0.941***	0.938***	0.964**	0.940***	**0.934*****	0.963**
	Nonfarm Empl.	0.853***	0.824***	0.805***	0.835***	0.807***	0.790***	0.833***	0.806***	**0.788*****	0.827***	0.802***	0.796***
	S&P 500	0.991	0.984	0.981*	0.990	0.983*	**0.981***	0.991	0.984*	0.982*	0.988	0.984*	0.982*
	Unemp.	1.107	1.098	1.114	1.096	1.091	1.104	1.095	1.093	1.105	1.089	1.087	1.098
	M2 (Real)	0.953*	0.955	0.967	0.930*	0.933*	0.944	**0.917****	0.920*	0.932*	0.931*	0.930*	0.941*
	PPI: FG	0.985	0.966***	0.967**	0.982	0.965***	0.966**	0.984	**0.964*****	0.964**	0.986	0.966***	0.969**
	10-Year T. Rate	0.973	0.940**	0.965	0.971	0.940**	0.959	0.963	**0.937****	0.955	0.975	0.942*	0.957
	CPI	1.013	1.000	1.006	1.012	1.001	1.007	1.010	1.001	1.006	1.025	1.012	1.023
	RPI ex. Rec.	1.002	1.001	1.004	1.004	1.004	1.007	1.011	1.012	1.015	1.037	1.039	1.041
12	Fed Funds	0.826***	0.815***	0.805***	0.813***	0.807***	**0.798*****	0.844***	0.818***	0.814***	0.823***	0.813***	0.815***
	Ind. Prod.	0.923***	0.894***	0.901***	0.920***	0.890***	0.898***	0.908***	**0.885*****	0.892***	0.926***	0.893***	0.910***
	Nonfarm Empl.	0.802***	**0.792*****	0.798***	0.805***	0.797***	0.804***	0.805***	0.797***	0.805***	0.845***	0.842***	0.856***
	S&P 500	0.989	0.986*	0.986*	0.985*	0.982**	0.983**	0.984**	0.981**	0.982**	0.980**	**0.977*****	0.978***
	Unemp.	1.081	1.070	1.092	1.082	1.071	1.091	1.084	1.073	1.094	1.064	1.054	1.070
	M2 (Real)	1.014	0.982	1.011	0.989	0.968	0.995	0.986	0.964	0.992	0.968	**0.946***	0.974
	PPI: FG	1.025	1.027	1.030	1.019	1.021	1.025	1.019	1.022	1.026	1.013	1.020	1.025
	10-Year T. Rate	1.015	0.988	0.991	1.018	**0.985**	0.988	1.016	0.987	0.991	1.010	0.986	0.990
	CPI	1.027	1.019	1.019	1.021	1.015	1.014	1.023	1.016	1.015	1.026	1.020	1.020
	RPI ex. Rec.	0.909***	0.896***	**0.891*****	0.911***	0.900***	0.895***	0.914***	0.903***	0.897***	0.940*	0.929**	0.923**

This table reports the MSE of CAVS-based forecasts relative to that of PCR, for each of the selected series and forecast horizons. The first row denotes the Absolute Value forecasting strategy, and the second row the sign forecasting strategy considered. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. If no method is highlighted, the PCR benchmark is the best performing method. DM tests of superior predictive ability relative to the PCR are carried out, and stars denote significance levels. ($^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$).

Taula 1.5: Machine Learning Forecasts

Series	PCR	Ridge	LASSO	RF	Adab.	k-NN	NN
Fed Funds	0.587	0.582	0.584	**0.473**	0.540	0.577	0.829
Ind. Prod.	0.877	**0.874**	0.903	0.880	0.892	0.937	1.115
Nonfarm Empl.	**0.898**	0.941	0.940	0.903	1.154	1.197	1.581
S&P 500	1.008***	**0.944*****	0.960***	0.986***	1.043***	0.988*	1.076
Unemp.	**0.956**	0.982	1.011	0.977	1.017	1.043	1.260
M2 (Real)	0.997	0.981	**0.971**	1.095	1.075	1.119	1.113
PPI: FG	1.142	**0.995**	0.996	1.215	1.256	1.233	1.173
10-Year T. Rate	1.034	**0.995**	1.009	1.018	1.025	1.036	1.251
CPI	1.193	1.051	1.068	1.107	1.123	1.165	1.212
RPI ex. Rec.	1.074	1.122	1.136	1.015	1.028	1.055	1.219

This table reports the results for select series. For each series, the table reports the ratios of the MSE of the CAVS specification considered relative to PCR forecasts. The absolute values employed in this CAVS forecast are obtained by Ridge, and the sign forecasts are based on the methods in the columns. Best performing methods are highlighted in boldface and stars represent significance of a superior predictive ability test relative to the PCR forecast at 5% significance level. If no method is highlighted, the PCR benchmark is the best performing method.

Taula 1.6: Predictability

		Linear			*CAVS*		
h	Series	PCR	Ridge	LASSO	PCR	Ridge	LASSO
1	Output	37.50	**68.75**	50.00	**68.75**	**68.75**	**68.75**
	Consumption	**85.71**	**85.71**	71.43	57.14	71.43	71.43
	Labor	82.76	89.66	82.76	**93.10**	**93.10**	86.21
	Money	50.00	57.14	57.14	35.71	**64.29**	57.14
	Stocks	25.00	**100.00**	75.00	25.00	**100.00**	**100.00**
	Interest and Exc.	45.00	55.00	45.00	60.00	**95.00**	85.00
	Prices	85.00	**95.00**	**95.00**	45.00	60.00	65.00
	FRED-MD	63.64	77.27	69.09	62.73	**79.09**	75.45
3	Output	12.50	43.75	31.25	56.25	**62.50**	50.00
	Consumption	28.57	28.57	28.57	**42.86**	**42.86**	**42.86**
	Labor	58.62	65.52	62.07	68.97	**72.41**	65.52
	Money	0.00	**7.14**	**7.14**	**7.14**	**7.14**	**7.14**
	Stocks	0.00	0.00	0.00	0.00	0.00	0.00
	Interest and Exc.	30.00	35.00	35.00	**40.00**	35.00	35.00
	Prices	0.00	0.00	0.00	0.00	0.00	0.00
	FRED-MD	24.55	32.73	30.00	37.27	**38.18**	34.55
6	Output	18.75	25.00	18.75	25.00	**43.75**	37.50
	Consumption	**28.57**	**28.57**	**28.57**	**28.57**	**28.57**	**28.57**
	Labor	68.97	**75.86**	65.52	68.97	72.41	72.41
	Money	0.00	0.00	0.00	0.00	0.00	0.00
	Stocks	0.00	0.00	0.00	0.00	0.00	0.00
	Interest and Exc.	25.00	30.00	30.00	**35.00**	**35.00**	25.00
	Prices	0.00	0.00	0.00	0.00	0.00	0.00
	FRED-MD	27.27	30.91	27.27	30.00	**33.64**	30.91
12	Output	0.00	6.25	6.25	6.25	**37.50**	31.25
	Consumption	0.00	14.29	0.00	**28.57**	**28.57**	**28.57**
	Labor	31.03	37.93	27.59	41.38	**48.28**	41.38
	Money	**7.14**	**7.14**	**7.14**	**7.14**	**7.14**	**7.14**
	Stocks	0.00	0.00	0.00	0.00	0.00	0.00
	Interest and Exc.	25.00	25.00	20.00	**30.00**	25.00	25.00
	Prices	**5.00**	**5.00**	0.00	**5.00**	**5.00**	0.00
	FRED-MD	14.55	18.18	12.73	20.91	**26.36**	22.73

This table reports the results for each group in the FRED-MD, as well as for the dataset as a whole. For each group and forecast horizon, this table reports the percentage of series within a group for which each method outperforms an unconditional mean benchmark according to a DM test at the 5% significance level. Best performing methods for each group are highlighted in boldface.

Taula 1.7: Forecast Combinations

h	Series	*Model Selection* Linear	*Model Selection* CAVS	*Forecast Averaging* Linear	*Forecast Averaging* CAVS	*Forecast Averaging* *Hybrid*
1	Fed Funds	1.000	0.580^{***}	0.719^{***}	0.574^{***}	$\mathbf{0.538}^{***}$
	Ind. Prod.	0.970^{**}	0.880^{***}	0.951^{***}	$\mathbf{0.867}^{***}$	0.891^{***}
	Nonfarm Empl.	1.000	0.946	0.953^{**}	$\mathbf{0.892}^{**}$	0.901^{***}
	S&P 500	0.969^{*}	$\mathbf{0.948}^{***}$	0.964^{***}	0.960^{***}	0.951^{***}
	Unemp.	0.987	0.996	0.969^{**}	0.973	$\mathbf{0.946}^{**}$
	M2 (Real)	0.978	1.001	0.967^{**}	**0.956**	0.973
	PPI: FG	$\mathbf{0.941}^{**}$	0.981	0.961^{**}	1.010	0.957^{**}
	10-Year T. Rate	0.994	1.001	0.954^{***}	1.000	$\mathbf{0.940}^{**}$
	CPI	$\mathbf{0.874}^{***}$	1.066	0.907^{***}	1.084	0.946
	RPI ex. Rec.	1.000	1.080	1.015	1.103	1.023
3	Fed Funds	1.000	0.792^{***}	0.930^{**}	$\mathbf{0.785}^{***}$	0.842^{***}
	Ind. Prod.	0.951^{***}	0.873^{***}	0.939^{***}	$\mathbf{0.863}^{***}$	0.897^{***}
	Nonfarm Empl.	1.003	$\mathbf{0.822}^{***}$	0.910^{***}	0.828^{***}	0.843^{***}
	S&P 500	0.995	0.987	0.990^{**}	**0.986**	0.986^{*}
	Unemp.	1.002	1.061	1.018	1.076	1.023
	M2 (Real)	0.973	0.982	0.969^{*}	0.981	$\mathbf{0.964}^{*}$
	PPI: FG	1.002	0.983	0.998	$\mathbf{0.977}^{**}$	0.986^{**}
	10-Year T. Rate	0.991	0.967^{**}	0.973^{**}	$\mathbf{0.950}^{**}$	0.962^{***}
	CPI	1.000	0.986	0.975^{**}	0.979	$\mathbf{0.973}^{**}$
	RPI ex. Rec.	1.003	**0.981**	1.006	0.998	0.989
6	Fed Funds	0.807^{***}	0.717^{***}	0.796^{***}	$\mathbf{0.705}^{***}$	0.730^{***}
	Ind. Prod.	0.980	0.949^{**}	0.985	$\mathbf{0.945}^{***}$	0.957^{***}
	Nonfarm Empl.	0.902^{***}	0.817^{***}	0.904^{***}	$\mathbf{0.796}^{***}$	0.818^{***}
	S&P 500	1.000	0.986	0.994	$\mathbf{0.983}^{*}$	0.987^{*}
	Unemp.	1.003	1.089	**0.996**	1.096	1.021
	M2 (Real)	$\mathbf{0.918}^{***}$	0.935^{*}	0.949^{***}	0.936^{*}	0.931^{**}
	PPI: FG	0.999	0.970^{**}	0.988^{**}	$\mathbf{0.970}^{**}$	0.975^{***}
	10-Year T. Rate	1.002	0.952^{*}	0.965^{**}	0.949^{*}	$\mathbf{0.946}^{**}$
	CPI	0.998	1.008	0.997	1.002	**0.993**
	RPI ex. Rec.	1.004	1.015	1.002	1.006	1.005
12	Fed Funds	1.000	0.820^{***}	0.869^{***}	$\mathbf{0.807}^{***}$	0.829^{***}
	Ind. Prod.	0.953^{***}	0.909^{***}	0.943^{***}	$\mathbf{0.896}^{***}$	0.920^{***}
	Nonfarm Empl.	0.997	0.791^{***}	0.984	$\mathbf{0.789}^{***}$	0.857^{***}
	S&P 500	0.996^{*}	$\mathbf{0.978}^{***}$	0.994^{**}	0.984^{**}	0.984^{***}
	Unemp.	1.019	1.076	1.006	1.077	1.021
	M2 (Real)	$\mathbf{0.929}^{***}$	0.948^{*}	0.948^{***}	0.984	0.938^{***}
	PPI: FG	1.000	1.023	1.010	1.020	1.010
	10-Year T. Rate	0.994	0.992	0.987^{*}	0.993	**0.984**
	CPI	1.003	1.014	1.002	1.018	**0.999**
	RPI ex. Rec.	0.980	0.911^{**}	0.977^{**}	$\mathbf{0.899}^{***}$	0.935^{***}

This table reports the MSE of each forecast combination strategy (columns) relative to that of PCR, for each of the selected series and forecast horizon. Numbers below 1 imply that the strategy considered outperforms PCR. Best performing strategies for each series and forecast horizon are highlighted in boldface. If no method is highlighted, the PCR benchmark is the best performing method. DM tests of superior predictive ability relative to the PCR are carried out, and stars denote significance level. ($^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$).

Apèndix A

Appendix

A.1 Proofs

Proof of proposition 1. We need to show that

$$\mathbb{E}\Big[\Big(Y_t - \mu_{\text{CAVS}}(X_t)\Big)^2\Big] \le \sigma_u^2 + a_1\mathbb{E}\Big[\Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2\Big]$$

where we assume that $\mu_A(X_t) = \mathbb{E}[|Y_t||X_t] \le \tilde{a}_1$ for all X_t. We begin by expanding

$$\begin{aligned}
\mathbb{E}\Big[\Big(Y_t - \mu_{\text{CAVS}}(X_t)\Big)^2\Big] &= \sigma_u^2 + \mathbb{E}\Big[\Big(\mu(X_t) - m_A^*(X_t)m_S^*(X_t) - c(X_t)\Big)^2\Big] \\
&= \sigma_u^2 + \mathbb{E}\Big[\Big(\mu_A(X_t)\mu_S(X_t) - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] \\
&= \sigma_u^2 + \mathbb{E}\Big[\Big(\mu_A(X_t)\big(\mu_S(X_t) - m_S^*(X_t)\big) + m_S^*(X_t)\big(\mu_A(X_t) - m_A^*(X_t)\big)\Big)^2\Big] \\
&= \sigma_u^2 + \mathbb{E}\Big[||M'\varepsilon||^2\Big]
\end{aligned}$$

with $M = (\mu_A(X_t)\ ,\ m_S^*(X_t))'$ and $\varepsilon = \Big(\mu_S(X_t) - m_S^*(X_t)\ ,\ \mu_A(X_t) - m_A^*(X_t)\Big)'$. Noting that

$$\begin{aligned}
||M||^2 &= \mu_A(X_t)^2 + m_S^*(X_t)^2, \text{ and} \\
||\varepsilon||^2 &= \Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2,
\end{aligned}$$

we can apply Cauchy-Schwarz's inequality to obtain

$$\begin{aligned}
||M'\varepsilon||^2 &\le ||M||^2||\varepsilon||^2 \\
&\le (1 + \mu_A(X_t)^2)||\varepsilon||_2^2 \\
&= \big(1 + \mu_A(X_t)^2\big)\Big(\big(\mu_S(X_t) - m_S^*(X_t)\big)^2 + \big(\mu_A(X_t) - m_A^*(X_t)\big)^2\Big)
\end{aligned}$$

where we have used the fact that $|m_S^*(X_t)| \leq 1$. Next, because $\mu_A(X_t) \leq c$ for all X_t by assumption, we have

$$\begin{aligned}
\mathbb{E}\Big[\Big(Y_t - \mu_{\text{CAVS}}(X_t)\Big)^2\Big] &\leq \sigma_u^2 + \mathbb{E}\Big[\big(1+\mu_A(X_t)^2\big)\Big(\big(\mu_S(X_t) - m_S^*(X_t)\big)^2 + \big(\mu_A(X_t) - m_A^*(X_t)\big)^2\Big)\Big] \\
&\leq \sigma_u^2 + \mathbb{E}\Big[\big(1+c^2\big)\Big(\big(\mu_S(X_t) - m_S^*(X_t)\big)^2 + \big(\mu_A(X_t) - m_A^*(X_t)\big)^2\Big)\Big] \\
&\leq \sigma_u^2 + a_1\mathbb{E}\Big[\Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2 + \Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2\Big] \,,
\end{aligned}$$

where $a_1 = 1 + c^2$. □

Proof of Proposition 2. Note that

$$\begin{aligned}
\mathbb{E}\Big[\Big(Y_t - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] &= \sigma_u^2 + \mathbb{E}\Big[\Big(\mu(X_t) - \mu_{\text{CAVS}}(X_t) + \mu_{\text{CAVS}}(X_t) - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] \\
&= \sigma_u^2 + \mathbb{E}\Big[\Big(\mu(X_t) - \mu_{\text{CAVS}}(X_t) + c(X_t)\Big)^2\Big]
\end{aligned}$$

Next using the fact that $(a-b)^2 \leq 2(a^2+b^2)$, we write

$$\mathbb{E}\Big[\Big(\mu(X_t) - \mu_{\text{CAVS}}(X_t) + c(X_t)\Big)^2\Big] \leq 2\mathbb{E}\Big[\Big(\mu(X_t) - \mu_{\text{CAVS}}(X_t)\Big)^2 + c(X_t)^2\Big]$$

Applying Proposition 1, we have

$$\begin{aligned}
\mathbb{E}\Big[\Big(Y_t - m_A^*(X_t)m_S^*(X_t)\Big)^2\Big] \leq \sigma_u^2 + 2\mathbb{E}\Big[c(X_t)^2\Big] + 2a_1\mathbb{E}\Big[\Big(\mu_S(X_t) - m_S^*(X_t)\Big)^2\Big] \\
+ 2a_1\mathbb{E}\Big[\Big(\mu_A(X_t) - m_A^*(X_t)\Big)^2\Big],
\end{aligned}$$

setting $a_2 = 2a_1$ yields the proposition. □

Proof of proposition 3. Let $u \sim t_\nu$ denote a t-distributed random variable with mean zero and $\nu > 2$ and define $h(x) : \mathbb{R} \to \mathbb{R}^+$, $v(x) : \mathbb{R} \to [-1,1]$ and $g(x) : \mathbb{R} \to \mathbb{R}$ as

$$\begin{aligned}
h(x,\nu) &= \mathbb{E}_u[|x+u|], \\
v(x,\nu) &= 2F_\nu(x) - 1 \text{ and,} \\
g(x,\nu) &= h(x,\nu)v(x,\nu) \,.
\end{aligned}$$

where F_ν is the cumulative distribution function of u. Note that

$$\begin{aligned}
h(x,\nu) &= \int_{-\infty}^{\infty} (x+u)\text{sign}(x+u) f_\nu(u)du \\
&= \int_{-x}^{\infty} (x+u) f_\nu(u)du - \int_{-\infty}^{-x} (x+u) f_\nu(u)du \\
&= 2\int_{-x}^{\infty} u f_\nu(u)du + xv(x) \\
&= 2\Big(\frac{\nu + x^2}{\nu - 1}\Big) f_\nu(x) + xv(x),
\end{aligned}$$

and the squared approximation error is given by

$$\begin{aligned} D(x,\nu) &= \big(x - g(x,\nu)\big)^2 \\ &= \Big(4xF_\nu(x)(1-F_\nu(x)) - 2\Big(\frac{\nu+x^2}{\nu-1}\Big)f_\nu(x)(2F_\nu(x)-1)\Big)^2 \,. \end{aligned}$$

Denote $D(x) = \lim_{\nu\to\infty} D(x,\nu)$, and let $x^*(\nu) = \arg\max_{x\in\mathbb{R}} D(x,\nu)$. Proceeding numerically, it can be verified that $D(x^*(\nu),\nu)$ is increasing in ν, which implies

$$\begin{aligned} D(x,\nu) = (x - v(x,\nu)h(x,\nu))^2 &\leq \big(4x^*\Phi(x^*)(1-\Phi(x^*)) - 2\phi(x^*)(2\Phi(x^*)-1)\big)^2 \\ &\approx 0.04187 \,, \end{aligned}$$

where $x^* = \arg\max_{x\in\mathbb{R}} \big(4x^*\Phi(x^*)(1-\Phi(x^*)) - 2\phi(x^*)(2\Phi(x^*)-1)\big)^2$. □

Proof of proposition 4. Let $X \sim \mathcal{N}(0,\sigma^2 I)$ and write

$$Y_t = |x_1\kappa\beta_1 + x_2\beta_2|\text{sign}(x_1\beta_1 + x_2\beta_2) + u_t \,, u_t \sim \mathcal{D}(0,1)$$

The OLS estimate for β_1^* is given by

$$\begin{aligned} \beta_1^* &= \frac{1}{\sigma^2}\mathbb{E}\Big[x_1|x_1\kappa\beta_1 + x_2\beta_2|\text{sign}(x_1\beta_1 + x_2\beta_2)\Big] \\ &= \frac{1}{\sigma^2}\mathbb{E}\Big[|x_1||x_1\kappa\beta_1 + x_2\beta_2|\text{sign}(x_1)\text{sign}(x_1\beta_1 + x_2\beta_2)\Big] \\ &= \frac{1}{\sigma^2}\mathbb{E}\Big[|x_1^2\kappa\beta_1 + x_1x_2\beta_2|\text{sign}(x_1^2\beta_1 + x_1x_2\beta_2)\Big] \\ &= \frac{1}{\sigma^2}\mathbb{E}\Big[(x_1^2\kappa\beta_1 + x_1x_2\beta_2)\text{sign}(x_1^2\beta_1^2\kappa + x_1x_2\beta_1\beta_2(1+k) + x_2^2\beta_2^2)\Big] \end{aligned}$$

Proceeding analogously for β_2^*, we have

$$\beta_2^* = \frac{1}{\sigma^2}\mathbb{E}\Big[(x_1x_2\kappa\beta_1 + x_2^2\beta_2)\text{sign}(x_1^2\beta_1^2\kappa + x_1x_2\beta_1\beta_2(1+k) + x_2^2\beta_2^2)\Big]$$

Further, under the assumption that $\beta_1 = \beta = \beta_2$, we have

$$\begin{aligned} \beta_1^* &= \frac{1}{\sigma_1^2}\beta\mathbb{E}\Big[(x_1^2\kappa + x_1x_2)\text{sign}(x_1^2\kappa + x_1x_2(1+\kappa) + x_2^2)\Big], \text{ and} \\ \beta_2^* &= \frac{1}{\sigma_2^2}\beta\mathbb{E}\Big[(x_1x_2\kappa + x_2^2)\text{sign}(x_1^2\kappa + x_1x_2(1+\kappa) + x_2^2)\Big] \end{aligned}$$

Let $S(\kappa) = \text{sign}(x_1^2\kappa + x_1x_2(1+\kappa) + x_2^2)$ and denote by $a = \mathbb{E}\Big[x_1^2S(\kappa)\Big]$, $b = \mathbb{E}\Big[x_2^2S(\kappa)\Big]$ and $c = \mathbb{E}\Big[x_1x_2S(\kappa)\Big]$.

$$\begin{aligned} \mathbb{E}\Big[\Big(Y_t - X_t'\beta^*\Big)^2\Big] - \sigma_u^2 = \mathbb{E}\Big[\Big(&\beta(x_1\kappa + x_2)S(\kappa) - \beta\frac{x_1}{\sigma^2}\mathbb{E}\Big[(x_1^2\kappa + x_1x_2)S(\kappa)\Big]\Big] \\ &- \mathbb{E}\Big[\beta\frac{x_2}{\sigma^2}\mathbb{E}\Big[(x_1x_2\kappa + x_2^2)S(\kappa)\Big]\Big)^2\Big] \end{aligned}$$

and we can write the scaled excess MSE as:

$$
\begin{aligned}
\beta^{-2}\Big(\mathbb{E}\Big[\Big(Y_t - X_t'\beta^*\Big)^2\Big] - \sigma_u^2\Big) &= \mathbb{E}\Big[\Big((x_1\kappa + x_2)S(\kappa) - \frac{x_1}{\sigma^2}\mathbb{E}\Big[(x_1^2\kappa + x_1x_2)S(\kappa)\Big] \\
&\quad - \frac{x_2}{\sigma^2}\mathbb{E}\Big[(x_1x_2\kappa + x_2^2)S(\kappa)\Big]\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(S(\kappa)x_1\kappa + S(\kappa)x_2 - \frac{x_1}{\sigma^2}\kappa a - \frac{c}{\sigma^2}x_1 - \frac{c}{\sigma^2}\kappa x_2 - \frac{x_2}{\sigma^2}b\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(\kappa\big(S(\kappa)x_1 - \frac{x_1}{\sigma^2}a - \frac{x_2}{\sigma^2}c\big) + x_2\big(S(\kappa) - \frac{b}{\sigma^2}\big) - \frac{x_1}{\sigma^2}c\Big)^2\Big].
\end{aligned}
$$

Noting that $S(\kappa) \leq 1$ for all κ and that $\mathbb{E}\Big[x_2^2 S(\kappa)\Big] \leq \mathbb{E}\Big[x_2^2\Big]$, it is clear that the order of the first term will determine the order of $\beta^{-2}\Big(\mathbb{E}\Big[\Big(Y_t - X_t'\beta^*\Big)^2\Big] - \sigma_u^2\Big)$ for large κ. For the first term, we have

$$
\begin{aligned}
\mathbb{E}\Big[\kappa^2\Big(S(\kappa)x_1 - \frac{x_1}{\sigma^2}a - \frac{x_2}{\sigma^2}c\Big)^2\Big] &= \kappa^2\mathbb{E}\Big[\big(S(\kappa)x_1 - \frac{x_1}{\sigma^2}\mathbb{E}[x_1^2S(\kappa)] - \frac{x_2}{\sigma^2}\mathbb{E}[x_1x_2S(\kappa)]\big)^2\Big] \\
&= \kappa^2\mathbb{E}\Big[x_1^2 + \frac{x_1^2}{\sigma^4}\mathbb{E}[x_1^2S(\kappa)]^2 + \frac{x_2^2}{\sigma^4}\mathbb{E}[x_1x_2S(\kappa)]^2\Big] \\
&\quad - \frac{2}{\sigma^2}\mathbb{E}[x_1^2S(\kappa)]^2 - \frac{2}{\sigma^2}\mathbb{E}[x_1x_2S(\kappa)]^2 \\
&= \kappa^2\big(\sigma^2 - \frac{1}{\sigma^2}\mathbb{E}[x_1^2S(\kappa)]^2 - \frac{1}{\sigma^2}\mathbb{E}[x_1x_2S(\kappa)]^2\big)\ .
\end{aligned}
$$

For the second term, we have

$$
\mathbb{E}\Big[x_2^2S(\kappa)^2 - x_2^2\frac{b^2}{\sigma^4}\Big] = \sigma^2 - \frac{b^2}{\sigma^2}\ ,
$$

for the third term, we have

$$
\mathbb{E}\Big[\frac{x_1^2}{\sigma^4}c^2\Big] = \frac{c^2}{\sigma^2}\ .
$$

For the cross-products, we have:

$$
\mathbb{E}\Big[\frac{c}{\sigma^2}x_1x_2S(\kappa) - \frac{b}{\sigma^2}\frac{x_1x_2}{\sigma^2}c\Big] = \frac{c^2}{\sigma^2}\ ,
$$

and

$$
\mathbb{E}\Big[\kappa\big(S(\kappa)x_1 - \frac{x_1}{\sigma^2}a - \frac{x_2}{\sigma^2}c\big)\frac{x_1}{\sigma^2}c\Big] = \kappa(\frac{ac}{\sigma^2} - \frac{ac}{\sigma^2}) = 0\ ,
$$

and finally

$$
\mathbb{E}\Big[\kappa\big(S(\kappa)x_1 - \frac{x_1}{\sigma^2}a - \frac{x_2}{\sigma^2}c\big)\big(x_2S(\kappa) - x_2\frac{b}{\sigma^2}\big)\Big] = -\frac{\kappa}{\sigma^2}c(a+b)\ .
$$

Combining everything, we have

$$\begin{aligned}
\beta^{-2}\Big(\mathbb{E}\Big[\Big(Y_t - X_t'\beta^*\Big)^2\Big] - \sigma_u^2\Big) &= \kappa^2\big(\sigma^2 - \frac{1}{\sigma^2}(a^2+c^2)\big) - 2\kappa\frac{c(a+b)}{\sigma^2} + \sigma^2 - \frac{b^2}{\sigma^2} - \frac{c^2}{\sigma^2} \\
&= \kappa^2\Big(\sigma^2 - \frac{\mathbb{E}[x_1^2 S(\kappa)]^2}{\sigma^2} - \frac{\mathbb{E}[x_1x_2 S(\kappa)]^2}{\sigma^2}\Big) \\
&- 2\kappa\mathbb{E}[x_1x_2S(\kappa)]\Big(\frac{\mathbb{E}[x_1^2S(\kappa)]}{\sigma^2} + \frac{\mathbb{E}[x_2^2S(\kappa)]}{\sigma^2}\Big) \\
&+ \Big(\sigma^2 - \frac{\mathbb{E}[x_2^2S(\kappa)]^2}{\sigma^2} - \frac{\mathbb{E}[x_1x_2S(\kappa)]^2}{\sigma^2}\Big)\,.
\end{aligned}$$

To compute the order of the MSE, note that for $\kappa > 1$ we have

$$\frac{1}{\sigma^2}\mathbb{E}[x_1^2S(\kappa)]^2 - \frac{1}{\sigma^2}\mathbb{E}[x_1x_2S(\kappa)]^2 < \frac{1}{\sigma^2}\mathbb{E}[x_1^2S(1)]^2 - \frac{1}{\sigma^2}\mathbb{E}[x_1x_2S(1)]^2 = \sigma^2$$

which implies that, for $\kappa > 1$,

$$0 < \big(\sigma^2 - \frac{1}{\sigma^2}\mathbb{E}[x_1^2S(\kappa)]^2 - \frac{1}{\sigma^2}\mathbb{E}[x_1x_2S(\kappa)]^2\big) < \sigma^2\,,$$

and $\mathbb{E}\Big[\kappa^2\Big(S(\kappa)x_1 - \frac{x_1}{\sigma^2}a - \frac{x_2}{\sigma^2}c\Big)^2\Big] \propto \kappa^2$ for large κ, which proves Corollary 2. □

Capítol 2

Backtesting Global Growth-at-Risk

2.1 Introduction

In recent years, the focus of policymakers on downside risk has substantially increased (Prasad et al., 2019; Caldera Sanchéz and Röhn, 2016), which has motivated the development of tools to assess the likelihood and extent of extreme events of key economic variables (Adrian et al., 2019; Ghysels et al., 2018). In particular, the International Monetary Fund (IMF) has recently popularized a risk measure for GDP growth called Growth-at-Risk (GaR), which is the worst-case scenario GDP growth at a given coverage level and is the analog of the classic Value-at-Risk (VaR) used in risk management. Several institutions currently routinely publish GaR for major world economies.[1] Despite GaR's rapid success, its out-of-sample predictive performance has not been extensively studied.[2]

The main objective of this paper is to conduct an out-of-sample empirical evaluation of GaR predictions at different horizons for a panel of OECD countries. We consider GaR predictions constructed from quantile regression and GARCH models. Our evaluation strategy is based on the classic backtesting methodology developed in the risk management literature. Our analysis uncovers a number of empirical findings and identifies best practices.

We explore different types of multi-country GaR forecasts that supervisory institutions may consider. Recall that the univariate $(1-p)\%$ GaR is defined as the lower one-sided prediction interval that contains future realizations of GDP growth of a given country with $(1-p)\%$ coverage probability. We consider two generalizations of univariate GaR called marginal GaR and joint GaR. The $(1-p)\%$

[1] See, for example, IMF (2017) and ECB (2019).

[2] Adrian et al. (2019, 2018) evaluate the accuracy of predictive densities for USA GDP growth. However, the accuracy of the prediction intervals for GDP growth remains relatively unexplored.

marginal GaR is defined as the prediction region that contains the GDP growth of *each* country with $(1-p)\%$ coverage probability. Stated simply, marginal GaR is the region obtained by stringing together the univariate GaR prediction intervals of each country. In particular, the forecasts published by the IMF may be interpreted as marginal GaR. The $(1-p)\%$ joint GaR is defined as the prediction region that contains the GDP growth of *all* countries with $(1-p)\%$ coverage probability. A number of methods are available to construct joint GaR and we rely on Bonferroni's method and the bootstrap joint prediction regions (BJPR) method of Wolf and Wunderli (2015). To the best of our knowledge, this is the first paper that addresses joint GaR prediction. We emphasize that marginal and joint GaR predictions serve different purposes and a detailed comparison of these predictions is provided in what follows. In this work we consider both types of predictions simply to provide a comprehensive assessment of GaR predictive ability.

Notably, the primary focus of this paper is to construct and evaluate prediction regions rather than predictive densities. Although both problems are of interest, we focus on prediction regions for two reasons. First, international organizations and central banks typically publish prediction regions, which should thus be evaluated properly. Second, evaluation metrics that assess the goodness-of-fit of the entire (or part of the) density do not necessarily identify the methods that are better suited to capture downside risk as measured by specific quantiles. That being said, for completeness, we also carry out a predictive density evaluation exercise.

We construct marginal and joint GaR on the basis of quantile regression and GARCH models. One of the appealing features of quantile regression, which has been used extensively in the GaR literature, is that it allows direct linkage of downside risk predictors to the quantiles of GDP growth. GARCH models are routinely used to construct VaR forecasts in the risk management literature yet, somehow surprisingly, are rarely used in the GaR literature. To the best of our knowledge, this is the first paper that compares the two approaches out-of-sample.[3] We consider a number of quantile regression specifications based on country-specific and global variables, including the national financial conditions index (NFCI) (Adrian et al., 2019), credit-to-GDP statistics (Borio and Lowe, 2002), housing prices (Claessens et al., 2008), economic uncertainty indexes (Baker et al., 2016; Ahir et al., 2018), a geopolitical risk index (Caldara and Iacoviello, 2018) and market risk measures (Faust et al., 2013; Estrella and Hardouvelis, 1991). Among these variables, the NFCI is considered to be one of the most relevant predictors of downside risk by the GaR literature. We also remark that most of the predictors used in this study (NFCI included) are not constructed in real-time and have look-

[3] Adrian et al. (2019) report that in their analysis they also employ a GARCH model; however, the paper only reports estimation and out-of-sample results for a conditionally heteroskedastic model in which the variance dynamics are driven by the NFCI (as opposed to past squared prediction errors as in a GARCH).

ahead bias. Furthermore, we consider a set of GARCH models: GARCH(1,1), GJR-GARCH(1,1), Factor GARCH(1,1) and a GARCH(1,1) augmented with the NFCI as an exogenous conditional variance predictor. The relative scarcity of GARCH applications in macroeconomics may be due to the short time series available for estimation. We overcome this hurdle by using composite estimation (Pakel et al., 2011), which exploits cross-sectional commonality in GARCH dynamics to obtain precise parameter estimates.

We backtest the marginal and joint GaR forecasts by using a battery of tools developed in the risk management literature. We employ several variants of the dynamic quantile test of Engle and Manganelli (2004) to assess whether the GaR predictions are efficient with respect to different information sets. Additionally, the marginal GaR forecasts are evaluated using the tick loss, which is a loss function commonly used to assess the accuracy of VaR predictions (Giacomini and Komunjer, 2005). We compare quantile regression and GARCH with benchmarks based on the historical unconditional distribution of GDP growth rates.

We study GaR predictive ability for a panel of 24 OECD countries from 1973Q1 to 2016Q4. We first conduct an in-sample analysis based on the entire sample and then recursively forecast and evaluate the marginal and joint GaR from 1983Q4 to 2016Q4.

Our backtesting results shows that quantile regression and GARCH forecasts have a similar performance. If anything, our evidence suggests that standard volatility models such as the GARCH(1,1) are more accurate, even though the GARCH(1,1) uses no information other than GDP growth. This is remarkable considering that most of the quantile regression predictors (including the NFCI) have a look-ahead bias that may not be negligible.[4] For marginal GaR, GARCH typically has better backtesting performance than quantile regression across all evaluation metrics considered. In particular, GARCH performs better than the quantile regression based on the NFCI, which is the best performing quantile regression specification. A superior predictive ability test comparison based on the tick loss shows that the GARCH(1,1) outperforms the quantile regression based on the NFCI for at most six countries across all horizons. In contrast, the quantile regression based on the NFCI outperforms the GARCH(1,1) for at most two countries across all horizons. For joint GaR, GARCH forecasts based on BJPR have substantially better backtesting performance than those based on quantile regression in conjunction with Bonferroni.[5] Additionally, in the robustness section we show that GARCH density forecasts are more accurate than quantile regression density forecasts constructed using the methodology of Adrian et al. (2019).

[4]We remark that the NFCIs are estimated from a dynamic a factor model with time-varying parameters and such models tend to have fairly high filtering uncertainty at the sample endpoints.

[5]The comparison with quantile regression should be taken with caution as the BJPR cannot be applied to quantile regression.

Overall, despite the popularity of quantile regression for downside risk prediction, this paper suggests caution against relying too heavily on this technique.

There are a number of possible explanations for our findings. First, quantile regression rely on specifying an appropriate set of downside risks predictors that measure specific sources of distress. Their relevance and predictive ability may vary over time. In fact, additional robustness checks in the Appendix show that the out-of-sample performance of quantile regression based on the NFCI deteriorates if the period starting from the Great Financial Crisis is excluded from the validation sample. Moreover, it is unclear whether financial conditions are a relevant downside risk predictor during the covid-19 pandemic of 2020. In contrast, a pure time series model such as a GARCH that is agnostic about the specific sources of distress in the economy may be more robust for prediction. Second, capturing the dynamics of conditional quantiles is empirically challenging in a macro environment where time series information is scarce, especially if the interest lies in extreme quantiles. Thus, a volatility model may be better suited for quantile forecasting even if it is misspecified.

This paper is related to various strands of literature. First, it is related to the rapidly growing literature on GaR. This includes work by Adrian et al. (2019), Plagborg-Møller et al. (2020), Carriero et al. (2020), De Nicolò (2019), Beutel (2019), Chavleishvili and Manganelli (2019), and Adrian et al. (2018). Our work also relates to the literature on interval forecast evaluation and backtesting in risk management. This includes work by Giacomini and Komunjer (2005) and Christoffersen (1998). This paper is also connected to the literature on dynamic quantile models; see, for example, Engle and Manganelli (2004) and White et al. (2015). Finally, our work is related to the literature on the impact of financial distress on real activity, including works by Allen et al. (2012) and Brownlees and Engle (2017).

The remainder of this paper is structured as follows. Section 2.2 details the methodology, and Section 2.3 presents the empirical evidence. Concluding remarks follow in Section 2.4. We provide additional methodological details and empirical results in the Appendix.

2.2 Forecasting Growth-at-Risk

2.2.1 Marginal and Joint Growth-at-Risk Definitions

Let Y_{it} denote the GDP growth rate of country $i = 1, \ldots, n$ for period $t = 1, \ldots, T$. The h-step-ahead $(1 - p)\%$ marginal GaR is defined as the prediction region $GaR^M_{t+h|t} = (GaR^M_{1\,t+h|t}, \infty) \times \cdots \times (GaR^M_{n\,t+h|t}, \infty)$ such that for each

i we have

$$\mathbb{P}_t(Y_{i\,t+h} \leq GaR^M_{i\,t+h|t}) = p,$$

where $\mathbb{P}_t(\cdot)$ is the probability measure conditional on the information set available in period t. That is, the $(1-p)\%$ marginal GaR is the prediction region that should contain the GDP growth of *each* country with $(1-p)\%$ probability. The h-step-ahead $(1-p)\%$ joint GaR is the prediction region $GaR^J_{t+h|t} = (GaR^J_{1\,t+h|t}, \infty) \times \cdots \times (GaR^J_{n\,t+h|t}, \infty)$ such that

$$\mathbb{P}_t \left(\text{at least one growth rate } Y_{i\,t+h} \text{ is not in } GaR^J_{t+h|t} \right) = p. \tag{2.1}$$

That is, the $(1-p)\%$ joint GaR should contain the GDP growth of *all* countries with $(1-p)\%$ probability. We suppress the dependence of GaR on the coverage level $(1-p)\%$ to avoid burdening notation. The marginal GaR prediction region is unique[6] and determined by the conditional quantiles of the GDP growth rate of each country. In contrast, the joint GaR prediction region is not uniquely determined by its definition, and a number of procedures are available to construct such regions. Marginal GaR is a natural generalization of univariate GaR for a panel of countries. However, marginal GaR may be considered to be a myopic global risk measurement tool as it measures the downside risks for each country individually. In contrast, joint GaR is designed to measure downside risk in the case of a system-wide event[7] and produces a prediction region that contains the growth rates of all countries simultaneously with the desired coverage probability.

Building upon Wolf and Wunderli (2015) and Romano and Wolf (2007), we introduce a more general joint GaR prediction region (see also Romano et al., 2010, p. 95). Defining the joint GaR prediction region on the basis of the event "*at least one growth rate $Y_{i\,t+h}$ is not in $GaR^J_{t+h|t}$*" as in (2.1) may be excessively restrictive, even for moderately large panels. This definition may, in turn, lead to prediction regions that are excessively large and of little practical use. Therefore, we introduce the h-step-ahead $(1-q)\%/(1-p)\%$ joint GaR as the prediction region $GaR^{Jq}_{t+h|t} = (GaR^{Jq}_{1\,t+h|t}, \infty) \times \cdots \times (GaR^{Jq}_{n\,t+h|t}, \infty)$ such that

$$\mathbb{P}_t \left(\text{at least } \lceil qn \rceil \text{ growth rates } Y_{i\,t+h} \text{ are not in } GaR^{Jq}_{t+h|t} \right) = p,$$

where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. That is, the $(1-q)\%/(1-p)\%$ joint GaR should contain the GDP growth of $(1-q)\%$ of the

[6]Throughout this paper, we assume that the conditional cumulative distribution function of the GDP growth rate of each country is invertible, which guarantees the uniqueness of the conditional quantiles and the corresponding marginal GaR prediction region.

[7]That is, the event "*at least one growth rate $Y_{i\,t+h}$ is not in $GaR^J_{t+h|t}$*".

countries with $(1-p)\%$ probability. Defining the prediction region on the basis of a less stringent event leads to smaller and potentially more informative regions. In practice, q may be chosen such that a large fraction of the entire system, for example, 95%, is within the prediction region with the prescribed coverage probability. Let us also emphasize that the event "*at least* $\lceil qn \rceil$ *growth rates* Y_{it+h} *are not in* $GaR^{Jq}_{t+h|t}$" may be more interesting from a global risk monitoring perspective, irrespective of the panel dimensions. For example, in 2019Q3, global recession fears were triggered by weak growth figures for 5 economies.

A number of remarks are in order. It is important to emphasize that unlike Adrian et al. (2019), we focus on predicting the h-step ahead GDP growth rate rather than the (average) cumulative h-step ahead GDP growth. Both are commonly encountered target variables of interest in the macro-econometrics literature (Faust and Wright, 2013; Stock and Watson, 2006). Here, we focus on the h-step ahead growth rate as we are interested in assessing for how many quarters ahead covariates and/or exploiting GDP dynamics deliver more accurate forecasts than predictions based solely on the historical unconditional distribution of GDP. We remark that the literature documents that the NFCI is significant (in-sample) for the cumulative h-step ahead GDP growth up to 1 year ahead (Adrian et al., 2019) or even (roughly) 2 years ahead (Adrian et al., 2018). This evidence prompts the question of how persistent the information content of the NFCI is. That being said, for completeness in the empirical application we also consider cumulative h-step ahead GaR prediction.

The forecasts published by the IMF can be interpreted as marginal GaR. By construction, when the number of countries is large, the probability of observing a GDP realization of *at least one* country outside the marginal GaR region is clearly much larger than one minus the nominal marginal coverage. In fact, in our empirical analysis, we document at least one GDP realization outside the 95% marginal GaR prediction region every 3 quarters for all forecasting methods considered. This feature may not be appealing from a global risk monitoring perspective. In contrast, the probability of observing a GDP realization of *any* country outside the joint GaR region is equal to one minus the nominal joint coverage. We think of joint GaR as a stress GaR measure that is well suited to monitor global downside risk.

It is also important to emphasize the difference between constructing the univariate GaR for the average of the countries in a panel versus constructing the joint GaR of all countries in the panel. We argue that a supervisory institution may be interested in tracking the GaR of all countries jointly rather than an average. This may be particularly important if a supervisory institution is concerned about the fact that the distress of a small yet systemic group of countries may jeopardize the entire system. In fact, in the 2011 European sovereign debt crisis, the turmoil in the Eurozone originated from smaller periphery economies.

The coverage properties of the considered GaR prediction regions are relative to a single, fixed horizon. Instead, one may be interested in constructing prediction regions that provide uniform coverage simultaneously across all horizons. Although we do not consider alternative GaR definitions, we note that the methodology of Wolf and Wunderli (2015) enables the construction of such prediction regions. Moreover, the joint GaR prediction region defined in this work is rectangular. Other possibilities may be explored (for example, one may define an elliptical joint GaR region). However, we believe that a rectangular joint GaR region is natural, easy to interpret and overall, well suited for global risk monitoring.

In conclusion, we remark that we do not take a strong stand with respect to which regions should be constructed and reported. Our work is mainly concerned with providing a comprehensive assessment of the predictive ability of different types of GaR predictions that may be entertained by international organizations such as the IMF, BIS or ECB.

2.2.2 Models for Growth-at-Risk

Quantile Regression

Quantile regression was popularized by Adrian and Brunnermeier (2016) in the aftermath of the great financial crisis as a tool to construct downside risk measures and is routinely used to construct GaR predictions. In the quantile regression framework, the conditional quantiles of GDP growth are modeled as linear functions of a set of quantile predictors. More precisely, the h-step-ahead $p\%$ conditional quantile of $\{Y_{i\,t}\}$ is given by

$$Q_p(Y_{i\,t+h}|\mathcal{I}_t) = \alpha_i^p + \beta_{i\,0}^p Y_{i\,t} + \beta_{i\,1}^p X_{i1\,t} + \ldots + \beta_{i\,K}^p X_{iK\,t}, \tag{2.2}$$

where $Q_p(Y_{i\,t+h}|\mathcal{I}_t)$ denotes the $p\%$ quantile of $Y_{i\,t+h}$ conditional on the information set available in period t, which is denoted by $\mathcal{I}_t$, and $X_{i\,k\,t}$ for $k = 1, \ldots, K$ denotes the set of predictors for country i.

To make the model in (A.6) operational, one must specify an appropriate set of predictors. In this work, we consider a moderately large set of candidates based on the evidence established in the literature. The precise set of variables that we explore is enumerated in Section 2.3.1, where we introduce the data used in our analysis.

The parameters in (A.6) are estimated by minimizing the tick loss, which is given by

$$\mathrm{TL}_p = \frac{1}{T}\sum_{t=1}^{T} \rho_p(\, Y_{i\,t} - Q_p(Y_{i\,t}|\mathcal{I}_{t-h})\,),$$

where $\rho_p(x) = x\left(p - \mathbb{1}_{\{x<0\}}\right)$. We refer to Koenker and Basset (1978) for the details of the estimation of quantile regression. In the forecasting application, we also rely on a LASSO-type estimation of quantile regression to regularize the estimates of more expensively parameterized specifications.

GARCH

GARCH models are the workhorse of volatility forecasting (Brownlees et al., 2011) and are routinely used to construct VaR predictions. In this work, we construct GaR forecasts on the basis of the following GARCH-type specification:

$$Y_{i\,t+1} = \mu_{i\,t+1|t} + \sqrt{\sigma^2_{i\,t+1|t}}\,Z_{i\,t+1} \quad Z_{i\,t+1} \overset{i.i.d}{\sim} \mathcal{D}_{Z_i}(0,1), \tag{2.3}$$

where $\mu_{i\,t+1|t}$ denotes the 1-step-ahead conditional mean, $\sigma^2_{i\,t+1|t}$ is the 1-step-ahead conditional variance and $\mathcal{D}_{Z_i}(\mu, \sigma^2)$ is a location-scale distribution with mean μ and variance σ^2. The 1-step-ahead conditional distribution of the GDP growth rates implied by (A.1) is

$$Y_{i\,t+1}|\mathcal{I}_t \sim \mathcal{D}_{Z_i}(\mu_{i\,t+1|t}, \sigma^2_{i\,t+1|t}),$$

which indicates that the innovation distribution $\mathcal{D}_{Z_i}$ determines the shape of the conditional distribution. The 1-step-ahead $p\%$ conditional quantile of $\{Y_{i\,t}\}$ is then given by

$$Q_p(Y_{i\,t+1}|\mathcal{I}_t) = \mu_{i\,t+1|t} + \sqrt{\sigma^2_{i\,t+1|t}}F^{-1}_{Z_i}(p), \tag{2.4}$$

where $F^{-1}_{Z_i}(\cdot)$ is the inverse cumulative distribution function of $\mathcal{D}_{Z_i}$. In general, the h-step-ahead conditional distribution for $h > 1$ is not available in closed form, and simulation techniques similar to the ones used in Brownlees and Engle (2017) must be applied to estimate the $p\%$ conditional quantile. In this case, the h-step-ahead $p\%$ conditional quantile is

$$Q_p(Y_{i\,t+h}|\mathcal{I}_t) = F^{-1}_{\widetilde{Y}_{i\,t+h|t}}(p), \tag{2.5}$$

where $\widetilde{Y}_{i\,t+h|t}$ denotes a simulated realization of the process in period $t+h$, given the path of the GDP growth rates observed up to period t. This value is obtained by iterating the dynamic model defined in (A.1); therefore, we label the quantile forecast in (2.5) as iterated. Section A.1 of the Appendix details the simulation algorithm for this computation.

To make the model in (A.1) operational, the conditional mean $\mu_{i\,t+1|t}$, conditional variance $\sigma^2_{i\,t+1|t}$ and innovation distribution $\mathcal{D}_{Z_i}$ must be specified. For the conditional mean, we rely on an AR(1) model. The conditional variance equation is key to quantile forecasting, and we entertain a number of GARCH specifications. First, we consider the standard GARCH(1,1),

$$\sigma^2_{i\,t+1|t} = \sigma^2_i(1 - \alpha_i - \beta_i) + \alpha_i \varepsilon^2_{i\,t} + \beta_i \sigma^2_{i\,t}, \quad (2.6)$$

where $\varepsilon_{i\,t}$ is the AR(1) residual, and the variance equation parameters satisfy the constraints $\sigma_i > 0$, $\alpha_i > 0$, $\beta_i \geq 0$ and $\alpha_i + \beta_i < 1$. Additionally, we consider the GJR-GARCH(1,1), which takes into account asymmetry in conditional volatility dynamics, the Factor GARCH(1,1), which decomposes volatility dynamics into systematic and idiosyncratic components, and a GARCH(1,1) whose conditional variance equation is augmented with the NFCI as an exogenous predictor. We describe the models in detail in the Appendix. Finally, the innovation distribution $\mathcal{D}_{Z_i}$ is estimated nonparametrically.

GARCH models are estimated by quasi-maximum likelihood. A challenge in the estimation of GARCH models when applied to macro time series is that moderately large samples are needed to obtain stable parameter estimates (Brownlees et al., 2011). This problem becomes more relevant when carrying out a recursive estimation for prediction. Thus, we choose the following estimation strategy. In the in-sample analysis, we estimate GARCH models individually for each country based on (standard) quasi-maximum likelihood. In the out-of-sample analysis, we rely on a panel GARCH estimation procedure known in the literature as composite likelihood (Pakel et al., 2011). Composite likelihood estimation enhances efficiency by pooling information across series. This is particularly advantageous in the beginning of the out-of-sample forecasting exercise where the length of the in-sample estimation window is small. We describe this procedure in the Appendix.

2.2.3 Constructing Marginal and Joint Growth-at-Risk

The $(1-p)\%$ marginal GaR prediction region is uniquely determined by the conditional quantiles of the GDP growth rates and is obtained by stringing together individual quantiles. That is, the lower endpoints of the prediction region $GaR^M_{t+h|t}$ are, for $i = 1, \ldots, n$,

$$GaR^M_{i\,t+h|t} = Q_p(Y_{i\,t+h}|\mathcal{I}_t),$$

where $Q_p(Y_{i\,t+h}|\mathcal{I}_t)$ is set equal to (A.6) for quantile regression or (2.4) and (2.5) for GARCH.

The $(1-p)\%$ joint GaR prediction region can be constructed using different methods. We note that Wolf and Wunderli (2015) contains an extensive discussion of the construction of joint prediction regions, and we refer the interested reader to that paper for additional background. The first method that we consider, for illustrative purposes, is labeled "joint marginal" and is the prediction region obtained by setting the joint GaR equal to the marginal GaR. The joint marginal GaR can be considered the naïve joint GaR that would be constructed if one ignored that the marginal and joint coverage properties differ. The second method that we consider is called Bonferroni's method. It follows from Bonferroni's inequality that joining univariate $p/n\%$ quantile forecasts will yield a region with uniform coverage of at least $(1-p)\%$. The Bonferroni-based $(1-p)\%$ joint GaR, which we denote by $GaR^{J,B}_{t+h|t}$, is the prediction region with lower endpoints for $i = 1, \ldots, n$ given by

$$GaR^{J,B}_{i\,t+h|t} = Q_{p/n}(Y_{i\,t+h}|\mathcal{I}_t).$$

Note that Bonferroni's inequality does not account for cross-sectional dependence information; therefore, the inequality usually provides overly conservative regions with joint coverage much greater than $(1-p)\%$ (thus, much larger regions). The third and final method that we consider is the BJPR of Wolf and Wunderli (2015), which is a bootstrap-based procedure that allows for the construction of joint prediction regions under fairly general assumptions. Despite its wide applicability, the BJPR requires a model for the GDP growth rates and in this work, can be applied only to GARCH models. The BJPR-based $(1-p)\%$ joint GaR, which is denoted by $GaR^{J,BJPR}_{t+h|t}$, is the prediction region with lower endpoints for $i = 1, \ldots, n$ given by

$$GaR^{J,BJPR}_{i\,t+h|t} = \mu_{i\,t+h|t} + d^1_p \sigma_{i\,t+h|t}, \tag{2.7}$$

where $\mu_{i\,t+h|t}$ is the h-step-ahead conditional mean, $\sigma_{i\,t+h|t}$ is the h-step-ahead conditional volatility and d^1_p is the $p\%$ quantile of U^1_t, with U^1_t being the smallest value of the vector $(\widetilde{Z}_{1\,t+h|t}, \ldots, \widetilde{Z}_{n\,t+h|t})'$, with $\widetilde{Z}_{i\,t+h|t} = (\widetilde{Y}_{i\,t+h|t} - \mu_{i\,t+h|t})/\sigma_{i\,t+h|t}$ and $\widetilde{Y}_{i\,t+h|t}$ defined as in Section 2.2.2. Typically, d^1_p is unknown and can be approximated by resampling. Section A.1 of the Appendix provides a bootstrap algorithm to estimate this quantile. Wolf and Wunderli (2015) show that BJPR-based regions have an asymptotic coverage of $(1-p)\%$.

The $(1-q)\%/(1-p)\%$ joint GaR can also be constructed on the basis of the BJPR method. The BJPR-based $(1-q)\%/(1-p)\%$ joint GaR, which is denoted by $GaR^{Jq,BJPR}_{t+h|t}$, is the prediction region with lower endpoints for $i = 1, \ldots, n$ given by

$$GaR^{Jq,BJPR}_{i\,t+h|t} = \mu_{i\,t+h|t} + d^{\lceil qn \rceil}_p \sigma_{i\,t+h|t},$$

where $\mu_{i\,t+h|t}$ is the h-step-ahead conditional mean, $\sigma_{i\,t+h|t}$ is the h-step-ahead conditional volatility and $d^{\lceil qn \rceil}_p$ is the $p\%$ quantile of $U^{\lceil qn \rceil}_t$, with $U^{\lceil qn \rceil}_t$ being the $\lceil qn \rceil$-th smallest value of the vector $(\widetilde{Z}_{1\,t+h|t}, \ldots, \widetilde{Z}_{n\,t+h|t})'$ and with $\widetilde{Z}_{i\,t+h|t}$ defined as above.

2.2.4 Backtesting

We measure the accuracy of GaR predictions using standard backtesting tools from the VaR evaluation literature. We define the average empirical coverage of the marginal GaR and the empirical coverage of the joint GaR, respectively, as

$$\hat{C}^M = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{\{Y_{i\,t} > GaR^M_{i\,t|t-h}\}}\right)$$

and

$$\hat{C}^J = \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{\bigcup_{i=1}^{n}\{Y_{i\,t} > GaR^J_{i\,t|t-h}\}}.$$

Accurate GaR forecasts are expected to have an empirical coverage close to the nominal coverage. The average lengths of the marginal and joint GaR predictions are defined as

$$\hat{L}^M = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{Q}_{0.99}(Y_i) - GaR^M_{i\,t|t-h}\right)$$

and

$$\hat{L}^J = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{Q}_{0.99}(Y_i) - GaR^J_{i\,t|t-h}\right),$$

where $\hat{Q}_{0.99}(Y_i)$ denotes the (unconditional) 99% empirical quantile of the i-th series estimated on the entire sample. All else being equal, GaR forecasts with a smaller length are typically preferred. We measure length with respect to the 99% quantile as most methods considered in this work generate values of $GaR^M_{i\,t|t-h}$ and $GaR^J_{i\,t|t-h}$ that are smaller than this quantity.[8] We also note that the “size” of a prediction region is usually measured by its volume. We report the average length instead, which is more natural in this context.

We backtest GaR forecasts on the basis of the dynamic quantile test of Engle and Manganelli (2004). To explain this test, we must first introduce the notion of a hit sequence. For marginal GaR, we define the hit sequence of the i-th series as $H^M_{i\,t} = \mathbb{1}_{\{Y_{i\,t} \leq GaR^M_{i\,t|t-h}\}} - p$, that is, a sequence of binary random variables that are

[8] In the empirical application, all the quantile regression models that we consider have at most 2 observations per country such that $GaR^M_{i\,t|t-h} > \hat{Q}_{0.99}(Y_i)$ or $GaR^J_{i\,t|t-h} > \hat{Q}_{0.99}(Y_i)$. In these cases, we set the length to 0. This is not the case for the GARCH models. This rule slightly biases the results in favor of the quantile regression models.

equal to $1-p$ when the t-th realization of the i-th series is below its corresponding marginal GaR and $-p$ otherwise. Analogously, for joint GaR, we define the joint hit sequence as $H_t^J = \mathbb{1}_{\bigcup_{i=1}^n \{Y_{i\,t} \leq GaR^J_{i\,t|t-h}\}} - p$. Let $W_{1\,t}, \ldots, W_{K\,t}$ denote a set of auxiliary predictors observed in period t, and consider the regression

$$H_{t+h} = c_0 + \sum_{k=1}^{K} c_k W_{k\,t} + u_{t+h}, \tag{2.8}$$

where H_t may denote either $H_{i\,t}^M$ or H_t^J, and u_t is an error term. Optimal GaR forecasts generate zero-mean m-dependent hit sequences with dependence parameter $m = h$[9] (Christoffersen, 1998). This result implies that if the GaR forecasts used to construct the hit sequence are optimal, the coefficients of the regression in (2.8) are zero. Thus, the dynamic quantile test is based on testing the null H_0 : $c_0 = \ldots = c_K = 0$ against the alternative $\mathrm{H}_1 : c_k \neq 0$ for some $k = 0, \ldots, K$. Following Engle and Manganelli (2004), the dynamic quantile test statistic is constructed using an appropriately tailored Wald test that takes into account the m-dependence structure of the hit sequence. We backtest GaR forecasts on the basis of variants of the dynamic quantile test that rely on different choices of auxiliary predictors to assess optimality with respect to different information sets. First, we consider the dynamic quantile test based on no auxiliary predictors, which allows us to test whether GaR forecasts are unconditionally optimal in the sense of having correct unconditional coverage. Next, we consider the dynamic quantile test based on setting the auxiliary predictors to the lags of the hit sequence. This is the most common form of the test used in the literature, and it allows us to assess whether the hit sequence is optimal with respect to the information set generated by the hit sequence itself. In addition to the standard dynamic quantile tests described above, in the empirical exercise, we employ dynamic quantile tests based on setting the auxiliary predictors to a set of downside risk predictors. In our analysis, we use the dynamic quantile test to assess the out-of-sample optimality of GaR forecasts as well as to evaluate the in-sample goodness-of-fit of our models.

Finally, marginal GaR forecasts are evaluated on the basis of a loss function. Noting that marginal GaR forecasts are determined by quantile forecasts, we evaluate the performance of competing predictions on the basis of the average tick loss, defined as

$$\mathrm{TL}_p^M = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{T} \sum_{t=1}^{T} \rho_p (Y_{i\,t} - GaR^M_{i\,t|t-h}) \right).$$

[9]A sequence of random variables $X_1, X_2, \ldots$ is m-dependent for some $m > 0$ if for any i, we have that $X_1, \ldots, X_i$ is independent of $X_{i+j}, X_{i+j+1}, \ldots$ when $j \geq m$.

The tick loss is a proper loss function to evaluate quantile forecasts (Giacomini and Komunjer, 2005) and is the loss minimized in the estimation of the quantile regression.

2.3 Empirical Analysis

2.3.1 Data

We construct GaR forecasts from a balanced panel of GDP growth rates for 24 OECD countries that spans from 1961Q1 to 2019Q1.[10] The sample comprises all countries for which GDP data are available since at least 1973Q1 to match some of the predictors used in the quantile regression analysis. GDP growth rates are defined as the quarterly percentage change in seasonally adjusted real GDP and are obtained from the OECD database.

Quantile regression requires specifying a set of downside risk predictors. The list of variables that we entertain builds on the evidence established in the literature and contains both country-specific and global predictors. We consider country-specific variables, namely, the national financial conditions index (NFCI), credit-to-GDP gap and growth (CG and CR), term spread (TS), housing prices (HP), the World Uncertainty Index (WUI), and economic policy uncertainty (EPU). Additionally, we consider global predictors such as the global real activity factor (GF), stock variance (SV), credit spread (CS), and the geopolitical risk index (GPR). The details on the data availability, construction and imputation can be found in the Appendix.

2.3.2 In-sample Analysis

In-sample Quantile Regression Analysis

We begin by reporting the estimation results of a set of baseline quantile regressions used to gauge the explanatory power of each predictor. For each country $i = 1, \ldots, n$, each forecast horizon $h = 1, \ldots, 4$ and each predictor $k = 1, \ldots, K$, we estimate the 5% quantile regression given by

$$Q_{0.05}(Y_{i\,t+h}|\mathcal{I}_t) = \alpha_i^{0.05} + \beta_{i\,0}^{0.05} Y_{i\,t} + \beta_{i\,1}^{0.05} X_{ik\,t}, \tag{2.9}$$

[10] Australia (AUS), Austria (AUT), Belgium (BEL), Canada (CAN), Denmark (DNK), Finland (FIN), France (FRA), Germany (DEU), Greece (GRC), Iceland (ISL), Ireland (IRL), Italy (ITA), Japan (JPN), South Korea (KOR), Luxembourg (LUX), Mexico (MEX), the Netherlands (NLD), Norway (NOR), Portugal (PRT), Spain (ESP), Sweden (SWE), Switzerland (CHE), the U.K. (GBR) and the U.S.A. (USA).

which is similar in spirit to the linear regression specifications used to test for Granger causality. The quantile regression in (2.9) based on the NFCI is used as a benchmark and is called QR-NFCI. To simplify comparisons, for each predictor X_{ikt}, the corresponding quantile regression given by (2.9) is estimated using the subset of observations in which both the predictor X_{ikt} and the NFCI are available, which in most cases, corresponds to the sample where the NFCI is available. The predictors are standardized to have a mean of zero and a variance of one throughout this section to ease the interpretation of the quantile regression coefficients.

TABLE 2.1 ABOUT HERE

Table 2.1 reports the summary estimation results. The table reports the following for each forecasting horizon and predictor: the sample period used for estimation; the quartiles of the estimated β_{i1} coefficient across countries; the percentage of countries for which the β_{i1} coefficient is significant at the 5% significance level; and the percentage increase in the in-sample average tick loss over the baseline QR-NFCI model. The models are sorted by the average percentage of significance of each predictor across horizons. We remark that quantile regression based on the CS and the uncertainty indexes (EPU, WUI and GPR) are estimated on shorter samples. The NFCI is a strong predictor of downside risk, especially at shorter horizons, which is consistent with the findings in Adrian et al. (2019). In particular, the NFCI has a negative impact on the 5% quantile for the vast majority of countries considered, and its coefficient is significant for up to 75% of the series. The TS improves the average tick loss over longer horizons and is relevant for up to 50% of the series. The GF is relevant for up to 30% of the series, despite not improving the average in-sample fit. The remaining variables only marginally improve the fit. Overall, our results confirm the prominence of the NFCI and show that a few additional regressors may be of value in predicting downside risks to GDP growth. In particular, the TS, GF, CS, HP and SV may be useful additional predictors, especially at longer horizons.

Next, we consider a number of multivariate quantile regression specifications to assess whether alternative combinations of the predictors emphasized above improve the fit relative to the baseline QR-NFCI. Despite its potential, we do not consider the CS due to its shorter sample availability. We estimate the multivariate quantile regression for the 5% quantile for each country in the panel from 1973Q1 to 2016Q4, the period for which NFCI data are available.

TABLE 2.2 ABOUT HERE

Table 2.2 reports the summary estimation results. The table reports the following for each specification: the quartiles of the estimated coefficients across countries; the percentage of countries for which the estimated coefficients are significant; the percentage of countries for which the dynamic quantile test based on the last four lags of the hit sequence is not rejected; and the average tick loss. All tests are performed at the 5% significance level. The NFCI maintains its prominence when additional regressors are included in the baseline QR-NFCI specification. TS and GF are relevant for a number of countries, and their relevance increases with the forecast horizon. HP and the SV provide only minor improvements. Overall, the in-sample results confirm that the NFCI is the main downside risk predictor, and a few other variables (the TS and GF) may improve prediction.

In-sample GARCH Analysis

Next, we estimate the four GARCH specifications for each country in the panel to obtain insights into the volatility dynamics of the series. The GARCH models are estimated by quasi-maximum likelihood using data from 1961Q1 to 2016Q4. After estimating these models, we compute the skewness and kurtosis of the GDP growth rates standardized by their corresponding conditional volatility to learn the shape of the conditional GDP growth distribution.

TABLE 2.3 ABOUT HERE

Table 2.3 reports the summary estimation results. The table reports the following for each forecasting horizon and GARCH specification: the quartiles of the estimated coefficients across countries; the quartiles of the skewness and kurtosis of the standardized residuals across countries; the average tick loss for $p = 0.05$ (computed over the dates for which the NFCI is available); the percentage of series for which a likelihood ratio test of the null hypothesis of no volatility dynamics is not rejected; the percentage of series for which the null hypothesis of no ARCH effects in the standardized residuals is not rejected; and the percentage of times for which the dynamic quantile test for $p = 0.05$ based on the last four lags of the hit sequence is not rejected. All tests are performed at the 5% significance level. The Factor GARCH(1,1) results refer to the parameters of the idiosyncratic volatility dynamics. In the following, we focus our discussion on the GARCH(1,1) results. The remaining specifications provide analogous findings.[11] The GARCH(1,1) estimation results show that the majority of countries exhibit persistent volatility dynamics, with the median persistence being 0.927. After GARCH filtering, the standardized GDP growth rates do not have residual ARCH

[11] The GJR-GARCH results show little evidence of asymmetries, and the GARCH-NFCI show that once the NFCI is included, the estimated persistence parameter $\alpha + \beta$ is smaller than in the case of the GARCH(1,1).

effects for the vast majority of countries. Additionally, the dynamic quantile test shows no evidence of residual tail dynamics after accounting for time-varying volatility for the vast majority of the series. Regarding the shape of the conditional distribution of GDP growth, we find no systematic pattern for skewness and that all series exhibit moderately fat tails. In particular, we find no significant evidence of conditional skewness for the USA.[12] This result may appear to starkly contradict the findings of Adrian et al. (2019) who emphasize that the conditional distribution of GDP growth is negatively skewed. However, the conditional distribution implied by the models considered in Adrian et al. (2019) and by the GARCH models in this section are based on different specifications and conditioning information sets. Our result implies that negative skewness is not a robust feature of the conditional distribution of GDP growth since it depends on the choice of the model and information set.

2.3.3 Out-of-sample Analysis

We recursively estimate the quantile regression and GARCH specifications considered in this study for each quarter from 1973Q1 to 2016Q4 and construct out-of-sample forecasts starting from 1983Q4. We consider a number of forecasting models. For quantile regression based forecasting, in addition to the specifications reported in Table 2.2, we consider a model that employs LASSO variable selection on specification (4) augmented with CR, which is the only remaining variable not previously considered and available throughout the entire estimation period.[13] For GARCH-based forecasting, we employ the four specifications previously described estimated by composite likelihood. We construct marginal and joint GaR forecasts at the 95% coverage level, which is the level used by the IMF. Starting the forecasting exercise from 1983Q4 implies that our out-of-sample validation is based on approximately 75% of the available data.

FIGURE 2.1 ABOUT HERE

Figure 2.1 displays a plot of different types of GaR regions for the USA. The figure displays the 1-step-ahead 95% marginal GaR, 90%/95% joint GaR, 95%/95% joint GaR and 95% joint GaR. All regions are constructed on the basis of the GARCH(1,1) model[14] introduced in Section 2.2.2, and the joint regions are based on the BJPR procedure described in Section 2.2.3. The regions depicted

[12] The skewness of the standardized GARCH(1,1) residuals for the USA is -0.0014 and is not significant.

[13] We do not consider CG, which has a slightly worse in-sample performance than CR.

[14] Figures A1 and A2 in the Appendix display further marginal and joint GaR plots for the G7 countries.

throughout this section are computed out-of-sample. A natural concern regarding joint prediction regions is their length. As shown in Figure 2.1, the joint GaR for the USA is, on average, 1.5% larger than the marginal GaR. In practice, this is a fairly wide and not particularly informative prediction region. By contrast, the $(1-q)\%$/95% joint GaR regions are substantially smaller than the 95% joint GaR. In particular, the 95%/95% joint GaR is, on average, 0.5% larger than the marginal GaR. Overall, the $(1-q)\%$/95% joint GaR prediction regions provide a balance between the slightly weaker notion of joint coverage and regions that are not excessively wide. This class of regions produces cautious lower bounds for GDP growth rates that simultaneously hold for a prespecified fraction of countries and a given coverage probability.

We backtest marginal and joint GaR forecasts using the tools introduced in Section 2.2.4. To assess the optimality of GaR forecasts with respect to different information sets, we consider four variants of the dynamic quantile test. The first test, labeled DQ Unc., is based on including no auxiliary predictors in the dynamic quantile test regression in (2.8). The second test, labeled DQ Hits, is based on setting the auxiliary predictors equal to the first four lags of the hit sequence generated by the GaR forecasts. The third test, labeled DQ NFCI, is based on setting the auxiliary predictors equal to the first four lags of the NFCI. The last test, labeled DQ Real, is based on setting the auxiliary predictors equal to the first four lags of GDP growth. The dynamic quantile tests employed for marginal GaR backtesting are based on each countries' individual auxiliary predictor series. To backtest joint GaR forecasts, we use the global NFCI as defined in Section B of the Appendix and the first principal component of the GDP growth rates. Some of the quantile regressions include the variables against which optimality is being tested; therefore, this battery of tests is particularly instructive to assess whether these quantile regressions effectively exploit the information content of the predictors.

We use historical benchmarks to evaluate the accuracy of the marginal and joint GaR forecasts. For marginal GaR, the benchmark is the recursively estimated unconditional quantile of the GDP growth rates for each country. For joint GaR, the benchmark is constructed based on the BJPR procedure described in Section 2.2.3, with the conditional mean and variance forecasts replaced by their recursively estimated unconditional counterparts.

Marginal GaR Forecasting

TABLE 2.4 AND 2.5 ABOUT HERE

Table 2.4 reports the summary backtesting results for 95% marginal GaR prediction. The table reports the following for each horizon and forecasting method:

the average empirical coverage; the average length; the percentage of series for which the backtesting tests are not rejected at the 5% significance level; and the average tick loss for the historical benchmark, with the percentage improvements for the remaining models considered. We begin by comparing the performance of each model according to the battery of backtesting tests. The unconditional dynamic quantile test shows that the majority of models provide adequate coverage for at least approximately 50% of the countries. The remaining dynamic quantile test results show that for the majority of countries and horizons, quantile regression and GARCH models are optimal with respect to information sets that include the lagged NFCI, the series of lag hit sequences or the lagged GDP growth rates. Additionally, the GARCH models typically have better backtesting performance than the quantile regression specifications. Remarkably, GaR forecasts based on GARCH models are efficient with respect to the information set that contains the NFCI more often than the QR-NFCI. At longer horizons, no model has a substantially better backtesting performance than the historical benchmark. Next, we compare the performance of each model according to the tick loss. At 1 step ahead, the baseline QR-NFCI performs best among the quantile regression models and improves tick loss by approximately 4% relative to the historical benchmark.[15] The largest specification considered has the worst tick loss performance. GARCH models exhibit gains of approximately 12% compared to the historical benchmark. At 2 steps ahead, among the quantile regression models, only the baseline specification modestly performs better than the historical benchmark. All GARCH specifications perform similarly and are approximately 7% better than the historical benchmark. At forecast horizons greater than 2, no quantile regression specification performs better than the historical benchmark, and GARCH models show only modest improvements. Table 2.5 complements these results with a Diebold-Mariano superior predictive ability test comparison analysis based on the tick loss. The table shows that the evidence of outperformance among the various methods is not strong. The QR-NFCI outperforms the GARCH(1,1) for at most two countries across all horizons, whereas the GARCH(1,1) outperforms the QR-NFCI for at most six countries across all horizons.

TABLE 2.6 ABOUT HERE

Table 2.6 provides detailed univariate GaR forecasting results for the intersection of countries in our sample and the countries considered in IMF (2017). We report the results for the historical benchmark, the QR-NFCI – the best-performing quantile regression – and the GARCH(1,1). The table reports, for each country

[15]Table A11 in the Appendix extends this finding to the set of bivariate quantile regressions including the lagged GDP and each potential predictor considered.

and forecast horizon, the empirical coverage, the length, the p-value of the unconditional dynamic quantile test and the tick loss. We also perform Diebold-Mariano equal predictive ability tests based on the tick loss between the GARCH(1,1) and QR-NFCI. In terms of empirical coverage, for the majority of countries, we cannot reject the null of correct unconditional coverage at the 5% significance level for all models and horizons. In terms of the tick loss, the GARCH(1,1) typically achieves the smallest loss. Additionally, the GARCH(1,1) significantly outperforms the QR-NFCI for approximately 3 countries out of 12 across all horizons, whereas the QR-NFCI never significantly outperforms the GARCH(1,1).

Overall, the results convey that quantile regression models and the QR-NFCI in particular do not have better performance than the standard GARCH(1,1). Moreover, the GARCH-NFCI results show that GARCH(1,1) predictions are not improved by including the NFCI. At longer horizons, no model outperform the historical benchmark.

Joint GaR Forecasting

TABLE 2.7 ABOUT HERE

Table 2.7 reports the summary backtesting results for 95% joint GaR predictions. The table reports the following for each model and forecast horizon: the empirical coverage; the average length; and the p-values of the backtesting tests considered. All tests are evaluated at the 5% significance level. The empirical coverage of all forecasts varies substantially and depends on the method used for the construction of the joint GaR. The joint marginal GaR forecasts and Bonferroni-based quantile regressions undercover by a wide margin. The poor performance of the Bonferroni-based quantile regression models is due to the small number of observations and the extreme quantile being forecast to construct the Bonferroni region. The Bonferroni-based GARCH region also undercovers by a small margin, especially at short horizons. The BJPR-based GARCH region produces regions with coverage close to the nominal for all horizons considered. In particular, the null hypothesis of correct unconditional coverage cannot be rejected for any horizon.

Overall, the joint GaR forecasts based on the GARCH(1,1) paired with the BJPR method have a better performance than the benchmark and forecasts based on quantile regressions.

Robustness Analysis

TABLE 2.8 ABOUT HERE

Real-time Forecasting One of the advantages of the GARCH(1,1) relative to the QR-NFCI is that real-time forecasting is relatively straightforward to implement. The NFCI is obtained from smoothed estimates of a dynamic factor model and is likely subject to filtering uncertainty at the sample endpoints. To provide more insights on the real-time performance of the GARCH(1,1), we compare the marginal GaR GARCH(1,1) forecasts against the historical benchmark. We construct quarterly real-time forecasts from 2000Q1 to 2019Q4 based on vintage data obtained from the OECD-MEI database. Specifically, we predict GDP growth at horizon $t + h$ by using the latest available data vintage that contains GDP at time t but not $t + 1$.[16] Table 2.8 contains the results of the forecasting exercise. The table reports the coverage and tick loss for the 12 IMF countries and the average coverage and average tick loss for the entire panel for both forward and cumulative growth rates. The results show that the GARCH(1,1) typically achieves the best performance in terms of the tick loss for most countries across all horizons. The Diebold-Mariano superior predictive ability test results show that the GARCH(1,1) outperforms the historical benchmark for 7 out 12 IMF countries one quarter ahead. However, the evidence of outperformance rapidly decays as the prediction horizon increases.

TABLE 2.9 ABOUT HERE

Cumulative Growth Forecasting We carry out GaR prediction for the (average) cumulative h-step ahead GDP growth based on GARCH and quantile regression models, in the spirit of Adrian et al. (2019). The exercise is designed following the same steps of the main forecasting exercise described in Section 2.3.3 with the h-step ahead growth rate replaced by the cumulative h-step ahead GDP growth.[17] Table 2.9 contains the marginal GaR prediction results for the cumulative growth. The backtesting tests and the tick loss provide essentially the same evidence reported for the h-step ahead growth rates. In particular, the QR-NFCI is the best performing quantile regression specification. However, it does not have better performance than any of the GARCH specifications. Table 2.5 complements these results with a Diebold-Mariano superior predictive ability test comparison analysis based on the tick loss. As in the case of the h-step ahead growth rates, the evidence of outperformance among the various methods is not strong, and in particular, the QR-NFCI outperforms the GARCH(1,1) for at most two series across all horizons. Tables A1 (in the Appendix) and 2.8 contain the

[16] If there is no data vintage that includes the value at t but not $t + 1$, we keep the first revision that includes the value at time t.

[17] It is straightforward to modify the algorithms for the computation of the marginal and joint GaR forecasts (provided in the Appendix) for the h-step ahead growth rate to compute the analogous forecasts for the h-step ahead cumulative growth.

joint GaR prediction results and the real-time marginal GaR prediction results for the cumulative growth, respectively. Again, the overall evidence is consistent with the findings reported for the h-step ahead growth rates.

Density Forecasting For each country, quarter and forecasting model (either based on the GARCH or quantile regression), we follow Adrian et al. (2019) and interpolate the 5%, 25%, 75% and 95% quantiles with the skewed Student's t distribution of Azzalini and Capitanio (2003). The densities obtained are then evaluated on the basis of the log predictive score (LPS), a proper scoring rule for density forecasts. For completeness, we assess each models performance on different parts of the distribution on the basis of the weighted LPS (Amisano and Giacomini, 2007).[18] Table A2 of the Appendix reports our findings. We find that the density forecasts obtained by quantile regression seldom perform better than the historical benchmark. In particular, the density implied by the QR-NFCI modestly improves over the historical density forecasts on the left tails for up to two quarters ahead. In contrast, GARCH-based densities perform better than the historical density benchmark uniformly across forecast horizons.

Additional Robustness Checks A number of additional robustness checks are included in the Appendix. In particular, we consider (i) alternative sample periods for the out-of-sample prediction exercise, (ii) alternative GaR coverage levels, (iii) alternative GARCH specifications, (iv) the backtesting performance of panel quantile regressions (as opposed to individual country estimation) and (v) the backtesting performance of GARCH models estimated via individual country estimation (as opposed to panel estimation). Overall, these checks provide evidence that is in line with the findings of this section.

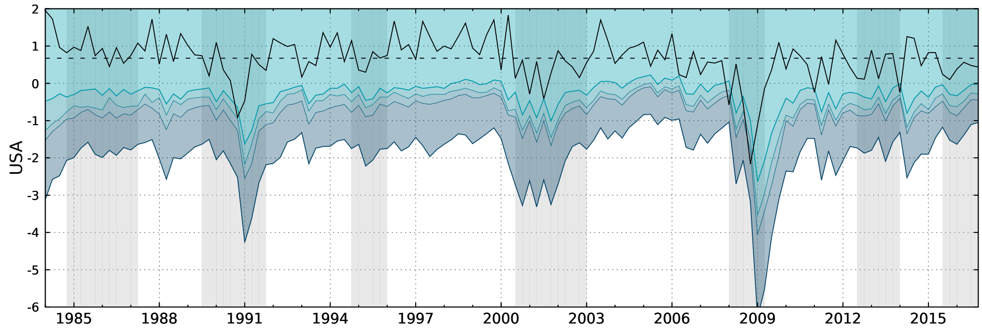
2.4 Conclusions

In this work, we conduct an out-of-sample backtesting exercise of GaR forecasts for a panel of major world economies on the basis of quantile regression and GARCH models. We rely on the standard battery of tools developed in the risk management literature to assess accuracy. Our backtesting results shows that quantile regression and GARCH forecasts have similar performance. If anything, our evidence suggests that standard volatility models such as the GARCH(1,1) are more accurate, even though the GARCH(1,1) uses no information other than GDP

[18]The different weighting functions allow the evaluation of the goodness-of-fit over different ranges of the support of the GDP growth rates.

growth. The majority of contributions in the GaR literature favor quantile regression. Quantile regression is a natural methodology in this context that directly links downside risk to predictors of interest to economists and policymakers. However, if interest lies in forecasting, then our evidence suggest caution against relying too heavily on this technique.

Figura 2.1: GaR Prediction Regions Comparison for the USA



This figure displays the 1-step-ahead 95% marginal GaR, 90%/95% joint GaR, 95%/95% joint GaR and 95% joint GaR for the USA. The lightest region represents the marginal GaR region, and the darkest region represents the 95% joint GaR for the 24 countries considered. The gray regions are OECD recession dates. We also plot the USA GDP growth rate (black line) and the average GDP growth rate over the sample period (dashed black line).

Taula 2.1: IN-SAMPLE BIVARIATE 5% QUANTILE REGRESSION ANALYSIS

	Estimation Window		h = 1	2	3	4
NFCI	1973Q1 2016Q4	Coef.	-0.49 $[-0.79\ -0.30]$	-0.39 $[-0.64\ -0.16]$	-0.33 $[-0.47\ 0.00]$	-0.22 $[-0.34\ 0.06]$
		Sig.	75.00	50.00	37.50	16.67
		TL	0.1170	0.1239	0.1276	0.1298
TS	1973Q1 2016Q4	Coef.	0.31 $[0.15\ 0.47]$	0.33 $[0.24\ 0.50]$	0.29 $[0.13\ 0.44]$	0.28 $[0.10\ 0.45]$
		Sig.	41.67	50.00	37.50	33.33
		ΔTL	-5.47	-1.54	0.00	1.49
GF	1973Q1 2016Q4	Coef.	0.06 $[0.02\ 0.11]$	0.07 $[0.04\ 0.12]$	0.08 $[0.00\ 0.12]$	0.04 $[-0.01\ 0.12]$
		Sig.	16.67	16.67	29.17	20.83
		ΔTL	-7.85	-4.16	-1.76	-0.68
HP	1973Q1 2016Q4	Coef.	0.13 $[-0.04\ 0.34]$	0.08 $[-0.02\ 0.25]$	0.11 $[-0.05\ 0.32]$	-0.02 $[-0.26\ 0.19]$
		Sig.	25.00	16.67	16.67	16.67
		ΔTL	-6.90	-4.27	-2.17	-0.94
SV	1973Q1 2016Q4	Coef.	-0.43 $[-0.50\ -0.24]$	-0.28 $[-0.66\ -0.13]$	0.03 $[-0.15\ 0.13]$	0.08 $[-0.06\ 0.19]$
		Sig.	29.17	29.17	4.17	4.17
		ΔTL	-1.42	-2.42	-2.24	-1.71
CG	1973Q1 2016Q4	Coef.	-0.12 $[-0.28\ 0.22]$	-0.17 $[-0.43\ -0.02]$	0.04 $[-0.25\ 0.23]$	-0.14 $[-0.30\ 0.06]$
		Sig.	8.33	20.83	12.50	12.50
		ΔTL	-8.42	-3.85	-3.42	-1.98
CS	1986Q2 2016Q4	Coef.	-0.27 $[-0.50\ -0.07]$	-0.15 $[-0.49\ -0.01]$	-0.04 $[-0.27\ 0.07]$	0.06 $[-0.05\ 0.17]$
		Sig.	33.33	8.33	8.33	4.17
		ΔTL	3.91	0.83	0.17	-0.26
EPU	1985Q1 2016Q4	Coef.	-0.18 $[-0.32\ 0.01]$	0.03 $[-0.07\ 0.20]$	0.04 $[-0.02\ 0.15]$	0.04 $[-0.22\ 0.28]$
		Sig.	16.67	8.33	12.50	12.50
		ΔTL	1.41	0.14	-0.10	-0.71
CR	1973Q1 2016Q4	Coef.	-0.11 $[-0.29\ 0.07]$	-0.05 $[-0.42\ 0.04]$	0.06 $[-0.17\ 0.18]$	-0.17 $[-0.38\ -0.02]$
		Sig.	8.33	12.50	12.50	12.50
		ΔTL	-7.54	-4.29	-1.98	-1.43
WUI	1996Q1 2016Q4	Coef.	0.10 $[-0.03\ 0.31]$	0.18 $[-0.01\ 0.30]$	0.15 $[-0.11\ 0.32]$	0.13 $[-0.17\ 0.42]$
		Sig.	12.50	8.33	4.17	4.17
		ΔTL	-0.42	0.01	0.54	-0.49
GPR	1985Q1 2016Q4	Coef.	0.04 $[-0.03\ 0.15]$	0.05 $[-0.01\ 0.19]$	0.07 $[-0.01\ 0.22]$	0.17 $[0.07\ 0.25]$
		Sig.	0.00	0.00	0.00	4.17
		ΔTL	-2.45	-0.32	-2.80	-3.32

This table reports the following for each forecast horizon and predictor: the period used for estimation; the quartiles of the estimated coefficients; the percentage improvement in the tick loss relative to a quantile regression with lagged GDP and NFCI; and the percentage of series for which the predictor is significant at the 5% significance level. All quantile regressions are of the form $Q_p(Y_{i\,t+h}|\mathcal{I}_t) = \alpha_0^p + \beta_0^p Y_{it} + \beta_1^p X_{ik\,t}$, where $X_{ik\,t}$ is the variable of interest. Significance is assessed on the basis of (block) bootstrap standard errors with blocks of length 4.

Taula 2.2: In-sample Multivariate 5% quantile regression Analysis

h		1				2				3				4			
		(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Y_{it}	Coef.	0.11 [−0.09 0.24]	0.14 [−0.11 0.28]	0.06 [−0.13 0.24]	0.04 [−0.04 0.16]	0.11 [−0.01 0.24]	0.11 [−0.02 0.33]	0.09 [−0.03 0.24]	0.13 [−0.06 0.18]	0.13 [−0.07 0.25]	0.08 [−0.05 0.22]	−0.01 [−0.08 0.12]	0.03 [−0.08 0.13]	0.01 [−0.17 0.14]	−0.01 [−0.13 0.13]	−0.06 [−0.22 0.13]	−0.00 [−0.24 0.15]
	Sig.	16.67	20.83	12.50	16.67	20.83	20.83	8.33	4.17	8.33	4.17	8.33	12.50	8.33	16.67	20.83	16.67
NFCI	Coef.	−0.49 [−0.79 −0.30]	−0.51 [−0.75 −0.21]	−0.46 [−0.64 −0.20]	−0.24 [−0.48 −0.05]	−0.39 [−0.64 −0.16]	−0.33 [−0.53 −0.09]	−0.18 [−0.47 −0.03]	−0.24 [−0.50 0.06]	−0.33 [−0.47 0.00]	−0.16 [−0.42 0.04]	−0.15 [−0.45 −0.01]	−0.11 [−0.45 0.11]	−0.22 [−0.34 0.06]	−0.04 [−0.29 0.13]	−0.03 [−0.27 0.14]	−0.04 [−0.30 0.23]
	Sig.	75.00	58.33	54.17	41.67	50.00	33.33	33.33	50.00	37.50	29.17	33.33	25.00	16.67	16.67	8.33	20.83
TS	Coef.	–	0.21 [0.02 0.34]	0.22 [0.07 0.30]	0.19 [0.07 0.37]	–	0.26 [0.14 0.34]	0.30 [0.18 0.38]	0.27 [0.14 0.36]	–	0.26 [0.11 0.40]	0.28 [0.09 0.40]	0.29 [0.10 0.38]	–	0.33 [0.09 0.47]	0.36 [0.03 0.46]	0.30 [0.10 0.48]
	Sig.		33.33	33.33	33.33		29.17	50.00	41.67		33.33	33.33	37.50		33.33	41.67	37.50
GF	Coef.	–	–	0.04 [−0.02 0.08]	0.00 [−0.04 0.03]	–	–	0.08 [0.01 0.11]	0.07 [0.00 0.11]	–	–	0.05 [0.01 0.14]	0.05 [−0.01 0.14]	–	–	0.03 [−0.03 0.12]	0.03 [−0.04 0.12]
	Sig.			12.50	12.50			25.00	25.00			37.50	25.00			25.00	16.67
SV	Coef.	–	–	–	−0.35 [−0.57 −0.03]	–	–	–	−0.01 [−0.27 0.16]	–	–	–	0.08 [−0.06 0.25]	–	–	–	0.13 [−0.04 0.32]
	Sig.				33.33				8.33				8.33				4.17
HP	Coef.	–	–	–	0.09 [−0.11 0.30]	–	–	–	−0.01 [−0.15 0.09]	–	–	–	0.06 [0.01 0.17]	–	–	–	−0.07 [−0.20 0.18]
	Sig.				20.83				16.67				8.33				20.83
TL		0.1170	0.1134	0.1114	0.1036	0.1239	0.1192	0.1163	0.1112	0.1276	0.1222	0.1186	0.1132	0.1298	0.1233	0.1198	0.1156
DQ Hits		70.83	83.33	79.17	87.50	83.33	95.83	91.67	91.67	79.17	87.50	91.67	83.33	87.50	87.50	87.50	83.33

This table reports the following for each forecast horizon and model considered: the quartiles of the estimated coefficients; the tick loss; and the percentage of series for which the dynamic quantile test based on the last four lags of the hit sequence does not reject the null of model optimality at the 5% significance level.

Taula 2.3: IN-SAMPLE GARCH ANALYSIS

	GARCH	GARCH-NFCI	GJR-GARCH	F-GARCH
ϕ	0.104 [−0.014 0.332]	0.104 [−0.014 0.332]	0.104 [−0.014 0.332]	0.104 [−0.014 0.332]
λ	–	–	0.512 [0.381 0.617]	–
σ_u^2	1.257 [0.935 1.833]	1.257 [0.935 1.833]	1.257 [0.935 1.833]	1.257 [0.935 1.833]
Pers.	0.927 [0.826 0.982]	0.778 [0.344 0.952]	0.936 [0.822 0.980]	0.973 [0.935 0.985]
β	0.696 [0.460 0.759]	0.519 [0.000 0.775]	0.684 [0.453 0.754]	0.751 [0.658 0.896]
NFCI	–	0.024 [0.000 0.085]	–	–
γ	–	–	0.086 [−0.097 0.146]	–
Skew.	−0.053 [−0.224 0.238]	−0.016 [−0.276 0.234]	−0.047 [−0.275 0.305]	−0.031 [−0.205 0.238]
Kurt.	4.552 [3.942 5.853]	4.180 [3.523 4.916]	4.346 [3.742 6.386]	4.669 [3.656 5.755]
LR Test	100.00	100.00	100.00	100.00
ARCH-LM	95.83	91.67	95.83	100.00
TL	0.1195	0.1175	0.1199	0.1165
DQ Hits	91.67	79.17	83.33	83.33

This table reports the following for each of the GARCH models considered: the quartiles of the estimated coefficients; the quartiles of the skewness and kurtosis of the standardized GARCH residuals; the tick loss; the percentage of countries for which a likelihood ratio test of the null hypothesis of no persistence is not rejected; the percentage of series for which the null hypothesis of no ARCH effects on each model's standardized residuals is not rejected; and the percentage of series for which the dynamic quantile test based on the last four lags of the hit sequence is not rejected. We perform an additional likelihood ratio test for the null hypothesis that $\gamma = 0$ on the GJR-GARCH model. The test is rejected for 16.67% of the series. Similarly, likelihood ratio tests for $\beta = 0$ and $\theta_{\text{NFCI}} = 0$ in the GARCH-NFCI specification are rejected for approximately 45.83 and 50% of the series, respectively. In particular, for the USA, both tests are rejected. All tests are performed at the 5% significance level. The likelihood ratio tests are computed based on the Gaussian likelihood. Additionally, the conditional volatility of the factor is described by $\beta = 0.195$, and the persistence is 0.714. The details of the estimated models can be found in the Appendix.

Taula 2.4: 95% MARGINAL GAR FORECAST EVALUATION

h	Method	Model	Cov.	Length	Unc.	Hits	NFCI	Real	TL
1	Benchmark	Historical	94.44	5.422	70.83	41.67	58.33	62.50	0.1398
	QR	NFCI	92.77	5.170	66.67	41.67	79.17	62.50	3.88
		NFCI+TS	91.13	5.079	54.17	45.83	50.00	54.17	-0.09
		NFCI+TS+GF	90.72	5.086	58.33	50.00	45.83	58.33	-1.19
		Full	89.39	5.147	37.50	29.17	29.17	33.33	-19.14
		LASSO	90.25	5.151	50.00	29.17	50.00	45.83	-6.86
	GARCH	GARCH	93.34	5.115	75.00	66.67	87.50	87.50	11.97
		GARCH-NFCI	94.44	5.200	95.83	70.83	83.33	91.67	11.68
		GJR-GARCH	93.31	5.116	75.00	66.67	87.50	87.50	11.79
		F-GARCH	93.47	5.130	62.50	54.17	87.50	79.17	**14.84**
2	Benchmark	Historical	94.47	5.427	75.00	87.50	91.67	75.00	0.1410
	QR	NFCI	92.75	5.257	75.00	75.00	79.17	70.83	0.47
		NFCI+TS	90.62	5.166	62.50	87.50	79.17	75.00	-3.99
		NFCI+TS+GF	91.13	5.187	66.67	83.33	83.33	75.00	-4.22
		Full	89.54	5.172	54.17	79.17	54.17	70.83	-36.40
		LASSO	90.55	5.157	54.17	66.67	66.67	62.50	-8.46
	GARCH	GARCH	94.15	5.245	87.50	91.67	87.50	91.67	7.80
		GARCH-NFCI	95.04	5.374	95.83	87.50	100.00	95.83	**8.76**
		GJR-GARCH	93.99	5.245	87.50	91.67	87.50	91.67	7.07
		F-GARCH	94.15	5.275	83.33	83.33	87.50	87.50	7.51
3	Benchmark	Historical	94.36	5.433	75.00	87.50	95.83	87.50	0.1420
	QR	NFCI	92.69	5.314	66.67	83.33	75.00	75.00	-3.74
		NFCI+TS	90.42	5.195	62.50	62.50	66.67	70.83	-9.40
		NFCI+TS+GF	90.80	5.238	66.67	62.50	75.00	75.00	-9.19
		Full	89.04	5.256	50.00	58.33	54.17	58.33	-31.19
		LASSO	89.46	5.231	54.17	66.67	54.17	58.33	-14.19
	GARCH	GARCH	93.85	5.316	91.67	87.50	95.83	87.50	3.66
		GARCH-NFCI	95.32	5.464	91.67	91.67	95.83	91.67	**4.40**
		GJR-GARCH	93.88	5.316	91.67	91.67	95.83	87.50	3.33
		F-GARCH	94.04	5.352	91.67	79.17	91.67	91.67	3.54
4	Benchmark	Historical	94.32	5.440	75.00	91.67	87.50	83.33	0.1427
	QR	NFCI	92.09	5.324	75.00	79.17	79.17	87.50	-11.09
		NFCI+TS	90.31	5.214	66.67	83.33	66.67	87.50	-12.44
		NFCI+TS+GF	89.79	5.174	66.67	79.17	70.83	75.00	-15.00
		Full	88.79	5.187	54.17	70.83	62.50	62.50	-48.39
		LASSO	88.82	5.194	58.33	70.83	62.50	70.83	-19.63
	GARCH	GARCH	93.86	5.341	91.67	91.67	87.50	91.67	2.85
		GARCH-NFCI	95.38	5.502	87.50	87.50	95.83	91.67	2.36
		GJR-GARCH	93.83	5.339	91.67	91.67	87.50	91.67	2.81
		F-GARCH	94.09	5.376	87.50	91.67	95.83	91.67	**3.00**

This table reports the following for each forecast horizon and forecasting method: the average empirical coverage; the average length; the percentage of series that pass GaR adequacy tests at the 5% significance level (DQ Unc, Hits, NFCI and Real); and the percentage improvement in each model's average tick loss relative to the historical benchmark. The performance of the best forecasting method in terms of the tick loss is highlighted in boldface.

Taula 2.5: Superior Predictive Ability Test Pairwise Comparison

h		*Forward*				*Cumulative*			
		Historical	QR-NFCI	GARCH	GARCH-NFCI	Historical	QR-NFCI	GARCH	GARCH-NFCI
1	Historical	-	0.00	4.17	0.00	-	0.00	4.17	0.00
	QR-NFCI	16.67	-	4.17	0.00	16.67	-	4.17	0.00
	GARCH	37.50	16.67	-	16.67	37.50	16.67	-	16.67
	GARCH-NFCI	37.50	16.67	4.17	-	37.50	16.67	4.17	-
2	Historical	-	12.50	0.00	0.00	-	12.50	0.00	0.00
	QR-NFCI	16.67	-	4.17	4.17	16.67	-	4.17	8.33
	GARCH	20.83	12.50	-	12.50	25.00	20.83	-	20.83
	GARCH-NFCI	16.67	12.50	12.50	-	20.83	16.67	4.17	-
3	Historical	-	20.83	0.00	0.00	-	4.17	0.00	0.00
	QR-NFCI	4.17	-	8.33	4.17	8.33	-	4.17	8.33
	GARCH	8.33	16.67	-	20.83	20.83	12.50	-	12.50
	GARCH-NFCI	8.33	12.50	16.67	-	16.67	4.17	16.67	-
4	Historical	-	20.83	4.17	8.33	-	4.17	0.00	8.33
	QR-NFCI	4.17	-	0.00	4.17	4.17	-	0.00	8.33
	GARCH	8.33	25.00	-	29.17	4.17	12.50	-	4.17
	GARCH-NFCI	8.33	12.50	8.33	-	0.00	4.17	4.17	-

This table reports the results of the pairwise Diebold-Mariano tests of superior predictive ability at the 5% significance level. Each entry represents the percentage of countries for which the model in the column is outperformed by the model in the row.

Taula 2.6: 95% MARGINAL GaR FORECAST EVALUATION: IMF COUNTRIES

h	Country	Historical				QR NFCI				GARCH				
		Cov.	Length	Unc.	TL	Cov.	Length	Unc.	TL	Cov.	Length	Unc.	TL	DM
1	AUS	99.24	5.016	0.025	0.0943	99.24	5.015	0.025	0.0958	96.21	4.431	0.523	**0.0734*****	-4.390
	CAN	95.45	3.728	0.811	0.0937	94.70	3.365	0.873	**0.0684**	90.91	3.263	0.031	0.0714	0.402
	DEU	96.21	4.792	0.523	0.1291	95.45	4.508	0.811	**0.1165**	95.45	4.714	0.811	0.1265	1.388
	ESP	90.15	5.182	0.011	0.1045	91.67	5.209	0.079	0.1032	92.42	5.052	0.175	**0.0700*****	-3.960
	FRA	93.18	3.804	0.338	0.0628	90.15	3.693	0.011	**0.0504**	93.94	3.872	0.576	0.0570	0.787
	GBR	97.73	5.359	0.151	0.1007	96.21	5.230	0.523	0.0813	96.21	4.844	0.523	**0.0803**	-0.071
	ITA	90.91	3.735	0.031	0.0922	95.45	3.730	0.811	0.0796	91.67	3.535	0.079	**0.0599**	-1.460
	JPN	87.12	4.616	0.000	0.1464	88.64	4.753	0.001	0.1357	90.91	4.994	0.031	**0.1245**	-1.096
	KOR	97.73	9.091	0.151	0.2191	96.97	8.501	0.299	0.2116	95.45	7.832	0.811	**0.1870**	-0.806
	MEX	93.18	4.201	0.338	0.2137	96.97	4.236	0.299	**0.1604**	90.15	3.878	0.011	0.1869	0.924
	SWE	97.73	5.796	0.151	0.1352	92.42	5.682	0.175	0.1577	96.21	5.514	0.523	**0.1251****	-2.088
	USA	97.73	3.377	0.151	0.0909	93.94	2.929	0.576	0.0705	95.45	2.841	0.811	**0.0685**	-0.316
2	AUS	99.24	5.023	0.026	0.0940	99.24	5.097	0.026	0.1006	96.95	4.500	0.276	**0.0718*****	-8.988
	CAN	96.18	3.733	0.649	0.0949	93.13	3.404	0.473	0.0930	93.13	3.343	0.412	**0.0859**	-0.827
	DEU	96.18	4.802	0.558	**0.1279**	94.66	4.680	0.869	0.1313	94.66	4.735	0.881	0.1421	0.884
	ESP	90.08	5.185	0.164	0.1076	88.55	5.050	0.052	0.0921	93.89	5.075	0.691	**0.0895**	-0.503
	FRA	93.13	3.811	0.505	0.0648	93.13	3.970	0.511	0.0615	95.42	3.940	0.882	**0.0590**	-0.441
	GBR	97.71	5.365	0.277	0.1019	94.66	5.092	0.916	**0.0861**	95.42	4.881	0.883	0.0945	0.445
	ITA	90.84	3.739	0.237	0.0939	94.66	3.878	0.901	0.0950	93.13	3.692	0.472	**0.0835***	-1.929
	JPN	87.02	4.617	0.000	0.1492	87.79	4.921	0.001	**0.1411**	86.26	4.872	0.000	0.1551	1.554
	KOR	97.71	9.084	0.141	0.2186	96.95	8.455	0.276	0.2082	96.18	8.206	0.491	**0.2073**	-0.091
	MEX	93.13	4.218	0.498	**0.2215**	93.89	4.619	0.584	0.2274	90.84	3.686	0.038	0.2300	0.128
	SWE	97.71	5.801	0.195	0.1349	96.95	5.869	0.400	0.1400	96.18	5.497	0.601	**0.1306**	-1.096
	USA	97.71	3.384	0.250	0.0921	91.60	3.020	0.204	0.0782	93.89	2.778	0.690	**0.0765**	-0.162
3	AUS	99.23	5.030	0.026	0.0942	100.00	5.015	0.009	0.0929	98.46	4.703	0.066	**0.0818*****	-3.297
	CAN	96.15	3.739	0.659	0.0954	92.31	3.581	0.373	**0.0897**	92.31	3.414	0.398	0.0969	0.494
	DEU	96.92	4.811	0.349	**0.1268**	96.92	4.659	0.349	0.1274	95.38	4.728	0.870	0.1334	0.510
	ESP	90.00	5.187	0.160	0.1098	84.62	5.084	0.038	0.1225	91.54	5.072	0.295	**0.1046***	-1.723
	FRA	93.08	3.816	0.495	**0.0663**	93.85	3.939	0.686	0.0722	94.62	3.931	0.896	0.0682	-1.458
	GBR	97.69	5.370	0.283	0.1035	96.15	5.242	0.693	**0.0985**	96.15	4.871	0.690	0.0996	0.074
	ITA	90.77	3.743	0.231	**0.0953**	84.62	3.632	0.005	0.1139	91.54	3.745	0.317	0.0979	-1.381
	JPN	86.92	4.617	0.000	0.1501	88.46	4.845	0.001	**0.1443**	85.38	4.769	0.000	0.1527	0.990
	KOR	97.69	9.093	0.144	0.2191	96.92	8.501	0.283	0.2106	96.92	8.308	0.283	**0.2048**	-0.520
	MEX	93.85	4.233	0.665	0.2229	96.15	4.733	0.567	**0.2092**	90.00	3.980	0.052	0.2347	0.877
	SWE	97.69	5.806	0.200	0.1362	98.46	5.936	0.105	0.1429	96.15	5.482	0.611	**0.1299***	-1.661
	USA	96.92	3.390	0.500	0.0923	96.92	3.262	0.408	0.0844	96.15	3.005	0.681	**0.0839**	-0.049
4	AUS	99.22	5.036	0.027	0.0945	99.22	5.081	0.027	0.0965	98.45	4.740	0.068	**0.0827*****	-5.113
	CAN	96.12	3.744	0.669	0.0960	96.12	3.845	0.669	**0.0953**	94.57	3.581	0.881	0.1033	0.636
	DEU	96.90	4.819	0.357	**0.1264**	95.35	4.612	0.859	0.1269	96.90	4.727	0.357	0.1319	0.852
	ESP	89.92	5.189	0.155	**0.1102**	81.40	5.043	0.026	0.1542	89.15	5.075	0.132	0.1152^{**}	-2.117
	FRA	93.02	3.822	0.485	**0.0662**	89.92	3.909	0.256	0.0781	93.80	3.905	0.678	0.0714	-0.684
	GBR	97.67	5.376	0.288	0.1039	95.35	5.159	0.906	**0.1006**	96.12	4.930	0.699	0.1048	0.320
	ITA	90.70	3.746	0.226	**0.0951**	86.82	3.606	0.009	0.1145	91.47	3.725	0.253	0.1029	-0.956
	JPN	86.82	4.617	0.000	0.1508	89.92	4.893	0.057	0.1525	87.60	4.834	0.010	**0.1492**	-0.274
	KOR	97.67	9.102	0.148	0.2194	96.12	8.537	0.582	0.2376	96.12	8.247	0.574	**0.2017****	-1.998
	MEX	93.80	4.248	0.655	0.2239	96.90	4.813	0.357	**0.2029**	89.92	3.975	0.051	0.2303	0.820
	SWE	97.67	5.811	0.205	0.1363	96.12	5.905	0.622	0.1560	97.67	5.605	0.261	**0.1361****	-2.510
	USA	97.67	3.396	0.288	0.0926	94.57	3.154	0.874	0.0946	95.35	2.987	0.906	**0.0864**	-1.278

This table reports detailed backtesting statistics for selected countries. The performance of the best model in terms of the tick loss is highlighted. We perform Diebold-Mariano tests of equal predictive ability between the tick loss of the quantile regressions based on the NFCI and that of the GARCH(1,1). $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Taula 2.7: 95% JOINT GAR FORECAST EVALUATION

h	Method	Model	Cov.	Length	Unc.	Hits	NFCI	Real
1	Benchmark	Historical	90.91	7.976	0.031	0.003	0.024	0.121
	QR + Bonf.	QR-NFCI	51.52	6.111	0.000	0.000	0.000	0.000
		NFCI+TS	34.85	5.834	0.000	0.000	0.000	0.000
		NFCI+GF+TS	34.09	5.673	0.000	0.000	0.000	0.000
		Full	25.00	5.563	0.000	0.000	0.000	0.000
		LASSO	38.64	7.111	0.000	0.000	0.000	0.000
	GARCH + Marg.	GARCH	35.61	5.115	0.000	0.000	0.000	0.000
		GARCH-NFCI	43.94	5.200	0.000	0.000	0.000	0.000
		GJR-GARCH	36.36	5.116	0.000	0.000	0.000	0.000
		F-GARCH	33.33	5.130	0.000	0.000	0.000	0.000
	GARCH + Bonf.	GARCH	87.12	7.320	0.000	0.000	0.000	0.000
		GARCH-NFCI	84.09	7.055	0.000	0.000	0.000	0.000
		GJR-GARCH	86.36	7.293	0.000	0.000	0.000	0.000
		F-GARCH	85.61	7.302	0.000	0.000	0.000	0.000
	GARCH + BJPR	GARCH	94.70	7.747	0.873	0.021	0.619	0.455
		GARCH-NFCI	95.45	7.783	0.811	0.004	0.535	0.828
		GJR-GARCH	94.70	7.708	0.873	0.021	0.619	0.455
		F-GARCH	95.45	7.605	0.811	0.319	0.575	0.872
2	Benchmark	Historical	88.55	7.886	0.016	0.999	0.026	0.108
	QR + Bonf.	QR-NFCI	54.20	6.595	0.000	0.999	0.031	0.000
		NFCI+TS	39.69	6.241	0.000	0.000	0.000	0.000
		NFCI+GF+TS	40.46	6.157	0.000	0.000	0.000	0.000
		Full	31.30	5.864	0.000	0.000	0.000	0.000
		LASSO	41.98	7.560	0.000	0.000	0.000	0.000
	GARCH + Marg.	GARCH	44.27	5.245	0.000	0.000	0.999	0.000
		GARCH-NFCI	53.44	5.374	0.000	0.000	0.000	0.999
		GJR-GARCH	42.75	5.245	0.000	0.999	0.999	0.000
		F-GARCH	44.27	5.275	0.000	0.000	0.000	0.000
	GARCH + Bonf.	GARCH	87.79	8.050	0.008	0.999	0.000	0.297
		GARCH-NFCI	86.26	7.534	0.001	0.999	0.000	0.098
		GJR-GARCH	87.02	8.034	0.006	0.999	0.000	0.326
		F-GARCH	89.31	7.893	0.029	0.998	0.256	0.077
	GARCH + BJPR	GARCH	94.66	8.464	0.899	0.999	0.597	0.999
		GARCH-NFCI	95.42	8.668	0.862	0.000	0.592	0.990
		GJR-GARCH	94.66	8.442	0.899	0.999	0.597	0.999
		F-GARCH	95.42	8.175	0.869	0.106	0.583	0.996
3	Benchmark	Historical	88.46	7.858	0.028	0.999	0.016	0.228
	QR + Bonf.	QR-NFCI	51.54	6.570	0.000	0.999	0.000	0.000
		NFCI+TS	33.85	6.262	0.000	0.000	0.000	0.000
		NFCI+GF+TS	32.31	6.122	0.000	0.000	0.000	0.000
		Full	30.77	5.930	0.000	0.999	0.000	0.000
		LASSO	39.23	7.953	0.000	0.000	0.000	0.000
	GARCH + Marg.	GARCH	46.15	5.316	0.000	0.000	0.999	0.000
		GARCH-NFCI	54.62	5.464	0.000	0.000	0.000	0.999
		GJR-GARCH	46.15	5.316	0.000	0.000	0.999	0.000
		F-GARCH	46.15	5.352	0.000	0.000	0.000	0.000
	GARCH + Bonf.	GARCH	89.23	8.049	0.055	0.999	0.999	0.244
		GARCH-NFCI	84.62	7.543	0.000	0.999	0.999	0.001
		GJR-GARCH	89.23	8.023	0.055	0.999	0.999	0.244
		F-GARCH	89.23	7.959	0.008	0.205	0.999	0.055
	GARCH + BJPR	GARCH	94.62	8.425	0.887	0.339	0.533	0.932
		GARCH-NFCI	94.62	8.452	0.887	0.339	0.533	0.932
		GJR-GARCH	94.62	8.392	0.887	0.339	0.533	0.932
		F-GARCH	95.38	8.247	0.859	0.969	0.467	0.742
4	Benchmark	Historical	87.60	7.761	0.021	0.999	0.105	0.049
	QR + Bonf.	QR-NFCI	48.06	6.587	0.000	0.000	0.000	0.000
		NFCI+TS	34.11	6.335	0.000	0.000	0.000	0.000
		NFCI+GF+TS	33.33	6.167	0.000	0.000	0.000	0.000
		Full	23.26	5.878	0.000	0.999	0.000	0.000
		LASSO	35.66	7.834	0.000	0.000	0.999	0.000
	GARCH + Marg.	GARCH	46.51	5.341	0.000	0.000	0.000	0.000
		GARCH-NFCI	57.36	5.502	0.000	0.000	0.000	0.000
		GJR-GARCH	45.74	5.339	0.000	0.000	0.000	0.000
		F-GARCH	45.74	5.376	0.000	0.000	0.003	0.000
	GARCH + Bonf.	GARCH	89.92	8.051	0.084	0.325	0.572	0.014
		GARCH-NFCI	85.27	7.657	0.005	0.000	0.999	0.100
		GJR-GARCH	89.15	8.035	0.061	0.360	0.393	0.217
		F-GARCH	88.37	7.943	0.015	0.171	0.999	0.999
	GARCH + BJPR	GARCH	95.35	8.393	0.886	0.316	0.631	0.900
		GARCH-NFCI	94.57	8.506	0.876	0.500	0.588	0.857
		GJR-GARCH	95.35	8.366	0.886	0.316	0.631	0.900
		F-GARCH	96.12	8.333	0.616	0.983	0.645	0.934

This table reports the following for each forecast horizon and forecasting method: the average empirical joint coverage; the average length; and the p-values of the GaR adequacy tests considered (DQ Unc., Hits, NFCI and Real).

Taula 2.8: 95% Marginal GaR Forecast Evaluation: Real-Time GARCH

		Forward				*Cumulative*			
		Historical		GARCH		Historical		GARCH	
h	Country	Cov.	TL	Cov.	TL	Cov.	TL	Cov.	TL
	AUS	100.00	0.0836	98.73	**0.0648*****	100.00	0.0837	98.73	**0.0648*****
	CAN	96.20	0.0976	94.94	**0.0761*****	96.20	0.0976	94.94	**0.0761*****
	DEU	94.94	0.1275	93.67	**0.1137***	94.94	0.1275	93.67	**0.1137***
	ESP	93.67	0.1220	86.08	**0.0861**	93.67	0.1254	86.08	**0.0861**
	FRA	93.67	0.0706	93.67	**0.0543**	94.94	0.0682	93.67	**0.0543**
	GBR	96.20	0.1079	96.20	**0.0978**	96.20	0.1080	96.20	**0.0978**
1	ITA	89.87	0.1215	89.87	**0.0830****	89.87	0.1208	89.87	**0.0830****
	JPN	89.87	0.1717	89.87	**0.1390**	89.87	0.1720	89.87	**0.1390**
	KOR	98.73	0.1323	96.20	**0.1188****	98.73	0.1316	96.20	**0.1188***
	MEX	97.47	0.2362	97.47	**0.1444****	97.47	0.2376	97.47	**0.1444****
	SWE	93.67	0.2931	92.41	**0.2913**	93.67	0.2933	92.41	**0.2913**
	USA	97.47	0.0874	96.20	**0.0752***	97.47	0.0876	96.20	**0.0752***
	World	94.41	0.1637	92.56	**0.1513**	94.46	0.1638	92.56	**0.1513**
	AUS	100.00	0.0837	98.73	**0.0666**	100.00	0.0603	98.72	**0.0447**
	CAN	96.20	0.0976	94.94	**0.0795*****	96.15	0.0932	94.87	**0.0736****
	DEU	94.94	0.1275	93.67	**0.1199**	92.31	0.1190	91.03	**0.1128**
	ESP	93.67	0.1254	89.87	**0.0822**	89.74	0.1289	84.62	**0.0912**
	FRA	94.94	0.0682	93.67	**0.0632**	94.87	0.0707	93.59	**0.0596**
	GBR	96.20	0.1080	96.20	**0.1003**	94.87	0.1009	94.87	**0.0970**
2	ITA	89.87	0.1208	89.87	**0.1018**	89.74	0.1203	88.46	**0.1026**
	JPN	89.87	0.1720	89.87	**0.1595**	88.46	0.1320	89.74	**0.1266**
	KOR	98.73	0.1316	98.73	**0.1231**	96.15	**0.1106**	96.15	0.1192
	MEX	97.47	0.2376	97.47	**0.2185**	96.15	0.2089	96.15	**0.1825****
	SWE	93.67	0.2933	89.87	**0.2734**	94.87	**0.1885**	93.59	0.1935
	USA	97.47	0.0876	94.94	**0.0768**	96.15	0.0786	97.44	**0.0683**
	World	94.46	0.1638	93.62	**0.1505**	92.79	0.1393	92.09	**0.1300**
	AUS	100.00	0.0833	98.73	**0.0724**	100.00	0.0486	100.00	**0.0372**
	CAN	96.20	0.0978	96.20	**0.0844*****	96.10	0.0910	96.10	**0.0706****
	DEU	94.94	0.1279	94.94	**0.1223**	92.21	0.1124	89.61	**0.1090**
	ESP	93.67	0.1269	93.67	**0.0998****	88.31	0.1346	88.31	**0.1022****
	FRA	94.94	0.0680	93.67	**0.0671**	93.51	0.0666	92.21	**0.0619**
	GBR	96.20	0.1081	96.20	**0.1022**	93.51	**0.0933**	93.51	0.0992
3	ITA	89.87	0.1205	92.41	**0.1114**	93.51	**0.1091**	87.01	0.1102
	JPN	89.87	**0.1685**	89.87	0.1723	89.61	**0.1220**	89.61	0.1231
	KOR	98.73	0.1296	98.73	**0.1208**	96.10	**0.0999**	93.51	0.1025
	MEX	97.47	**0.2282**	97.47	0.2291	94.81	0.1968	94.81	**0.1909**
	SWE	93.67	0.2939	93.67	**0.2094**	93.51	0.1970	87.01	**0.1866**
	USA	94.94	0.0874	94.94	**0.0851**	96.10	0.0752	94.81	**0.0720**
	World	94.25	0.1632	94.15	**0.1549**	92.32	0.1341	90.48	**0.1334**
	AUS	100.00	0.0834	98.73	**0.0728**	100.00	0.0433	100.00	**0.0346**
	CAN	96.20	0.0972	96.20	**0.0867**	94.74	0.0758	92.11	**0.0698**
	DEU	94.94	0.1282	94.94	**0.1225**	90.79	0.1104	86.84	**0.1063**
	ESP	93.67	0.1277	93.67	**0.1049****	86.84	0.1421	86.84	**0.1104****
	FRA	94.94	**0.0674**	93.67	0.0707	92.11	0.0686	93.42	**0.0644**
	GBR	96.20	0.1082	96.20	**0.1062**	93.42	**0.0902**	92.11	0.1015
4	ITA	91.14	0.1197	89.87	**0.1158**	89.47	0.1021	86.84	**0.1007**
	JPN	89.87	**0.1683**	89.87	0.1754	90.79	**0.1090**	92.11	0.1161
	KOR	98.73	0.1299	98.73	**0.1241**	96.05	**0.0717**	93.42	0.0817
	MEX	97.47	**0.2118**	97.47	0.2342	94.74	**0.1751**	94.74	0.1811
	SWE	93.67	0.2943	93.67	**0.2177**	92.11	**0.1251**	90.79	0.1382
	USA	94.94	0.0876	94.94	**0.0863**	94.74	**0.0739**	93.42	0.0756
	World	94.20	0.1620	93.99	**0.1537**	90.84	**0.1320**	89.75	0.1371

This table reports detailed backtesting statistics for the cumulative and forward growth rates for selected countries. The performance of the best model in terms of the tick loss is highlighted. We also report the Diebold-Mariano test of superior predictive ability on the tick loss of each model against the historical benchmark. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.10$.

Taula 2.9: 95% MARGINAL GAR FORECAST EVALUATION: CUMULATIVE GROWTH

h	Method	Model	Cov.	Length	Unc.	Hits	NFCI	Real	TL
1	Benchmark	Historical	94.44	5.422	70.83	41.67	58.33	62.50	0.1398
	QR	NFCI	92.77	5.170	66.67	41.67	79.17	62.50	3.88
		NFCI+TS	91.13	5.079	54.17	45.83	50.00	54.17	-0.09
		NFCI+TS+GF	90.72	5.086	58.33	50.00	45.83	58.33	-1.19
		Full	89.39	5.147	37.50	29.17	29.17	33.33	-19.14
		LASSO	90.25	5.151	50.00	29.17	50.00	45.83	-6.86
	GARCH	GARCH	93.34	5.115	75.00	66.67	87.50	87.50	11.97
		GARCH-NFCI	94.44	5.200	95.83	70.83	83.33	91.67	11.68
		GJR-GARCH	93.31	5.116	75.00	66.67	87.50	87.50	11.79
		F-GARCH	93.47	5.130	62.50	54.17	87.50	79.17	**14.84**
2	Benchmark	Historical	93.73	4.899	87.50	66.67	70.83	83.33	0.1085
	QR	NFCI	91.98	4.730	70.83	87.50	75.00	83.33	2.26
		NFCI+TS	88.17	4.598	50.00	66.67	70.83	75.00	-3.28
		NFCI+TS+GF	88.96	4.635	58.33	50.00	62.50	66.67	-4.10
		Full	87.69	4.640	29.17	58.33	33.33	54.17	-18.71
		LASSO	88.36	4.635	50.00	75.00	70.83	54.17	-10.74
	GARCH	GARCH	93.10	4.769	87.50	79.17	62.50	87.50	10.17
		GARCH-NFCI	94.66	4.887	95.83	95.83	79.17	100.00	9.63
		GJR-GARCH	93.10	4.772	87.50	79.17	62.50	95.83	9.63
		F-GARCH	93.61	4.785	95.83	83.33	75.00	79.17	**11.68**
3	Benchmark	Historical	93.14	4.722	95.83	75.00	70.83	87.50	0.0995
	QR	NFCI	91.09	4.616	70.83	62.50	70.83	87.50	-3.70
		NFCI+TS	87.63	4.493	62.50	62.50	66.67	79.17	-10.70
		NFCI+TS+GF	88.62	4.524	58.33	79.17	79.17	83.33	-10.02
		Full	86.89	4.514	41.67	70.83	79.17	75.00	-32.36
		LASSO	86.63	4.479	45.83	79.17	66.67	70.83	-16.62
	GARCH	GARCH	92.37	4.639	91.67	83.33	75.00	87.50	4.68
		GARCH-NFCI	94.20	4.789	95.83	83.33	91.67	95.83	**5.12**
		GJR-GARCH	92.31	4.641	87.50	83.33	83.33	91.67	3.96
		F-GARCH	92.60	4.666	87.50	87.50	83.33	91.67	3.78
4	Benchmark	Historical	92.44	4.611	95.83	79.17	91.67	87.50	0.0948
	QR	NFCI	90.31	4.557	75.00	83.33	83.33	87.50	-7.13
		NFCI+TS	86.37	4.417	45.83	66.67	66.67	70.83	-15.93
		NFCI+TS+GF	86.14	4.421	50.00	54.17	79.17	66.67	-17.81
		Full	84.46	4.397	37.50	58.33	41.67	66.67	-49.40
		LASSO	85.05	4.379	37.50	50.00	58.33	70.83	-28.74
	GARCH	GARCH	91.63	4.571	87.50	70.83	87.50	91.67	**1.92**
		GARCH-NFCI	93.35	4.715	91.67	83.33	95.83	91.67	1.82
		GJR-GARCH	91.51	4.575	87.50	79.17	83.33	87.50	1.10
		F-GARCH	91.96	4.593	83.33	75.00	87.50	91.67	0.41

This table reports the following for each forecast horizon and forecasting method: the average empirical coverage; the average length; the percentage of series that pass GaR adequacy tests at the 5% significance level (DQ Unc, Hits, NFCI and Real); and the percentage improvement in each model's average tick loss relative to the historical benchmark. The performance of the best forecasting method in terms of the tick loss is highlighted in boldface.

Apèndix A

Appendix

A.1 Methodology

A.1.1 Simulation Algorithms

This section describes the simulation based procedure employed to construct GaR prediction regions. Algorithm 1 contains the pseudo-code of the procedure that we employ to construct marginal GaR predictions (in the case of the AR(1)-GARCH(1,1)). Algorithm 2 contains the pseudo-code of the procedure that we employ to construct joint GaR predictions on the basis of the BJPR method (in the case of the AR(1)-GARCH(1,1)).

Algorithm 1 MARGINAL GAR ALGORITHM FOR THE AR(1)-GARCH(1,1) MODEL

INPUTS:

(i) Y: the $T \times n$ matrix of GDP growth rates with (t, i) entry given by $Y_{i\,t}$

(ii) μ: the $T \times n$ matrix of conditional means with (t, i) entry given by $\mu_{i\,t}$

(iii) σ^2: the $T \times n$ matrix conditional variances with (t, i) entry given by $\sigma^2_{i\,t}$

(iv) θ: parameter vector the AR(1)-GARCH(1,1) $\theta = (\theta_1', \ldots, \theta_n')$ with $\theta_i = (\phi_{i\,0}, \phi_{i\,1}, \sigma_i^2, \alpha_i, \beta_i)$

(v) p: the quantile of interest

(vi) h: forecast horizon

(vii) S: number of bootstrap replications.

PROCEDURE:

1. Construct the $T \times n$ matrix of residuals $\widehat{Z}$ with (t, i) entry given by $\widehat{Z}_{it} = (Y_{i\,t} - \mu_{i\,t})/\sqrt{\sigma^2_{i\,t}}$
2. Construct the $S \times 1$ vector B with (s) entry given by b_s, uniform draw from $\{1, \ldots, T-h\}$
3. For i in $\{1, \ldots, n\}$ do

For s in $\{1, \ldots, S\}$ do

Construct the bootstrap innovation $\widetilde{Z}^s_{i\,j}$ for $j = 1, \ldots, h$ where $\widetilde{Z}^s_{i\,j} = \widehat{Z}_{i\,b_s+j-1}$

For j in $\{1, \ldots, h\}$ do

If j = 1 do

$$\mu_{i\,T+1|T} = \phi_{i\,0} + \phi_{i\,1} Y_{i\,T}$$

$$\sigma^{2\,s}_{i\,T+1|T} = \sigma_i^2(1 - \alpha_i - \beta_i) + \alpha_i(Y_{i\,T} - \mu_{i\,T})^2 + \beta_i \sigma^2_T,$$

$$\widetilde{Y}^s_{i\,T+1|T} = \mu_{i\,T+1|T} + \sqrt{\sigma^2_{i\,T+1|T}}\widetilde{Z}^s_{i\,j}$$

else

$$\mu^s_{i\,T+j|T} = \phi_{i\,0} + \phi_{i\,1}\widetilde{Y}^s_{i\,T+j-1|T}$$

$$\sigma^{2\,s}_{i\,T+j|T} = \sigma_i^2(1 - \alpha_i - \beta_i) + \alpha_i(\widetilde{Y}^s_{i\,T+j-1|T} - \mu^s_{i\,T+j-1|T})^2 + \beta_i \sigma^{2\,s}_{T+j-1|T}$$

$$\widetilde{Y}^s_{i\,T+j|T} = \mu^s_{i\,T+j|T} + \sqrt{\sigma^{2\,s}_{i\,T+j|T}}\widetilde{Z}^s_{i\,j}$$

4. Construct the marginal GaR region as $GaR^M_{T+h|T} = (\widehat{F}^{-1}_{\widetilde{Y}_{1\,T+h|T}}(p), \infty) \times \ldots \times (\widehat{F}^{-1}_{\widetilde{Y}_{n\,T+h|T}}(p), \infty)$, where $\widehat{F}^{-1}_{\widetilde{Y}_{i\,T+h|T}}(p)$ is the p-th empirical quantile of $\{\widetilde{Y}^s_{i\,T+h|T}\}^S_{s=1}$ for $i = 1, \ldots, n$.

OUTPUT: The h-step-ahead $(1-p)\%$ marginal GaR region $GaR^M_{T+h|T}$

Algorithm 2 JOINT GAR ALGORITHM FOR THE AR(1)-GARCH(1,1) MODEL

INPUTS:

(i) Y: the $T \times n$ matrix of GDP growth rates with (t, i) entry given by $Y_{i\,t}$

(ii) μ: the $T \times n$ matrix of conditional means with (t, i) entry given by $\mu_{i\,t}$

(iii) σ^2: the $T \times n$ matrix conditional variances with (t, i) entry given by $\sigma^2_{i\,t}$

(iv) θ: parameter vector the AR(1)-GARCH(1,1) $\theta = (\theta'_1, \ldots, \theta'_n)$ with $\theta_i = (\phi_{i\,0}, \phi_{i\,1}, \sigma^2_i, \alpha_i, \beta_i)$

(v) q: the percentage of the system for which coverage is desired; $q = 0$ is $(1 - p)\%$ joint GaR.

(vi) p: the quantile of interest

(vii) h: forecast horizon

(viii) S: number of bootstrap replications.

PROCEDURE:

1. Carry out steps 1 to 3 of Algorithm 1
2. Construct the $S \times n$ matrix $\widetilde{W}$ with (s, i) entry given by $\tilde{w}_{s\,i} = (\widetilde{Y}^s_{i\,T+h|T} - \hat{\mu}_{i\,T+h|T})/\sqrt{\hat{\sigma}^2_{i\,T+h|T}}$ where $\hat{\mu}_{i\,T+h|T}$ and $\hat{\sigma}^2_{i\,T+h|T}$ denote the mean and variance of $\widetilde{Y}_{i\,T+h|T}$ across the S simulations.
3. Define $U_s^{\lceil qn \rceil}$ as the $\lceil qn \rceil$-smallest element of $\{\tilde{w}_{s\,i}\}_{i=1}^n$
4. Compute $d_p^{\lceil qn \rceil}$ as the p% empirical quantile of $\{U_s^{\lceil qn \rceil}\}_{s=1}^S$
5. Construct the $(1 - q)$%/$(1 - p)$% joint GaR region as $GaR^{Jq,BJPR}_{T+h|T} = (GaR^{Jq,BJPR}_{i\,T+h|T}, \infty) \times \cdots \times (GaR^{Jq,BJPR}_{n\,T+h|T}, \infty)$ where $GaR^{Jq,BJPR}_{i\,T+h|T} = \hat{\mu}_{i\,T+h|T} + d_p^{\lceil qn \rceil}\sqrt{\sigma^2_{i\,T+h|T}}$ for $i = 1, \ldots, n$.

OUTPUT: The $(1 - q)$%/$(1 - p)$% joint GaR region $GaR^{Jq,BJPR}_{T+h|T}$

A.1.2 Additional GARCH Methodology

GARCH Specifications. We rely on the following GARCH-type specification for the GDP growth rates

$$Y_{i\,t+1} = \mu_{i\,t+1|t} + \sqrt{\sigma^2_{i\,t+1|t}}\,Z_{i\,t+1} \quad Z_{i\,t+1} \overset{i.i.d}{\sim} \mathcal{D}_{Z_i}(0,1), \tag{A.1}$$

where $\mu_{i\,t+1|t}$ denotes the 1-step-ahead conditional mean, $\sigma^2_{i\,t+1|t}$ is the 1-step-ahead conditional variance and $\mathcal{D}_{Z_i}(\mu,\sigma^2)$ is a location-scale distribution with mean μ and variance σ^2. For the conditional mean, we rely on an AR(1) model, that is,

$$\mu_{i\,t+1|t} = \phi_{i\,0} + \phi_{i\,1} Y_{i\,t}.$$

For the conditional variance, in addition to the GARCH(1,1) model described in the text, we rely on the different conditional variance specifications described below.

The GJR-GARCH model of **?** models the conditional volatility as

$$\sigma^2_{i\,t+1|t} = \sigma^2_i(1-\alpha_i-\beta_i-1/2\gamma_i) + \left(\alpha_i + \gamma_i \mathbb{1}_{\{\varepsilon_{i\,t}<0\}}\right)\varepsilon^2_{i\,t} + \beta_i \sigma^2_{i\,t+1|t},$$

where $\varepsilon_{i\,t} = Y_{i\,t} - \phi_{i\,0} - \phi_{i\,1} Y_{i\,t-1}$.

The Factor GARCH (**?**) assumes that $\varepsilon_{i\,t}$ can be decomposed into two orthogonal components, namely, a common factor and an idiosyncratic shock series. In particular, we have

$$\varepsilon_{i\,t} = \lambda_{i\,h}\varepsilon_{F\,t} + \xi_{i\,t},$$

where $\lambda_{i\,h}$ is the loading of each series on the common factor, $\varepsilon_{F\,t}$ is the common factor and $\xi_{i\,t}$ is an idiosyncratic shock that is uncorrelated with the common factor. In the Factor GARCH model, these shocks have time-varying volatility

$$\mathbb{V}ar_t(\varepsilon_{F\,t+1}) = \sigma^2_{F\,t+1} \text{ and } \mathbb{V}ar_t(\xi_{i\,t+1}) = \sigma^2_{I\,i\,t+1},$$

which follow a GARCH(1,1) process

$$\begin{aligned}
\sigma^2_{I\,i\,t+1|t} &= \sigma^2_i(1-\alpha_i-\beta_i) + \alpha_i \xi^2_{i\,t} + \beta_i \sigma^2_{I\,i\,t|t-1} \\
\sigma^2_{F\,t+1|t} &= \sigma^2_F(1-\alpha_F-\beta_F) + \alpha_F \varepsilon^2_{F\,t} + \beta_F \sigma^2_{F\,t|t-1}.
\end{aligned}$$

We refer the reader to **?** for details of the estimation of the Factor GARCH. We note that $\varepsilon_{F\,t}$ is estimated as the first principal component of $(\varepsilon_{1\,t},\dots,\varepsilon_{n\,t})'$.

Lastly, we consider a model labelled GARCH-NFCI, that is a GARCH-X model that includes the NFCI as a predictor of conditional volatility. The conditional volatility in the GARCH-NFCI model is given by

$$\sigma^2_{i\,t+1|t} = \sigma^2_i(1-\alpha_i-\beta_i-\kappa_i\theta_i) + \alpha_i \varepsilon^2_{i\,t} + \beta_i \sigma^2_{i\,t+1|t} + \theta_i X_{i\,t},$$

where $\varepsilon_{i\,t} = Y_{i\,t} - \phi_{i\,0} - \phi_{i\,1} Y_{i\,t-1}$, $\kappa_i = \mathbb{E}[X_{i\,t}]/\sigma_i^2$ and $X_{i\,t}$ is the NFCI of country i at time t.[1]

Direct GARCH Modeling. We define the direct GARCH model as

$$Y_{i\,t+h} = \mu_{i\,t+h|t} + \sqrt{\sigma^2_{i\,t+h|t}}\, Z_{i\,t+h} \quad Z_{i\,t+h}|\mathcal{F}_t \sim \mathcal{D}_{Z_i}(0,1), \tag{A.2}$$

where $\mu_{i\,t+h|t}$ is the h-step-ahead conditional mean, $\sigma^2_{i\,t+h|t}$ is the h-step-ahead conditional variance and $\mathcal{D}_{Z_i}(0,1)$ denotes a zero-mean and unit-variance location-scale distribution. The specification in equation (A.2) is labeled "direct", as it directly models the h-step-ahead growth rate $Y_{i\,t+h}$ as a function of the information set available in period t (**?**). The h-step-ahead conditional distribution implied by (A.2) is

$$Y_{i\,t+h}|\mathcal{I}_t \sim \mathcal{D}_{Z_i}(\mu_{i\,t+h|t}, \sigma^2_{i\,t+h|t}),$$

and the corresponding h-step-ahead $p\%$ conditional quantile is given by

$$Q_p(Y_{i\,t+h}|\mathcal{I}_t) = \mu_{i\,t+h|t} + \sqrt{\sigma^2_{i\,t+h|t}} F^{-1}_{Z_i}(p), \tag{A.3}$$

where $F^{-1}_{Z_i}(\cdot)$ is the inverse cumulative distribution function of $\mathcal{D}_{Z_i}$.

To make the model in (A.2) operational, one must specify the conditional mean $\mu_{i\,t+h|t}$, the conditional variance $\sigma^2_{i\,t+h|t}$ and the innovation distribution $\mathcal{D}_{Z_i}$. The conditional mean equation is an h-step-ahead direct AR(1) process, that is,

$$\mu_{i\,t+h|t} = \phi_{i\,0} + \phi_{i\,h} Y_{i\,t}.$$

The conditional variance is a direct version of the standard GARCH(1,1), that is,

$$\sigma^2_{i\,t+h|t} = \sigma_i^2(1 - \alpha_i - \beta_i) + \alpha_i \varepsilon^2_{i\,t} + \beta_i \sigma^2_{i\,t+h-1|t-1}, \tag{A.4}$$

where $\varepsilon_{i\,t} = Y_{i\,t} - \mu_{i\,t|t-1}$ with $\mu_{i\,t|t-1}$ being the 1-step-ahead conditional mean implied by an AR(1) process, and the variance equation parameters satisfy the constraints $\omega_i > 0$, $\alpha_i > 0$, $\beta_i \geq 0$ and $\alpha_i + \beta_i < 1$. Note that when $h = 1$, the model in (A.4) corresponds to the standard GARCH(1,1) specification. It is straightforward to see via recursive substitution that this is equivalent to

$$\sigma^2_{i\,t+h|t} = \frac{\omega_i}{1 - \beta_i} + \alpha_i \sum_{k=0}^{\infty} \beta_i^k \varepsilon^2_{i\,t-k}, \tag{A.5}$$

[1] Clearly, this model relies on NFCI data to be estimated. Therefore, the estimation window may differ from the remaining methods.

which emphasizes that the direct GARCH(1,1) h-step-ahead conditional variance is an exponentially weighted average of present and past GDP growth shocks.

We estimate the direct GARCH(1,1) by a naïve QML estimator that maximizes the Gaussian pseudo-log-likelihood given by

$$\hat{\theta} = \arg\max_{\theta} \sum_{t=1}^{T-h} -\frac{1}{2} \log \sigma^2_{i\,t+h|t} - \frac{1}{2} \frac{(Y_{i\,t+h} - \mu_{t+h|t})^2}{\sigma^2_{i\,t+h|t}} .$$

We remark that we label this estimator "naïve" since it does not take into account the serial dependence of the multi-step innovations of the model.

Pooled GARCH Estimation. In this work, we rely on a panel GARCH estimation approach named composite likelihood (CL), which is described in detail in Pakel et al. (2011). We describe the approach for the panel GARCH(1,1) specification given by

$$\begin{aligned} Y_{i\,t+1} &= \sqrt{\sigma^2_{i\,t+1|t}} Z_{i\,t+1} \quad Z_{i\,t+1} \sim \mathcal{D}_{Z_i}(0,1) \\ \sigma^2_{i\,t+1|t} &= \sigma^2_i(1-\alpha-\beta) + \alpha Y^2_{i\,t} + \beta \sigma^2_{i\,t|t-1}, \end{aligned}$$

where $\sigma^2_i > 0$, $\alpha, \beta > 0$ and $\alpha + \beta < 1$. In particular, we emphasize that in the equation above, the dynamic coefficients α and β are common across series, whereas the unconditional variance parameter σ^2_i is series specific. We are interested in estimating the common parameters $\theta = (\alpha, \beta)'$ and the nuisance parameter $\gamma = (\sigma^2_1, \ldots, \sigma^2_n)$. CL estimation consists of relying on a two-step procedure to estimate these parameters. In the first step, the γ parameter vector is estimated by the sample variance of each series, that is,

$$\hat{\gamma}_i = \hat{\sigma}^2_i = \frac{1}{T} \sum_{i=1}^{T} Y^2_{i\,t} ,$$

and in the second step, the θ parameter vector is estimated by maximizing the so-called composite likelihood given by

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{T-1} \frac{1}{n} \sum_{i=1}^{n} \left(-\frac{1}{2} \log \hat{\sigma}^2_{i\,t+1|t} - \frac{1}{2} \frac{Y^2_{i\,t+1}}{\hat{\sigma}^2_{i\,t+1|t}} \right) ,$$

where

$$\hat{\sigma}^2_{i\,t+1|t} = \hat{\sigma}^2_i(1-\alpha-\beta) + \alpha Y^2_{i\,t} + \beta \sigma^2_{i\,t|t-1}.$$

The properties of this estimator are studied in Pakel et al. (2011).

A.2 Data Construction Details

In this section, we provide details of the data construction and imputation. Note that some of the predictors that we consider are not available for the entire sample period, and some of the country-specific predictors are not available for all countries. Moreover, some of the predictors have missing values.

National Financial Conditions: The IMF has constructed and made available the NFCIs for 11 advanced and 10 emerging market economies. Indexes for advanced economies start in 1973Q1, whereas indexes for emerging market economies start in 1991Q1. The last observation available for all indexes is 2016Q4. Similarly to IMF (2017), we construct a measure of Global NFCI as the average of the country-specific available observations at each point in time. For countries for which the NFCI exists but starts later than 1973Q1, we impute the initial observations with the Global NFCI, which is standardized to have unit variance and multiplied by the standard deviation of the series for which the imputation is needed. This procedure guarantees that the imputation maintains the original series scale properties. We perform this procedure for Canada, Italy, South Korea, Mexico, Spain, Switzerland and Sweden. Additionally, we assign the Global NFCI to countries that do not have their own NFCI published. This is the case for Austria, Belgium, Denmark, Finland, Greece, Iceland, Ireland, Luxembourg, the Netherlands, Norway and Portugal. For the remaining countries – Australia, France, Germany, Japan, the United Kingdom and the United States – we directly use the data provided by the IMF. The NFCI data are available from the IMF's website (https://www.imf.org/∼/media/Files/Publications/GFSR/2017/October/chapter-3/csv-data/data-appendix.ashx?la=en).

Credit Spread: We use the US credit spread, which is available in the St.Louis Fed website (https://fred.stlouisfed.org/series/BAA10Y) as a proxy for the credit spreads for all countries considered. The data are available at a daily frequency: we use the last observation in each quarter to represent the quarterly value for this series. Missing observations are replaced by the average of the two adjacent non-missing observations.

Credit Gap and Credit to GDP Ratio: We use the credit statistics available on the BIS website (https://www.bis.org/statistics/c_gaps.htm). The data start from 1961 and cover all countries in our sample, with the exception of Iceland. We consider the first differences of the credit gaps and the log differences of the credit ratios. We construct a measure of the global credit gap and global credit ratio as the averages of available observations at each point in time. For Iceland, we impute the global credit gap. To handle missing observations in all other countries, we use the global credit gap standardized to have unit variance and multiplied by the standard deviation of each series.

Term Spread: We construct the term spread as long-term interest rates minus short-term interest rates. We obtain long and short term interest rates from the OECD website (https://data.oecd.org/interest/long-term-interest-rates.htm and https://data.oecd.org/interest/short-term-interest-rates.htm). Term spreads are available for all countries in our sample. We construct a measure of global term spreads, which we scale appropriately to impute the missing data for individual countries.

Stock Variance: We construct a measure of realized volatility as the mean of the squared daily returns of the S&P500 obtained from the CRSP.

House Prices: We use property price statistics available on the BIS website (https://www.bis.org/statistics/pp/pp_detailed.xlsx) to construct our measure of housing prices. We use the data from each countries' statistical agency whenever possible. We consider pure prices of all residential dwellings whenever possible. For Japan and Greece, we resort to the price per square meter. We construct a measure of global property prices, which we use to impute missing observations. We note that the BIS does not publish property prices for Iceland; therefore, we use the global property prices for this country. The property statistics for Japan are available on a semiannual basis, and we thus impute quarterly data as the average of two adjacent observations.

World Uncertainty Index: We use the World Uncertainty Index made available on the economic policy uncertainty website (https://www.policyuncertainty.com/wui_quarterly.html). For Iceland and Luxembourg, we use a Global World Uncertainty Index constructed as the average WUI of all countries in our sample. The WUI makes up for a balanced panel, and, therefore, there is no need to impute data on the time series.

Economic Policy Uncertainty Index: We consider Baker et al. (2016)'s *Economic Policy Uncertainty index* for Australia, Canada, France, Germany, Italy, Mexico, South Korea, the United Kingdom and the United States. Similarly constructed indexes are available for Japan (**?**), Greece (**?**), Ireland (**?**), the Netherlands (**?**), Spain (**?**) and Sweden (**?**). For the remaining countries, we use a Global EPU constructed as the average of all available observations at each point in time. The quarterly index is computed as the mean of the daily index over each quarter.

Geopolitical Risk Index: We consider Caldara and Iacoviello (2018)'s monthly Geopolitical Risk Index. We construct the quarterly index by averaging daily observations.

A.2.1 Real-time GaR Prediction Data

We construct quarterly real-time forecasts from 2000Q1 to 2019Q4 based on vintage data obtained from the OECD-MEI revisions database (https://www.oecd.org/sdd/

oecdmaineconomicindicatorsmei.htm) .[2] For each country and quarter t, we construct a dataset that contains GDP at time t but not $t+1$.[3] For some countries, some editions of GDP data are shorter – i.e., they have less observations – than previous releases. In such cases, the dataset is constructed by merging the short edition with the observations lost from the long edition. If this procedure generates breaks in the series, then we drop all observations up to the breakpoint.[4] Finally, we use the latest available vintage at the time of writing (see footnoote 2) to backtest the forecasts.

A.3 Additional Tables and Figures

A.3.1 Additional Robustness Checks

This section reports the tables described in Section 2.3.3 of the paper and a number of additional robustness checks.

Alternative Sample Periods We assess the robustness of our results to alternative out-of-sample specifications. First, we consider the period from 1983Q4 to 2007Q4 as our out-of-sample. Stopping short of the great financial crisis allows us to evaluate the extent to which our findings are influenced by the latest crisis episode. Next, we consider a second out-of-sample specification from 1996Q1 to 2016Q4. Based on a longer estimation window, this exercise is intended to verify whether our findings are driven by short estimation periods. Table A.3 reports the backtesting results for the historical benchmark, QR-NFCI and GARCH(1,1). First, for marginal GaR, we find that in the pre-crisis sample specification, the QR-NFCI performs worse than the historical benchmark across all horizons considered, whereas the GARCH(1,1) displays strong performance gains. In the second sample specification, the QR-NFCI performs better than the historical benchmark for up to 2 quarters ahead. Moreover, GARCH(1,1) retains its good forecasting performance and improves the tick loss over the tick loss of the historical benchmark for all horizons considered. Next, for the joint GaR, we find that the BJPR-based GARCH(1,1) is the only method that provides satisfactory prediction regions across all sample specifications and forecast horizons considered.

[2] Data downloaded on 06/06/2020

[3] If there is no data vintage that includes GDP at time t but not $t+1$, we keep the first revision that includes GDP at time t.

[4] We define a break as a consecutive change in GDP larger than 20 times the average change within a given revision, both in absolute values.

Alternative Coverage Levels. We assess the robustness of our findings to alternative coverage levels. In particular, we consider the 99%, 90%, 85% and 80% coverage levels. Table A.4 of the Appendix highlights that the GARCH models provide a smaller average tick loss than the QR-NFCI across all quantiles and forecast horizons considered. Moreover, after two quarters ahead, the QR-NFCI is outperformed by the historical benchmark for all coverage levels considered. In addition, the GARCH models outperform the historical benchmark for forecasting the 99% quantile at one quarter ahead but are outperformed for all longer horizons. The performance of the QR-NFCI deteriorates at more extreme quantiles. As shown in Table A.5 of the Appendix, the standard GARCH(1,1) model in conjunction with the BJPR produces GaR regions with the correct coverage for all coverage levels.

Alternative GARCH Specifications. We investigate the robustness of our findings to alternative specifications of the conditional mean, the variance and innovation distribution, and alternative forecasting strategies (direct or iterated). First, we model the conditional mean of the GDP growth rates with an AR(4) process and regress the series of the GDP growth rates on a constant (labeled Constant). Second, we entertain a model where the conditional mean of GDP is modeled as an AR(1) process and the conditional variance is constant. Third, we explore the impact of parametrizing the GARCH innovation distribution to a skewed Student's t distribution.[5] Lastly, we consider direct forecasts of the conditional mean and volatility dynamics. Section A.1.2 provides details of the direct GARCH model. Table A.6 reports the summary backtesting results for the marginal and joint GaR forecasting with the alternative model specifications considered. According to the backtesting tests, all models that allow for time-varying volatility dynamics exhibit similar performance. In terms of the tick loss, the time-varying volatility models dominate the model with constant variance. Overall, the GARCH-based GaR regions constructed with parsimonious conditional mean specifications perform better than those constructed with larger models, and iterated GARCH forecasts generally exhibit lower tick loss and produce smaller regions than their direct counterparts. We emphasize that all GARCH specifications considered outperform the QR-NFCI in terms of the tick loss.

Panel QR Specification. We follow Adrian et al. (2018) and model the conditional quantiles with the following specification:

$$Q_p(Y_{i\,t+h}|\mathcal{I}_t) = \alpha_i^p + \beta_{i\,0}^p Y_{i\,t} + \beta_1^p X_{i1\,t} + \ldots + \beta_K^p X_{iK\,t}\,. \tag{A.6}$$

[5] We use the skewed Student's t distribution introduced in **?**.

Accordingly, we allow each country to have its own intercept and autoregressive parameters, and we restrict the remaining coefficients to be the same across all countries. We do this in an attempt to improve the QR forecasts by estimating the model parameters from the entire panel rather than from individual countries. The model is estimated by minimizing the sum of individual tick losses. Table A.7 reports our findings. In particular, we find that the panel quantile specifications are outperformed by their country-specific counterparts in forecasting one quarter ahead. For longer horizons, the panel quantile specifications present modest performance gains but are nonetheless outperformed by GARCH models.

Univariate GARCH Estimation. We explore the extent to which composite likelihood estimation is relevant in the construction of accurate GARCH-based GaR forecasts. We construct and backtest GaR forecasts based on GARCH models that are estimated individually for each country. Tables A.8 and A.9 report our findings. We keep the QR models for comparison purposes, and the QR results are those reported in Tables 2.4 and 2.7 of the main paper. We find that the GARCH models that use cross-sectional information produce GaR forecasts with slightly better performance according to the tick loss than their univariate counterparts. Generally, the results reported in Section 2.3.3 remain qualitatively the same. In particular, GaR forecasts based on a univariate GARCH estimation display correct coverage and are optimal with respect to the information sets that include the NFCI for most series considered.

$(1-q)$**%/95% Joint GaR Forecasting.** As a final robustness check, we construct and backtest the $(1-q)$%/95% joint GaR regions. The joint GaR regions considered in this exercise are constructed on the basis of the BJPR introduced in Section 2.2 and the forecasting procedure described in Section 2.3.3. Table A.10 reports the summary backtesting results for $q = 5\%$ and the $q = 10\%$ BJPR-based historical and GARCH joint GaR. For comparison purposes, we also report the results for the joint GaR. The table reports the following for each forecast horizon, model and GaR region considered: the empirical coverage; the average length; and the p-values of the backtesting tests considered. We find that both models exhibit empirical coverage close to the nominal level. In fact, we cannot reject that all $(1-q)$%/95% joint GaR regions provide the correct unconditional coverage for either of the models. However, the GARCH-based regions are more accurate according to the majority of the backtesting tests and thus prevail over the historical benchmark.

Taula A.1: 95% JOINT GAR FORECAST EVALUATION: CUMULATIVE

h	Method	Model	Cov.	Length	Unc.	Hits	NFCI	Real
1	Benchmark	Historical	90.91	7.976	0.031	0.003	0.024	0.121
	QR + Bonf.	QR-NFCI	51.52	6.111	0.000	0.000	0.000	0.000
		NFCI+TS	34.85	5.834	0.000	0.000	0.000	0.000
		NFCI+GF+TS	34.09	5.673	0.000	0.000	0.000	0.000
		Full	25.00	5.563	0.000	0.000	0.000	0.000
		LASSO	38.64	7.111	0.000	0.000	0.000	0.000
	GARCH + Marg.	GARCH	35.61	5.115	0.000	0.000	0.000	0.000
		GARCH-NFCI	43.94	5.200	0.000	0.000	0.000	0.000
		GJR-GARCH	36.36	5.116	0.000	0.000	0.000	0.000
		F-GARCH	33.33	5.130	0.000	0.000	0.000	0.000
	GARCH + Bonf.	GARCH	87.12	7.320	0.000	0.000	0.000	0.000
		GARCH-NFCI	84.09	7.055	0.000	0.000	0.000	0.000
		GJR-GARCH	86.36	7.293	0.000	0.000	0.000	0.000
		F-GARCH	85.61	7.302	0.000	0.000	0.000	0.000
	GARCH + BJPR	GARCH	94.70	7.747	0.873	0.021	0.619	0.455
		GARCH-NFCI	95.45	7.783	0.811	0.004	0.535	0.828
		GJR-GARCH	94.70	7.708	0.873	0.021	0.619	0.455
		F-GARCH	95.45	7.605	0.811	0.319	0.575	0.872
2	Benchmark	Historical	98.47	7.886	0.102	0.781	0.707	0.677
	QR + Bonf.	QR-NFCI	46.56	5.406	0.000	0.000	0.000	0.000
		NFCI+TS	29.77	5.163	0.000	0.000	0.000	0.000
		NFCI+GF+TS	34.35	5.125	0.000	0.999	0.000	0.000
		Full	23.66	4.957	0.000	0.000	0.000	0.000
		LASSO	36.64	5.950	0.000	0.000	0.000	0.000
	GARCH + Marg.	GARCH	40.46	4.769	0.000	0.999	0.000	0.000
		GARCH-NFCI	50.38	4.887	0.000	0.000	0.999	0.000
		GJR-GARCH	41.98	4.772	0.000	0.000	0.000	0.999
		F-GARCH	43.51	4.785	0.000	0.000	0.999	0.000
	GARCH + Bonf.	GARCH	87.79	6.377	0.035	0.588	0.234	0.999
		GARCH-NFCI	87.79	6.215	0.038	0.391	0.000	0.668
		GJR-GARCH	87.79	6.365	0.035	0.588	0.234	0.999
		F-GARCH	87.02	6.289	0.007	0.000	0.999	0.999
	GARCH + BJPR	GARCH	93.89	6.383	0.715	0.641	0.990	0.999
		GARCH-NFCI	93.13	6.483	0.585	0.804	0.844	0.895
		GJR-GARCH	93.89	6.377	0.715	0.641	0.990	0.999
		F-GARCH	93.89	6.358	0.715	0.641	0.990	0.999
3	Benchmark	Historical	97.69	7.858	0.204	0.787	0.713	0.684
	QR + Bonf.	QR-NFCI	47.69	5.114	0.000	0.000	0.000	0.000
		NFCI+TS	29.23	4.812	0.000	0.000	0.000	0.000
		NFCI+GF+TS	31.54	4.815	0.000	0.000	0.000	0.000
		Full	23.08	4.770	0.000	0.000	0.000	0.000
		LASSO	36.92	5.326	0.000	0.000	0.000	0.000
	GARCH + Marg.	GARCH	40.77	4.639	0.000	0.999	0.000	0.000
		GARCH-NFCI	50.00	4.789	0.000	0.000	0.999	0.000
		GJR-GARCH	40.77	4.641	0.000	0.000	0.000	0.000
		F-GARCH	44.62	4.666	0.000	0.000	0.000	0.000
	GARCH + Bonf.	GARCH	84.62	5.949	0.004	0.007	0.917	0.080
		GARCH-NFCI	83.85	5.798	0.005	0.000	0.747	0.258
		GJR-GARCH	84.62	5.943	0.004	0.007	0.917	0.080
		F-GARCH	83.08	5.903	0.001	0.000	0.001	0.000
	GARCH + BJPR	GARCH	90.00	6.015	0.134	0.463	0.003	0.522
		GARCH-NFCI	90.00	6.105	0.134	0.463	0.003	0.522
		GJR-GARCH	90.00	6.003	0.134	0.463	0.003	0.522
		F-GARCH	90.00	6.008	0.134	0.463	0.003	0.522
4	Benchmark	Historical	97.67	7.761	0.295	0.000	0.254	0.254
	QR + Bonf.	QR-NFCI	44.19	4.945	0.000	0.000	0.000	0.000
		NFCI+TS	31.78	4.685	0.000	0.000	0.000	0.000
		NFCI+GF+TS	33.33	4.659	0.000	0.000	0.000	0.000
		Full	24.03	4.598	0.000	0.000	0.000	0.000
		LASSO	35.66	5.078	0.000	0.000	0.000	0.999
	GARCH + Marg.	GARCH	40.31	4.571	0.000	0.000	0.999	0.000
		GARCH-NFCI	50.39	4.715	0.000	0.000	0.000	0.000
		GJR-GARCH	40.31	4.575	0.000	0.000	0.999	0.999
		F-GARCH	42.64	4.593	0.000	0.000	0.000	0.000
	GARCH + Bonf.	GARCH	84.50	5.713	0.011	0.177	0.181	0.052
		GARCH-NFCI	80.62	5.569	0.001	0.044	0.000	0.999
		GJR-GARCH	83.72	5.708	0.004	0.108	0.396	0.000
		F-GARCH	84.50	5.701	0.019	0.470	0.349	0.999
	GARCH + BJPR	GARCH	89.92	5.989	0.130	0.325	0.150	0.540
		GARCH-NFCI	86.05	5.811	0.026	0.303	0.066	0.745
		GJR-GARCH	89.15	5.981	0.105	0.278	0.201	0.666
		F-GARCH	89.15	5.928	0.107	0.326	0.006	0.080

This table reports the following for each forecast horizon and forecasting method: the average empirical joint coverage; the average length; and the p-values of the GaR adequacy tests considered (DQ Unc., Hits, NFCI and Real).

Taula A.2: PREDICTIVE DENSITY EVALUATION

h	Method	Model	Unweighted	Left Tails	Body	Right Tails
1	Benchmark	Historical	1.4933	0.8191	0.3722	0.6742
	QR	QR-NFCI	-0.78	3.40	5.84	-5.86
		QR-NFCI+TS	-9.87	-7.88	-4.41	-12.27
		QR-NFCI+TS+GF	-15.92	-14.32	-4.79	-17.86
		Full	-21.41	-23.59	-10.44	-18.77
		Lasso	-23.36	-26.88	-12.81	-19.08
	GARCH	GARCH	9.89	10.41	11.87	9.26
		GARCH-NFCI	**11.14**	**14.47**	**15.23**	7.10
		GJR-GARCH	9.84	10.29	11.95	9.30
		F-GARCH	10.23	10.56	10.50	**9.84**
2	Benchmark	Historical	1.4999	0.8248	0.3735	0.6752
	QR	QR-NFCI	-2.71	0.93	-0.91	-7.17
		QR-NFCI+TS	-11.99	-10.57	-8.15	-13.71
		QR-NFCI+TS+GF	-14.20	-13.80	-7.35	-14.69
		Full	-32.43	-31.10	-24.87	-34.05
		Lasso	-29.55	-27.70	-21.77	-31.82
	GARCH	GARCH	6.71	5.46	8.53	8.24
		GARCH-NFCI	**9.42**	**11.66**	**12.82**	6.68
		GJR-GARCH	6.73	5.31	8.59	**8.46**
		F-GARCH	6.89	5.86	7.30	8.15
3	Benchmark	Historical	1.5016	0.8271	0.3741	0.6745
	QR	QR-NFCI	-13.36	-10.51	-7.29	-16.86
		QR-NFCI+TS	-23.58	-20.71	-15.10	-27.10
		QR-NFCI+TS+GF	-26.06	-17.16	-15.97	-36.98
		Full	-33.30	-29.74	-26.51	-37.66
		Lasso	-29.15	-25.33	-22.98	-33.82
	GARCH	GARCH	4.38	2.89	4.81	6.21
		GARCH-NFCI	**6.93**	**9.47**	**9.69**	3.82
		GJR-GARCH	4.28	2.63	4.92	**6.30**
		F-GARCH	4.14	2.91	3.78	5.66
4	Benchmark	Historical	1.5045	0.8284	0.3751	0.6761
	QR	QR-NFCI	-16.47	-16.61	-9.77	-16.30
		QR-NFCI+TS	-34.13	-31.25	-25.45	-37.67
		QR-NFCI+TS+GF	-38.80	-37.61	-25.81	-40.24
		Full	-48.55	-43.31	-40.70	-54.97
		Lasso	-40.90	-40.97	-30.14	-40.82
	GARCH	GARCH	3.02	1.21	3.66	5.25
		GARCH-NFCI	**5.31**	**7.57**	**7.70**	2.55
		GJR-GARCH	3.09	1.23	3.78	**5.36**
		F-GARCH	3.15	1.67	3.13	4.95

This table reports the percentage improvement in each model's average weighted log predictive score relative to the historical benchmark for each forecast horizon, forecasting method and weight function. The performance of the best forecasting method in terms of the score functions considered is highlighted in boldface.

Taula A.3: Robustness Check: Alternative Sample specifications

Panel A: Marginal GaR

		1983Q4-2007Q4							1996Q4-2016Q4						
h	Model	Cov.	Length	Unc.	Hits	NFCI	Real	TL	Cov.	Length	Unc.	Hits	NFCI	Real	TL
	Historical	96.26	5.410	75.00	58.33	79.17	79.17	0.1285	93.98	5.409	75.00	29.17	58.33	62.50	0.1446
1	QR-NFCI	93.99	5.144	83.33	45.83	83.33	70.83	-1.76	92.57	5.112	75.00	50.00	79.17	66.67	8.66
	GARCH	94.80	5.023	95.83	83.33	87.50	91.67	11.83	93.02	5.112	83.33	58.33	83.33	83.33	13.25
	Historical	96.35	5.419	66.67	75.00	95.83	91.67	0.1287	93.85	5.413	79.17	66.67	87.50	70.83	0.1468
2	QR-NFCI	93.58	5.210	70.83	75.00	100.00	83.33	-1.52	92.68	5.183	79.17	79.17	83.33	83.33	4.13
	GARCH	95.70	5.167	87.50	91.67	87.50	95.83	11.19	93.65	5.242	91.67	62.50	75.00	79.17	9.52
	Historical	96.32	5.428	66.67	66.67	87.50	91.67	0.1294	93.72	5.417	79.17	83.33	87.50	79.17	0.1480
3	QR-NFCI	94.12	5.261	70.83	58.33	83.33	100.00	-2.83	92.23	5.246	75.00	75.00	75.00	75.00	-1.06
	GARCH	95.57	5.261	91.67	95.83	91.67	91.67	8.61	93.36	5.305	91.67	91.67	95.83	83.33	4.58
	Historical	96.28	5.437	62.50	62.50	91.67	95.83	0.1302	93.75	5.422	79.17	95.83	83.33	83.33	0.1490
4	QR-NFCI	93.88	5.307	79.17	79.17	83.33	83.33	-8.68	91.61	5.227	79.17	83.33	83.33	87.50	-10.89
	GARCH	95.61	5.301	83.33	79.17	95.83	95.83	8.59	93.44	5.319	91.67	95.83	91.67	100.00	3.98

Panel B: Joint GaR

		1983Q4-2007Q4						1996Q4-2016Q4					
h	Model	Cov.	Length	Unc.	Hits	NFCI	Real	Cov.	Length	Unc.	Hits	NFCI	Real
	Historical	89.69	7.788	0.016	0.079	0.001	0.186	92.77	8.321	0.351	0.017	0.136	0.085
1	QR-NFCI	45.36	5.992	0.000	0.000	0.000	0.000	62.65	6.249	0.000	0.000	0.000	0.000
	GARCH + BJPR	94.85	7.571	0.944	0.018	0.810	0.597	93.98	7.770	0.669	0.000	0.624	0.468
	Historical	86.46	7.676	0.002	0.999	0.001	0.037	91.46	8.224	0.267	0.000	0.391	0.441
2	QR-NFCI	47.92	6.373	0.000	0.999	0.000	0.000	58.54	6.775	0.000	0.999	0.999	0.000
	GARCH + BJPR	95.83	8.240	0.763	0.999	0.836	0.968	95.12	8.642	0.971	0.999	0.831	0.999
	Historical	86.32	7.639	0.007	0.999	0.001	0.094	92.59	8.251	0.422	0.000	0.426	0.779
3	QR-NFCI	45.26	6.346	0.000	0.999	0.000	0.000	58.02	6.737	0.000	0.000	0.000	0.000
	GARCH + BJPR	95.79	8.199	0.777	0.999	0.837	0.983	95.06	8.590	0.985	0.994	0.793	0.936
	Historical	85.11	7.497	0.005	0.999	0.000	0.000	91.25	8.193	0.246	0.999	0.287	0.304
4	QR-NFCI	43.62	6.294	0.000	0.000	0.000	0.000	53.75	6.864	0.000	0.000	0.000	0.000
	GARCH + BJPR	95.74	8.156	0.790	0.999	0.972	0.847	96.25	8.597	0.651	0.979	0.851	0.855

This table reports the forecasting exercise performed in two alternative samples that span 1983Q1-2007Q4 and 1996Q4-2016Q4. Panel A reports the results for the marginal GaR, and Panel B reports the results for the joint GaR. For each forecast horizon, forecasting method and sample split considered, we report the average empirical coverage, the average length, the percentage of the series that pass GaR adequacy tests at the 5% significance level (DQ Unc., Hits, NFCI and Real) for the marginal GaR and the p-values of each test for the joint GaR. Additionally, for the marginal GaR, we report the percentage improvement in the average tick loss relative to the historical benchmark.

Taula A.4: Alternative Coverage Levels, Marginal

		99% Coverage							90% Coverage						
h	Model	Cov.	Length	Unc.	Hits	NFCI	Real	TL	Cov.	Length	Unc.	Hits	NFCI	Real	TL
	Historical	98.61	6.720	83.33	58.33	75.00	79.17	0.0467	89.02	4.834	45.83	33.33	50.00	54.17	0.2136
1	QR-NFCI	96.21	6.018	41.67	25.00	45.83	41.67	-30.59	87.50	4.745	70.83	58.33	70.83	66.67	4.01
	GARCH	98.45	6.329	87.50	66.67	83.33	91.67	9.52	86.90	4.684	66.67	62.50	70.83	79.17	9.97
	Historical	98.63	6.732	91.67	79.17	95.83	70.83	0.0471	88.99	4.832	70.83	70.83	70.83	62.50	0.2141
2	QR-NFCI	96.25	6.456	58.33	83.33	70.83	66.67	-25.78	88.33	4.783	79.17	66.67	62.50	62.50	2.07
	GARCH	98.22	6.736	91.67	87.50	83.33	83.33	-1.47	88.20	4.769	83.33	58.33	79.17	79.17	6.90
	Historical	98.62	6.743	91.67	79.17	87.50	83.33	0.0476	89.17	4.838	70.83	70.83	79.17	75.00	0.2152
3	QR-NFCI	95.45	6.435	54.17	66.67	75.00	83.33	-41.22	88.08	4.836	83.33	70.83	70.83	79.17	-2.15
	GARCH	97.98	6.819	95.83	91.67	95.83	91.67	-9.10	88.46	4.809	83.33	83.33	83.33	79.17	4.17
	Historical	98.55	6.755	95.83	79.17	87.50	83.33	0.0480	89.15	4.843	66.67	75.00	70.83	75.00	0.2159
4	QR-NFCI	95.28	6.457	70.83	79.17	75.00	83.33	-45.44	87.89	4.870	79.17	83.33	62.50	79.17	-6.01
	GARCH	97.80	6.862	91.67	91.67	79.17	91.67	-10.38	88.63	4.834	87.50	91.67	83.33	87.50	2.78

		85% Coverage							80% Coverage						
h	Model	Cov.	Length	Unc.	Hits	NFCI	Real	TL	Cov.	Length	Unc.	Hits	NFCI	Real	TL
	Historical	83.24	4.478	45.83	29.17	41.67	37.50	0.2653	76.96	4.228	58.33	29.17	45.83	33.33	0.3051
1	QR-NFCI	81.98	4.461	66.67	50.00	62.50	62.50	2.39	77.02	4.247	66.67	50.00	66.67	66.67	1.47
	GARCH	80.49	4.398	66.67	50.00	66.67	66.67	6.87	73.61	4.185	50.00	37.50	50.00	58.33	5.21
	Historical	83.08	4.477	66.67	41.67	70.83	58.33	0.2654	77.00	4.231	66.67	33.33	54.17	50.00	0.3054
2	QR-NFCI	82.86	4.487	79.17	54.17	66.67	66.67	1.26	77.19	4.280	79.17	50.00	62.50	62.50	0.44
	GARCH	81.11	4.452	66.67	54.17	75.00	79.17	4.92	74.24	4.210	62.50	54.17	75.00	58.33	3.69
	Historical	83.24	4.482	70.83	66.67	66.67	66.67	0.2655	77.18	4.236	66.67	54.17	58.33	54.17	0.3056
3	QR-NFCI	82.31	4.547	83.33	62.50	62.50	70.83	-1.96	76.54	4.322	83.33	58.33	62.50	75.00	-1.92
	GARCH	81.22	4.460	75.00	75.00	75.00	83.33	2.56	74.58	4.223	66.67	45.83	70.83	62.50	1.38
	Historical	83.14	4.488	62.50	66.67	66.67	62.50	0.2668	77.07	4.241	66.67	54.17	62.50	58.33	0.3066
4	QR-NFCI	82.69	4.566	79.17	75.00	66.67	66.67	-5.11	78.04	4.347	87.50	83.33	66.67	66.67	-3.54
	GARCH	81.59	4.475	79.17	70.83	87.50	79.17	1.92	74.94	4.239	66.67	54.17	70.83	50.00	0.57

This table reports the summary marginal GaR forecasting results for alternative coverage choices. Each panel reports the results for the respective coverage level. For each forecast horizon and forecasting method, we report the average empirical coverage, the average length, the percentage of the series that pass GaR adequacy tests at the 5% significance level (DQ Unc., Hits, NFCI and Real) and the percentage improvement in the average tick loss relative to the historical benchmark.

Taula A.5: Alternative Coverage Levels, Joint

		99% Coverage						90% Coverage					
h	Model	Cov.	Length	Unc.	Hits	NFCI	Real	Cov.	Length	Unc.	Hits	NFCI	Real
	Historical	98.48	10.538	0.552	0.000	0.931	0.774	80.30	6.797	0.000	0.000	0.010	0.029
1	QR-NFCI	51.52	6.111	0.000	0.000	0.000	0.000	51.52	6.111	0.000	0.000	0.000	0.000
	GARCH+Bonf. + BJPR	99.24	9.968	0.780	0.999	0.995	0.990	90.15	6.879	0.954	0.151	0.658	0.673
	Historical	98.47	10.550	0.681	0.998	0.984	0.966	78.63	6.748	0.020	0.000	0.310	0.026
2	QR-NFCI	54.20	6.595	0.000	0.999	0.000	0.000	54.20	6.595	0.000	0.000	0.000	0.000
	GARCH+Bonf. + BJPR	97.71	10.705	0.430	0.999	0.965	0.951	90.84	7.093	0.795	0.999	0.549	0.318
	Historical	97.69	10.541	0.290	0.924	0.999	0.189	77.69	6.665	0.029	0.000	0.031	0.999
3	QR-NFCI	51.54	6.570	0.000	0.000	0.000	0.000	51.54	6.570	0.000	0.000	0.000	0.999
	GARCH+Bonf. + BJPR	97.69	10.449	0.290	0.924	0.999	0.189	93.08	7.324	0.380	0.999	0.399	0.681
	Historical	97.67	10.580	0.298	0.997	0.983	0.865	77.52	6.614	0.022	0.000	0.005	0.000
4	QR-NFCI	48.06	6.587	0.000	0.000	0.000	0.000	48.06	6.587	0.000	0.000	0.000	0.000
	GARCH+Bonf. + BJPR	97.67	10.756	0.286	0.915	0.999	0.293	94.57	7.295	0.163	0.588	0.311	0.606

		85% Coverage						80% Coverage					
h	Model	Cov.	Length	Unc.	Hits	NFCI	Real	Cov.	Length	Unc.	Hits	NFCI	Real
	Historical	72.73	6.339	0.000	0.000	0.000	0.008	59.85	5.759	0.000	0.000	0.000	0.000
1	QR-NFCI	51.52	6.111	0.000	0.000	0.000	0.000	50.76	6.068	0.000	0.000	0.000	0.000
	GARCH+Bonf. + BJPR	85.61	6.397	0.845	0.641	0.997	0.673	75.76	6.108	0.223	0.414	0.603	0.559
	Historical	73.28	6.274	0.041	0.004	0.109	0.336	59.54	5.696	0.003	0.000	0.006	0.427
2	QR-NFCI	54.20	6.595	0.000	0.000	0.000	0.000	53.44	6.548	0.000	0.000	0.000	0.999
	GARCH+Bonf. + BJPR	84.73	6.645	0.945	0.000	0.720	0.541	80.92	6.309	0.832	0.007	0.832	0.578
	Historical	72.31	6.234	0.039	0.001	0.113	0.301	58.46	5.658	0.003	0.000	0.009	0.004
3	QR-NFCI	51.54	6.570	0.000	0.000	0.000	0.000	51.54	6.519	0.000	0.000	0.000	0.000
	GARCH+Bonf. + BJPR	87.69	6.821	0.501	0.133	0.213	0.719	79.23	6.399	0.867	0.461	0.937	0.706
	Historical	71.32	6.165	0.034	0.010	0.033	0.000	55.04	5.598	0.001	0.000	0.011	0.001
4	QR-NFCI	48.06	6.587	0.000	0.000	0.000	0.000	47.29	6.523	0.000	0.000	0.000	0.003
	GARCH+Bonf. + BJPR	91.47	6.834	0.095	0.486	0.323	0.432	79.84	6.422	0.972	0.152	0.160	0.463

This table reports the summary joint GaR forecasting results for alternative coverage choices. Each panel reports the results for the respective coverage level. For each forecast horizon and forecasting method, we report the average empirical coverage, the average length, and the p-values of the GaR adequacy tests considered (DQ Unc., Hits, NFCI and Real).

Taula A.6: Robustness Check: Alternative GARCH Specifications

Panel A: Marginal GaR

h	Model	Cov.	Length	Unc.	Hits	NFCI	Real	TL
1	Constant Mean	93.69	5.141	83.33	70.83	83.33	87.50	0.1214
	AR(4)	93.37	5.096	70.83	62.50	75.00	79.17	0.1281
	AR(1)+Cons. Var.	94.19	5.357	66.67	45.83	58.33	70.83	0.1455
	Skew-t	93.06	5.071	70.83	54.17	83.33	79.17	0.1222
	Direct	93.34	5.115	75.00	66.67	87.50	87.50	0.1231
2	Constant Mean	94.12	5.248	87.50	83.33	87.50	87.50	0.1319
	AR(4)	93.83	5.212	79.17	83.33	70.83	79.17	0.1304
	AR(1)+Cons. Var.	93.89	5.343	70.83	70.83	91.67	75.00	0.1420
	Skew-t	93.07	5.123	75.00	91.67	79.17	75.00	0.1298
	Direct	93.19	5.156	87.50	91.67	83.33	83.33	0.1298
3	Constant Mean	94.01	5.317	91.67	70.83	91.67	95.83	0.1366
	AR(4)	94.13	5.321	91.67	87.50	95.83	83.33	0.1409
	AR(1)+Cons. Var.	93.11	5.334	70.83	87.50	87.50	83.33	0.1468
	Skew-t	92.66	5.132	87.50	91.67	79.17	83.33	0.1347
	Direct	93.04	5.180	95.83	79.17	83.33	91.67	0.1352
4	Constant Mean	94.12	5.334	91.67	87.50	91.67	95.83	0.1382
	AR(4)	94.22	5.360	91.67	100.00	87.50	83.33	0.1392
	AR(1)+Cons. Var.	92.96	5.336	70.83	79.17	75.00	83.33	0.1483
	Skew-t	92.51	5.139	79.17	100.00	79.17	87.50	0.1375
	Direct	92.47	5.148	79.17	87.50	87.50	79.17	0.1393

Panel B: Joint GaR

h	Model	Cov.	Length	Unc.	Hits	NFCI	Real
1	Constant Mean	96.97	7.863	0.299	0.330	0.558	0.786
	AR(4)	93.94	7.536	0.576	0.745	0.637	0.470
	AR(1) + Const. Var.	87.12	7.912	0.000	0.005	0.000	0.001
	Skew-t	93.18	7.199	0.338	0.002	0.343	0.661
	Direct	94.70	7.747	0.873	0.021	0.619	0.455
2	Constant Mean	95.42	8.442	0.862	0.000	0.592	0.990
	AR(4)	94.66	8.313	0.885	0.000	0.592	0.990
	AR(1) + Const. Var.	87.79	7.894	0.003	0.999	0.999	0.015
	Skew-t	94.66	7.633	0.899	0.999	0.597	0.999
	Direct	94.66	7.750	0.899	0.999	0.597	0.999
3	Constant Mean	93.85	8.351	0.609	0.475	0.187	0.827
	AR(4)	93.85	8.464	0.614	0.225	0.388	0.802
	AR(1) + Const. Var.	83.08	7.875	0.000	0.015	0.000	0.011
	Skew-t	94.62	7.916	0.887	0.339	0.533	0.932
	Direct	94.62	7.950	0.887	0.339	0.533	0.932
4	Constant Mean	95.35	8.297	0.886	0.316	0.631	0.900
	AR(4)	94.57	8.338	0.859	0.301	0.631	0.900
	AR(1) + Const. Var.	83.72	7.868	0.000	0.001	0.000	0.000
	Skew-t	96.12	8.113	0.616	0.983	0.645	0.934
	Direct	94.57	7.707	0.876	0.500	0.588	0.857

This table reports the forecasting exercise performed with alternative models for the conditional mean, variance and distribution innovation, and the forecasting exercise performed with direct forecasts. Panel A of this table reports the results for the marginal GaR, and Panel B reports the results for the joint GaR. For each forecast horizon and forecasting method, we report the average empirical coverage, the average length, the percentage of the series that pass GaR adequacy tests at the 5% significance level (DQ Unc., Hits, NFCI and Real) for the marginal GaR and the p-values of each test for the joint GaR. Additionally, for the marginal GaR, we report the average tick loss of each model considered.

Taula A.7: Robustness Check: Panel Quantile Regressions

h	Method	Model	Cov.	Length	Unc.	Hits	NFCI	Real	TL
1	Benchmark	Historical	94.44	5.422	70.83	41.67	58.33	62.50	0.1398
	QR	QR-NFCI	94.26	5.315	75.00	41.67	70.83	75.00	1.47
		QR-NFCI+TS	93.18	5.227	62.50	37.50	70.83	50.00	1.88
		QR-NFCI+TS+GF	91.45	5.373	66.67	50.00	58.33	58.33	-26.14
		Full	93.72	5.306	66.67	45.83	70.83	58.33	4.04
	GARCH	GARCH	93.34	5.115	75.00	66.67	87.50	87.50	**11.97**
2	Benchmark	Historical	94.47	5.427	75.00	87.50	91.67	75.00	0.1410
	QR	QR-NFCI	94.34	5.419	70.83	79.17	91.67	87.50	0.40
		QR-NFCI+TS	92.81	5.220	75.00	70.83	83.33	79.17	-1.02
		QR-NFCI+TS+GF	91.06	5.491	58.33	62.50	58.33	75.00	-24.42
		Full	92.49	5.221	75.00	79.17	83.33	75.00	0.44
	GARCH	GARCH	94.15	5.245	87.50	91.67	87.50	91.67	**7.80**
3	Benchmark	Historical	94.36	5.433	75.00	87.50	95.83	87.50	0.1420
	QR	QR-NFCI	93.78	5.424	79.17	79.17	87.50	83.33	-4.23
		QR-NFCI+TS	93.04	5.253	79.17	70.83	79.17	79.17	-2.65
		QR-NFCI+TS+GF	92.63	5.555	79.17	70.83	70.83	75.00	-19.38
		Full	92.18	5.281	62.50	87.50	79.17	83.33	-2.17
	GARCH	GARCH	93.85	5.316	91.67	87.50	95.83	87.50	**3.66**
4	Benchmark	Historical	94.32	5.440	75.00	91.67	87.50	83.33	0.1427
	QR	QR-NFCI	94.32	5.482	79.17	83.33	79.17	83.33	-3.00
		QR-NFCI+TS	92.64	5.317	66.67	66.67	79.17	87.50	-3.08
		QR-NFCI+TS+GF	93.35	5.525	79.17	91.67	66.67	75.00	-18.98
		Full	91.67	5.318	70.83	66.67	79.17	83.33	-8.80
	GARCH	GARCH	93.86	5.341	91.67	91.67	87.50	91.67	**2.85**

This table reports the results for the panel QR specifications, along with the GARCH results to ease comparison. For each forecast horizon and forecasting method, this table reports the average empirical coverage, the average length, the percentage of the series that pass GaR adequacy tests at the 5% significance level (DQ Unc, Hits, NFCI and Real) and the percentage improvement in each model's average tick loss relative to the historical benchmark. The performance of the best forecasting method in terms of the tick loss is highlighted.

Taula A.8: 95% MARGINAL GAR FORECAST EVALUATION: UNIVARIATE GARCH

h	Method	Model	Cov.	Length	Unc.	Hits	NFCI	Real	TL
1	Benchmark	Historical	94.44	5.422	70.83	41.67	58.33	62.50	0.1398
	QR	NFCI	92.77	5.170	66.67	41.67	79.17	62.50	3.88
		NFCI+TS	91.13	5.079	54.17	45.83	50.00	54.17	-0.09
		NFCI+TS+GF	90.72	5.086	58.33	50.00	45.83	58.33	-1.19
		Full	89.39	5.147	37.50	29.17	29.17	33.33	-19.14
		LASSO	90.25	5.151	50.00	29.17	50.00	45.83	-6.86
	GARCH	GARCH	93.02	5.086	83.33	58.33	79.17	87.50	9.53
		GARCH-NFCI	93.94	5.164	95.83	70.83	87.50	95.83	13.48
		GJR-GARCH	92.96	5.122	83.33	66.67	75.00	79.17	7.42
		F-GARCH	93.09	5.119	70.83	50.00	79.17	79.17	**13.99**
2	Benchmark	Historical	94.47	5.427	75.00	87.50	91.67	75.00	0.1410
	QR	NFCI	92.75	5.257	75.00	75.00	79.17	70.83	0.47
		NFCI+TS	90.62	5.166	62.50	87.50	79.17	75.00	-3.99
		NFCI+TS+GF	91.13	5.187	66.67	83.33	83.33	75.00	-4.22
		Full	89.54	5.172	54.17	79.17	54.17	70.83	-36.40
		LASSO	90.55	5.157	54.17	66.67	66.67	62.50	-8.46
	GARCH	GARCH	93.64	5.255	87.50	79.17	87.50	95.83	4.08
		GARCH-NFCI	94.47	5.362	91.67	79.17	91.67	87.50	**8.10**
		GJR-GARCH	93.48	5.267	83.33	95.83	87.50	95.83	3.62
		F-GARCH	96.63	5.751	91.67	75.00	95.83	95.83	2.41
3	Benchmark	Historical	94.36	5.433	75.00	87.50	95.83	87.50	0.1420
	QR	NFCI	92.69	5.314	66.67	83.33	75.00	75.00	-3.74
		NFCI+TS	90.42	5.195	62.50	62.50	66.67	70.83	-9.40
		NFCI+TS+GF	90.80	5.238	66.67	62.50	75.00	75.00	-9.19
		Full	89.04	5.256	50.00	58.33	54.17	58.33	-31.19
		LASSO	89.46	5.231	54.17	66.67	54.17	58.33	-14.19
	GARCH	GARCH	93.56	5.318	87.50	87.50	91.67	83.33	0.37
		GARCH-NFCI	95.06	5.479	95.83	87.50	100.00	95.83	**2.95**
		GJR-GARCH	93.72	5.327	87.50	87.50	87.50	91.67	-0.05
		F-GARCH	97.34	6.155	91.67	91.67	100.00	95.83	-6.73
4	Benchmark	Historical	94.32	5.440	75.00	91.67	87.50	83.33	0.1427
	QR	NFCI	92.09	5.324	75.00	79.17	79.17	87.50	-11.09
		NFCI+TS	90.31	5.214	66.67	83.33	66.67	87.50	-12.44
		NFCI+TS+GF	89.79	5.174	66.67	79.17	70.83	75.00	-15.00
		Full	88.79	5.187	54.17	70.83	62.50	62.50	-48.39
		LASSO	88.82	5.194	58.33	70.83	62.50	70.83	-19.63
	GARCH	GARCH	93.73	5.364	95.83	87.50	95.83	95.83	**0.51**
		GARCH-NFCI	95.03	5.541	87.50	91.67	87.50	100.00	0.43
		GJR-GARCH	93.83	5.379	87.50	83.33	95.83	100.00	-0.21
		F-GARCH	97.74	6.414	75.00	87.50	100.00	95.83	-12.36

This table reports the following for each forecast horizon and forecasting method: the average empirical coverage; the average length; the percentage of the series that pass GaR adequacy tests at the 5% significance level (DQ Unc, Hits, NFCI and Real); and the percentage improvement in each model's average tick loss relative to the historical benchmark. The performance of the best forecasting method in terms of the tick loss is highlighted in boldface.

Taula A.9: 95% JOINT GAR FORECAST EVALUATION: UNIVARIATE GARCH

h	Method	Model	Cov.	Length	Unc.	Hits	NFCI	Real
1	Benchmark	Historical	90.91	7.943	0.031	0.003	0.024	0.121
	QR + Bonf.	QR-NFCI	51.52	6.111	0.000	0.000	0.000	0.000
		NFCI+TS	34.85	5.834	0.000	0.000	0.000	0.000
		NFCI+GF+TS	34.09	5.673	0.000	0.000	0.000	0.000
		Full	25.00	5.563	0.000	0.000	0.000	0.000
		LASSO	38.64	7.111	0.000	0.000	0.000	0.000
	GARCH + Marg.	GARCH	32.58	5.086	0.000	0.000	0.000	0.000
		GARCH-NFCI	35.61	5.164	0.000	0.000	0.000	0.000
		GJR-GARCH	31.06	5.122	0.000	0.000	0.000	0.000
		F-GARCH	31.82	5.119	0.000	0.000	0.000	0.000
	GARCH + Bonf.	GARCH	84.09	7.149	0.000	0.000	0.000	0.000
		GARCH-NFCI	81.82	6.719	0.000	0.000	0.000	0.000
		GJR-GARCH	82.58	7.123	0.000	0.000	0.000	0.000
		F-GARCH	85.61	7.168	0.000	0.000	0.000	0.000
	GARCH + BJPR	GARCH	91.67	7.656	0.079	0.001	0.004	0.057
		GARCH-NFCI	93.18	7.250	0.338	0.002	0.016	0.597
		GJR-GARCH	91.67	7.650	0.079	0.001	0.004	0.057
		F-GARCH	93.94	7.484	0.576	0.141	0.093	0.332
2	Benchmark	Historical	89.31	7.901	0.024	0.999	0.036	0.126
	QR + Bonf.	QR-NFCI	54.20	6.595	0.000	0.999	0.031	0.000
		NFCI+TS	39.69	6.241	0.000	0.000	0.000	0.000
		NFCI+GF+TS	40.46	6.157	0.000	0.000	0.000	0.000
		Full	31.30	5.864	0.000	0.000	0.000	0.000
		LASSO	41.98	7.560	0.000	0.000	0.000	0.000
	GARCH + Marg.	GARCH	40.46	5.255	0.000	0.000	0.000	0.000
		GARCH-NFCI	43.51	5.362	0.000	0.000	0.000	0.999
		GJR-GARCH	38.17	5.267	0.000	0.000	0.000	0.000
		F-GARCH	64.89	5.751	0.000	0.999	0.999	0.000
	GARCH + Bonf.	GARCH	88.55	7.975	0.061	0.999	0.999	0.000
		GARCH-NFCI	83.21	7.206	0.000	0.999	0.999	0.000
		GJR-GARCH	86.26	7.975	0.007	0.999	0.999	0.973
		F-GARCH	93.13	8.759	0.477	0.257	0.097	0.495
	GARCH + BJPR	GARCH	93.89	8.373	0.718	0.999	0.601	0.896
		GARCH-NFCI	93.89	8.419	0.718	0.999	0.601	0.896
		GJR-GARCH	93.89	8.357	0.718	0.999	0.601	0.896
		F-GARCH	97.71	9.287	0.195	0.902	0.715	0.809
3	Benchmark	Historical	88.46	7.797	0.028	0.999	0.016	0.228
	QR + Bonf.	QR-NFCI	51.54	6.570	0.000	0.999	0.000	0.000
		NFCI+TS	33.85	6.262	0.000	0.000	0.000	0.000
		NFCI+GF+TS	32.31	6.122	0.000	0.000	0.000	0.000
		Full	30.77	5.930	0.000	0.999	0.000	0.000
		LASSO	39.23	7.953	0.000	0.000	0.000	0.000
	GARCH + Marg.	GARCH	42.31	5.318	0.000	0.000	0.999	0.999
		GARCH-NFCI	50.00	5.479	0.000	0.999	0.999	0.999
		GJR-GARCH	40.77	5.327	0.000	0.000	0.999	0.999
		F-GARCH	71.54	6.155	0.000	0.000	0.999	0.999
	GARCH + Bonf.	GARCH	86.15	7.955	0.018	0.000	0.000	0.154
		GARCH-NFCI	83.08	7.329	0.000	0.938	0.999	0.243
		GJR-GARCH	83.85	7.989	0.002	0.999	0.999	0.000
		F-GARCH	93.85	9.699	0.626	0.420	0.162	0.684
	GARCH + BJPR	GARCH	93.85	8.409	0.708	0.999	0.560	0.971
		GARCH-NFCI	93.85	8.302	0.708	0.999	0.560	0.971
		GJR-GARCH	93.08	8.417	0.522	0.999	0.515	0.817
		F-GARCH	97.69	10.301	0.200	0.907	0.596	0.811
4	Benchmark	Historical	84.50	7.732	0.031	0.999	0.044	0.170
	QR + Bonf.	QR-NFCI	48.06	6.587	0.000	0.000	0.000	0.000
		NFCI+TS	34.11	6.335	0.000	0.000	0.000	0.000
		NFCI+GF+TS	33.33	6.167	0.000	0.000	0.000	0.000
		Full	23.26	5.878	0.000	0.999	0.000	0.000
		LASSO	35.66	7.834	0.000	0.000	0.999	0.000
	GARCH + Marg.	GARCH	44.96	5.364	0.000	0.999	0.000	0.000
		GARCH-NFCI	52.71	5.541	0.000	0.000	0.000	0.000
		GJR-GARCH	41.09	5.379	0.000	0.000	0.000	0.000
		F-GARCH	76.74	6.414	0.000	0.999	0.999	0.000
	GARCH + Bonf.	GARCH	86.82	7.985	0.017	0.004	0.999	0.715
		GARCH-NFCI	83.72	7.447	0.001	0.999	0.999	0.090
		GJR-GARCH	84.50	8.031	0.003	0.000	0.999	0.223
		F-GARCH	94.57	10.242	0.849	0.921	0.714	0.721
	GARCH + BJPR	GARCH	93.80	8.346	0.723	0.082	0.512	0.818
		GARCH-NFCI	93.02	8.408	0.512	0.526	0.705	0.620
		GJR-GARCH	92.25	8.367	0.420	0.999	0.456	0.743
		F-GARCH	99.22	10.911	0.027	0.477	0.447	0.406

This table reports the following for each forecast horizon and forecasting method: the average empirical joint coverage; the average length; and the p-values of the GaR adequacy tests considered (DQ Unc., Hits, NFCI and Real).

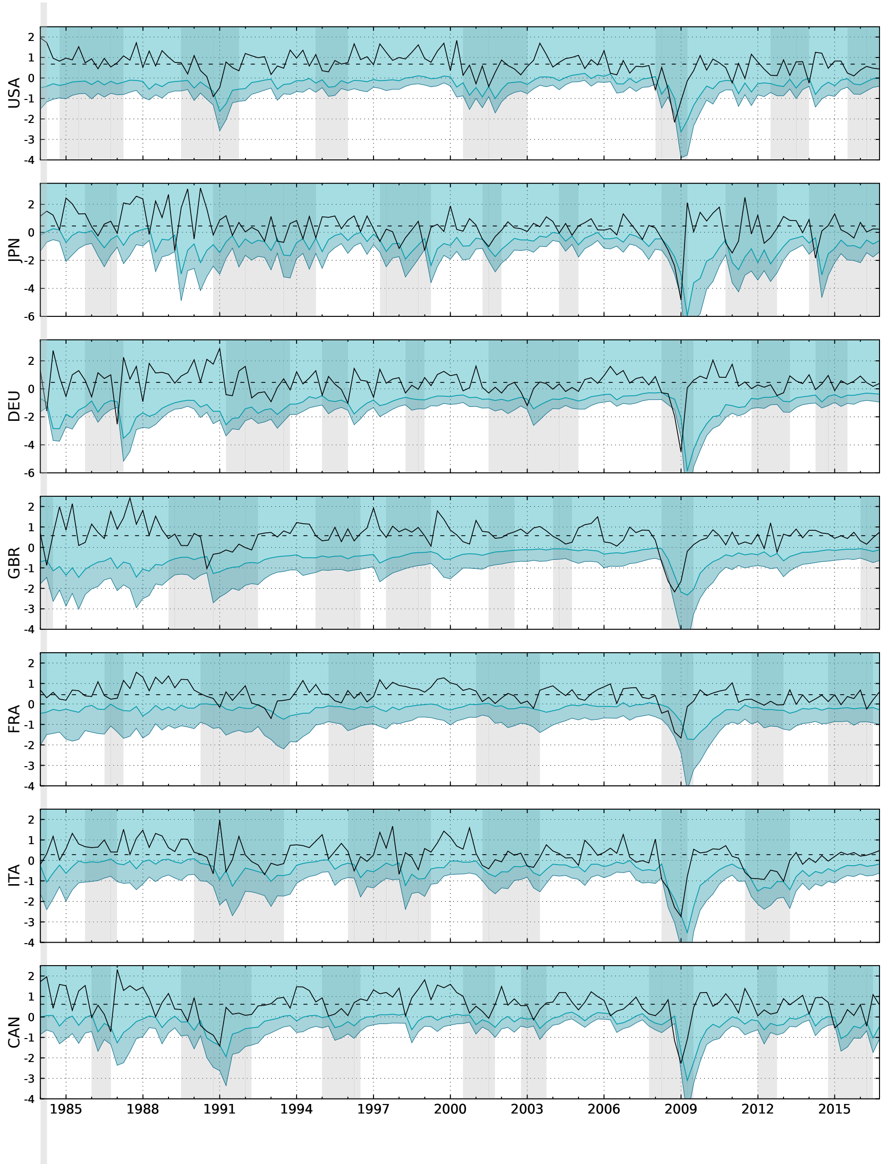
Taula A.10: Robustness Check: $(1-q)\%$/95% Joint GaR

h	Region	Model	Cov.	Length	Unc.	Hits	NFCI	Real
1	95% Joint GaR	Historical	90.15	7.952	0.011	0.005	0.004	0.034
		GARCH	93.94	7.749	0.576	0.000	0.446	0.535
	95% / 95% Joint GaR	Historical	93.18	6.106	0.338	0.005	0.000	0.014
		GARCH	90.91	5.996	0.031	0.000	0.001	0.162
	90% / 95% Joint GaR	Historical	92.42	5.741	0.175	0.000	0.000	0.019
		GARCH	93.94	5.642	0.576	0.000	0.005	0.275
2	95% Joint GaR	Historical	90.08	7.933	0.042	0.999	0.037	0.133
		GARCH	95.42	8.503	0.862	0.000	0.592	0.990
	95% / 95% Joint GaR	Historical	92.37	6.106	0.343	0.999	0.000	0.000
		GARCH	95.42	6.308	0.899	0.965	0.717	0.001
	90% / 95% Joint GaR	Historical	93.13	5.741	0.487	0.000	0.999	0.012
		GARCH	92.37	5.747	0.406	0.999	0.138	0.002
3	95% Joint GaR	Historical	90.00	7.930	0.054	0.999	0.012	0.999
		GARCH	95.38	8.426	0.874	0.033	0.508	0.917
	95% / 95% Joint GaR	Historical	93.08	6.106	0.424	0.756	0.079	0.000
		GARCH	94.62	6.479	0.907	0.524	0.515	0.999
	90% / 95% Joint GaR	Historical	93.08	5.742	0.477	0.999	0.286	0.030
		GARCH	93.08	5.813	0.608	0.073	0.698	0.999
4	95% Joint GaR	Historical	88.37	7.928	0.019	0.999	0.105	0.049
		GARCH	95.35	8.377	0.886	0.316	0.631	0.900
	95% / 95% Joint GaR	Historical	92.25	6.101	0.327	0.116	0.363	0.185
		GARCH	95.35	6.575	0.903	0.987	0.710	0.836
	90% / 95% Joint GaR	Historical	92.25	5.741	0.347	0.326	0.454	0.163
		GARCH	92.25	5.926	0.455	0.134	0.762	0.415

This table reports the summary backtesting results for the $(1-q)\%$/95% joint GaR for $q = 5\%$ and $q = 10\%$, alongside the joint GaR. For each forecast horizon and forecasting method, we report the average empirical coverage, the average length, and the p-values of the GaR adequacy tests (DQ Unc., Hits, NFCI and Real).

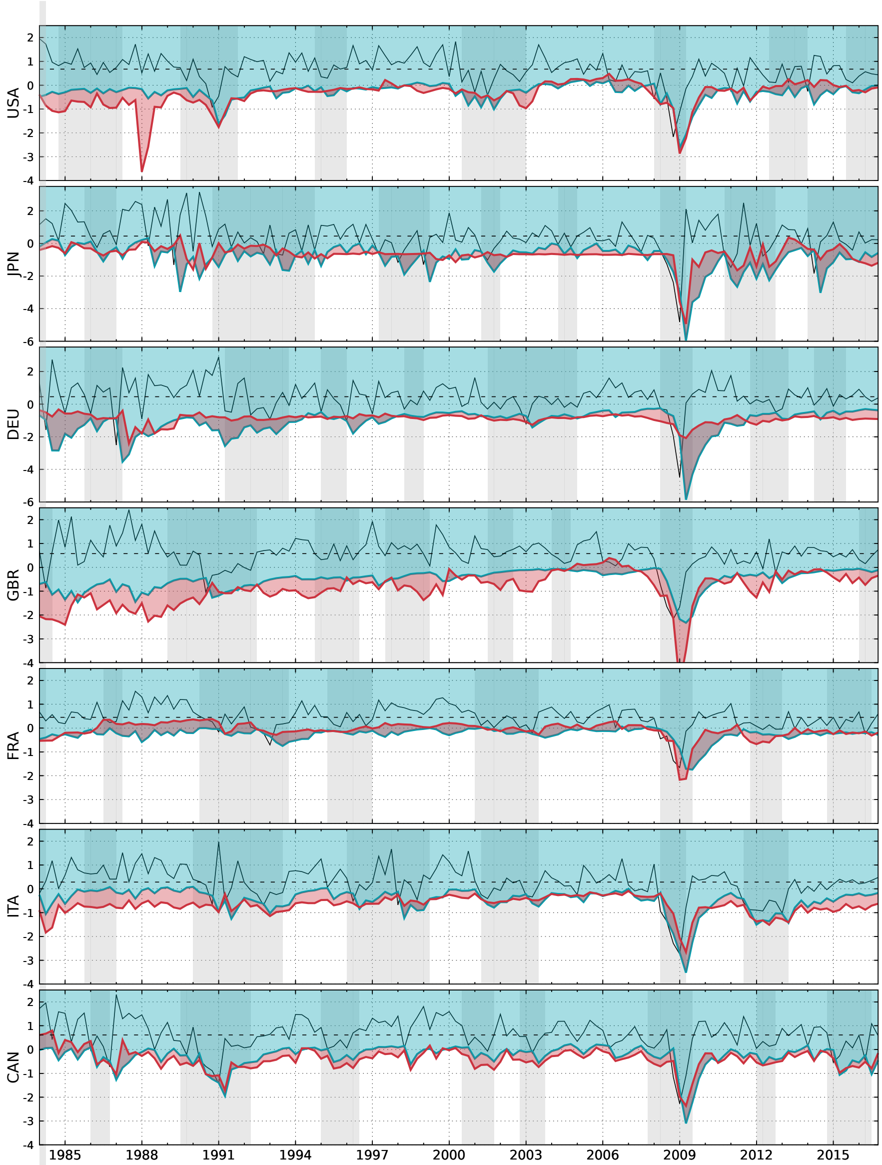
A.3.2 Additional Tables and Figures

Figura A.1: 95% MARGINAL GAR AND 95%/95% JOINT GAR FOR G7 COUNTRIES



This figure displays the time series plots of the sequence of the 1-step-ahead 95% joint and marginal GaR predictions obtained from the GARCH(1,1) model for the G7 countries from 1983Q4 to 2016Q4. The lightest shaded region is the marginal GaR, and the darkest shaded region represents the 95%/95% joint GaR. The gray regions are the OECD recession dates. We also plot the GDP growth rate of each country (black line) and the average GDP growth rate over the sample period (dashed black line).

Figura A.2: 95% MARGINAL GAR FOR G7 COUNTRIES: QR AND GARCH



This figure displays the times series plots of the sequence of the 1-step-ahead 95% joint and marginal GaR predictions obtained from the QR by using the NFCI for the G7 countries from 1983Q4 to 2016Q4. The blue shaded region is the marginal GaR constructed with GARCH(1,1), and the red shaded region is the GaR constructed with QR-NFCI. The gray regions are the OECD recession dates. We also plot the GDP growth rate of each country (black line) and the average GDP growth rate over the sample period (dashed black line).

Taula A.11: Out-of-sample QR Screening

	Date Availability		h: 1	2	3	4
Historical	1961Q2 2016Q4	$\Delta\%TL$ $\%DM$	-4.036 0.000	-0.476 12.500	3.603 20.833	9.987 20.833
CS	1986Q2 2020Q1	$\Delta\%TL$ $\%DM$	-5.143 8.333	-10.930 0.000	-14.764 0.000	0.185 4.167
HP	1970Q3 2019Q3	$\Delta\%TL$ $\%DM$	-5.896 0.000	-3.023 8.333	0.460 0.000	7.647 8.333
EPU	1985Q1 2019Q3	$\Delta\%TL$ $\%DM$	-6.042 8.333	-12.085 4.167	-14.906 8.333	-1.381 4.167
AR(1)	1961Q2 2016Q4	$\Delta\%TL$ $\%DM$	-7.245 0.000	-3.413 12.500	2.330 16.667	7.474 16.667
CG	1962Q3 2019Q1	$\Delta\%TL$ $\%DM$	-8.189 0.000	-4.439 4.167	0.262 8.333	6.051 16.667
TS	1964Q3 2019Q2	$\Delta\%TL$ $\%DM$	-8.548 4.167	-4.531 16.667	-3.270 12.500	4.328 25.000
SV	1950Q1 2019Q3	$\Delta\%TL$ $\%DM$	-8.947 0.000	-10.844 0.000	-7.268 0.000	-5.323 4.167
GF	1961Q2 2016Q4	$\Delta\%TL$ $\%DM$	-9.108 0.000	-3.187 4.167	0.795 12.500	8.281 16.667
CR	1952Q3 2019Q1	$\Delta\%TL$ $\%DM$	-10.927 8.333	-4.755 0.000	-1.281 12.500	5.178 16.667
GPR	1985Q1 2016Q4	$\Delta\%TL$ $\%DM$	-12.800 4.167	-9.635 4.167	-8.735 12.500	-0.728 8.333
WUI	1996Q1 2019Q3	$\Delta\%TL$ $\%DM$	-19.966 8.333	-20.144 0.000	-29.381 0.000	-26.952 0.000

This table reports the summary out-of-sample results for the bivariate QR models. For each candidate predictor and forecast horizon, we report the dates for which the predictor is available, the percentage improvement in the average tick loss relative to the baseline QR-NFCI and the percentage of countries for which the predictor outperforms the QR-NFCI, according to a DM test conducted at the 5% significance level.

Bibliografia

Adrian, T., Boyarchenko, N., and Giannone, D. (2019). Vulnerable growth. *American Economic Review*, 109(4):1263–1289.

Adrian, T. and Brunnermeier, M. K. (2016). Covar. *American Economic Review*, 106(7):1705–1741.

Adrian, T., Grinberg, F., Liang, N., and Malik, S. (2018). The term structure of growth-at-risk. WP.

Ahir, H., Bloom, N., and Furceri, D. (2018). The world uncertainty index. WP.

Allen, L., Bali, T. G., and Tang, Y. (2012). Does Systemic Risk in the Financial Sector Predict Future Economic Downturns? *Review of Financial Studies*, 25:3000–3036.

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

Anatolyev, S. and Gerko, A. (2005). A trading approach to testing for predictability. *Journal of Business & Economic Statistics*, 23(4):455–461.

Anatolyev, S. and Gospodinov, N. (2010). Modeling financial return dynamics via decomposition. *Journal of Business & Economic Statistics*, 28(2):232–245.

Anatolyev, S. and Gospodinov, N. (2019). Multivariate return decomposition: Theory and implications. *Econometric Reviews*, 38(5):487–508.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). Volatility and correlation forecasting. Handbook of Economic Forecasting.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):367–389.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146:304–317.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal Of Economics*, 131(4):1593–1636.

Barone-Adesi, G., Engle, R. F., and Mancini, L. (2008). A GARCH option pricing model with filtered historical simulation. *Review of Financial Studies*, 21(3):1223–1258.

Beutel, J. (2019). Forecasting growth at risk. Working Paper.

Boot, T. and Nibbering, D. (2019). Forecasting using random subspace methods. *Journal of Econometrics*, 209(2):391 – 406.

Borio, C. and Lowe, P. (2002). Asset prices, financial and monetary stability: Exploring the nexus. BIS working paper 114.

Breiman, L. (2001). Random forests. *Machine Learning*, 45:5–32.

Brockwell, P. J. and Davis, R. A. (1991). *Time series: Theorey and Methods*. Springer.

Brownlees, C. and Engle, R. F. (2017). Srisk: A conditional capital shortfall measure of systemic risk. *Review of Financial Studies*, 30(1):48–79.

Brownlees, C. T., Engle, R. F., and Kelly, B. (2011). A practical guide to volatility forecasting through calm and storm. *Journal of Risk*, 14(2):3–22.

Brownlees, C. T. and Souza, A. B. (2020). Backtesting global growth-at-risk. *Forthcoming Journal of Monetary Economics*.

Caldara, D. and Iacoviello, M. (2018). Measuring geopolitical risk. International Finance Discussion Papers 1222.

Caldera Sanchéz, A. and Röhn (2016). How do policies influence gdp tail risks? WP.

Carriero, A., Clark, T. E., and Marcellino, M. G. (2020). Capturing macroeconomic tail risks with bayesian vector autoregressions. FRB of Cleveland Working Paper no.20-02R.

Chavleishvili, S. and Manganelli, S. (2019). Forecasting and stress testing with quantile vector autoregression. WP.

Cheng, X. and Hansen, B. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics*, 186(2):280–293.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4):841–862.

Claessens, S., Kose, A. M., and Terrones, M. E. (2008). What happens during recessions, crunches and busts? IMF Working Paper.

Clements, M. P., Franses, P. H., and Swanson, N. R. (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, 20:169–183.

De Nicolò, G. (2019). An early warning system for tail risks. Working Paper.

Devroye, L., Laszlo, G., and Lugosi, G. (1996). *A probabilistic theory of pattern recognition*. Springer.

Diebold, F. X. and Christoffersen, P. F. (2006). Financial asset returns, direction-of-change forecasting and volatility dynamics. *Management Science*, 52:1273–1287.

Diebold, F. X., Christoffersen, P. F., Mariano, R. S., Tay, A. S., and Tse, Y. K. (2007). Direction-of-change forecasts based on conditional variance, skewness and kurtosis dynamics: international evidence. *Journal of Financial Forecasting*, 1(2):1–22.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economics Statistics*, 13:253–263.

Diebold, F. X. and Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives. *International Journal of Forecasting*, 35:1679–1691.

ECB (2019). Financial stability review, may 2019. Financial Stability Review. European Central Bank.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Engle, R. F. and Manganelli, S. (2004). Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economics Statistics*, 22(4):367–381.

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

Faust, J., Gilchrist, S., Wright, J. H., and Zakrajsek, E. (2013). Credit spreads as predictors of real time economic activity. *Review of Economics and Statistics*, 95(5):1501–1519.

Faust, J. and Wright, J. (2013). Forecasting inflation. *Handbook of Economic Forecasting*, 2A:3 – 56.

Freund, Y. and Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. *European Conference on computation learning theory*, pages 23–37.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.

Ghysels, E., Iania, L., and Striaukas, J. (2018). Quantile-based inflation risk models. Technical Report 349, National Bank of Belgium.

Giacomini, R. and Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business & Economics Statistics*, 23(4):416–431.

Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

IMF (2017). Global financial stability report: is growth at risk? volume 3 of *Global Financial Stability Report*, pages 91 – 118. International Monetary Fund.

Kim, H. H. and Swanson, N. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, 178(P2):352–367.

Koenker, R. W. and Basset, Jr., G. (1978). Regression quantiles. *Econometrica*, 46(1):33–50.

Leung, M. T., Daouk, H., and Chen, A.-S. (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecsating*, 16(2):173–190.

Marcellino, M. (2002). Instability and non-linearity in the emu. *Centre for Economic Policy Research*, Discussion paper No. 3312.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Ng, S. (2013). Variable selection in predictive regressions. Handbook of Economic Forecasting.

Nyberg, H. (2011). Forecasting the direction of the us stock market with dynamic binary probit models. *International Journal of Forecasting*, (27):561–578.

Pakel, C., Shephard, N., and Sheppard, K. (2011). Nuisance parameters, composite likelihoods and a panel of GARCH models. *Statistica Sinica*, 21(1):307–329.

Pesaran, H. M. and Timmerman, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economics Statistics*, 10:461–465.

Plagborg-Møller, M., Reichlin, L., Ricco, G., and Hasenzagl, T. (2020). When is growth at risk? *Working Paper*.

Prasad, A., Elekdag, S., Jeasakul, P., Lafarguette, R., Alter, A., Feng Xiaochen, A., and Wang, C. (2019). Growth at risk: Concept and application in imf country surveillance. WP.

Romano, J. P., Shaikh, A. M., and Wolf, M. (2010). Hypothesis testing in econometrics. *Annual Review of Economics*, 2(1):75–104.

Romano, J. P. and Wolf, M. (2007). Control of generalized error rates in multiple testing. *Annals of Statistics*, 35(4):1378–1408.

Rossi, B. (2013). Exchange rate predictability. *Journal of Economic Literature*, 51(4):1063–1119.

Rydberg, T. H. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: Decomposition and models. *Journal of Financial Econometrics*, 1(1):2–25.

Sinclair, T. M., Stekler, H. O., and Kitzinger, L. (2010). Directional forecasts of gdp and inflation: a joint evaluation with an application to federal reserve predictions. *Applied Economics*, 42(18):2289–2297.

Stock, J. and Watson, M. (1999). *A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series*, pages 1–44. Oxford University Press, Oxford.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2006). Forecasting with many predictors. *Handbook of Economic Forecasting*, 1:516 – 554.

Stock, J. H. and Watson, M. W. (2007). Why has u.s inflation become harder to forecast? *Journal of Money, Credt and Banking*, 39(1):4–33.

Stock, J. H. and Watson, M. W. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business & Economic Statistics*, 30(4):481–493.

Terasvirta, T. (2006). Forecasting economic variables with nonlinear models. Handbook of Economic Forecasting.

Timmerman, A. (2006). Forecast combinations. Handbook of Economic Forecasting.

White, H. (1990). Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, 3(5):169–183.

White, H. (2006). Approximate nonlinear forecasting. Handbook of Economic Forecasting.

White, H. and Swanson, N. (1997). Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models. *International Journal of Forecasting*, 13:439–461.

White, H., Tae-Hwan, K., and Manganelli, S. (2015). Var for var: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics*, 187:169–188.

Wolf, M. and Wunderli, D. (2015). Bootstrap joint prediction regions. *Journal of Time Series Analysis*, 36(1):352–376.