Journal of Monetary Economics 118 (2021) 312–330

Contents lists available at ScienceDirect

Journal of Monetary Economics

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

Backtesting global Growth-at-Risk[☆]

Christian Brownlees[a,b,*], André B.M. Souza[a,b]

[a] Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25–27, Barcelona 08005, Spain
[b] Barcelona GSE, Spain

ARTICLE INFO

Article history:
Received 28 March 2020
Revised 8 November 2020
Accepted 9 November 2020
Available online 21 November 2020

JEL classification:
C22
C23
C52
C53
C58

Keywords:
Growth-at-Risk
Backtesting
Quantile regression
GARCH

ABSTRACT

We conduct an out-of-sample backtesting exercise of Growth-at-Risk (GaR) predictions for 24 OECD countries. We consider forecasts constructed from quantile regression and GARCH models. The quantile regression forecasts are based on a set of recently proposed measures of downside risks to GDP, including the national financial conditions index. The backtesting results show 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.

1. Introduction

In recent years, the focus of policymakers on downside risk has substantially increased (Caldera Sanchéz and Röhn, 2016; Prasad et al., 2019), 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 insti-

☆ We would like to thank Simone Manganelli, Gianni de Nicolò, Katerina Petrova, Barbara Rossi, Bernd Schwaab, Michael Wolf and participants of the "Vulnerable Growth in the Euro Area" workshop (ECB, Frankfurt, 17 October 2019), the 2019 French Econometrics Conference (AMSE, Marseille, 25–26 November 2019) and the UPF internal seminars for providing numerous helpful comments. We would also like to thank associate editor Jonathan Wright and an anonymous referee for their useful comments. Christian Brownlees acknowledges support from the Spanish Ministry of Science and Technology (Grant MTM2012-37195) and the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (SEV-2011-0075). André B.M. Souza acknowledges support from the Spanish Ministry of Science and Technology (FPI Grant BES-2017-080615). The replication files for this paper can be found at github.com/ctbrownlees/gar-replication.

* Corresponding author. Tel.: (+34) 93542 2750.
E-mail address: christian.brownlees@upf.edu (C. Brownlees).

https://doi.org/10.1016/j.jmoneco.2020.11.003
0304-3932/

tutions 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 $(1-p)\%$ 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 (Ahir et al., 2018; Baker et al., 2016), a geopolitical risk index (Caldara and Iacoviello, 2018) and market risk measures (Estrella and Hardouvelis, 1991; Faust et al., 2013). 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-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 back-

[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.

[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).

[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.

testing 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). 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 relies 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 Online 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 details the methodology, and Section 3 presents the empirical evidence. Concluding remarks follow in Section 4. We provide additional methodological details and empirical results in the Online Appendix.

2. Forecasting growth-at-Risk

2.1. Marginal and joint growth-at-Risk definitions

Let Y_{it} denote the GDP growth rate of country $i = 1, \dots, n$ for period $t = 1, \dots, T$. 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

$$\mathbb{P}_t\left(\text{ at least one growth rate } Y_{i\,t+h} \text{ is not in } GaR^J_{t+h|t} \right) = p. \tag{1}$$

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 (1) may be excessively restrictive, even for moderately large panels. This definition

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

[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 $GaR^J_{t+h|t}$*".

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

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

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 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. Models for Growth-at-Risk

2.2.1. 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 $\{Y_{it}\}$ is given by

$$Q_p(Y_{it+h}|\mathcal{I}_t) = \alpha_i^p + \beta_{i0}^p Y_{it} + \beta_{i1}^p X_{i1\,t} + \ldots + \beta_{iK}^p X_{iK\,t}, \quad (2)$$

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

To make the model in (2) 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 3.1, where we introduce the data used in our analysis.

The parameters in (2) are estimated by minimizing the tick loss, which is given by

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

where $\rho_p(x) = x\big(p - \mathbb{1}_{\{x<0\}}\big)$. 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.

2.2.2. 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:

$$Y_{it+1} = \mu_{it+1|t} + \sqrt{\sigma^2_{it+1|t}}\, Z_{it+1} \quad Z_{it+1} \overset{i.i.d}{\sim} \mathcal{D}_{Z_i}(0,1), \tag{3}$$

where $\mu_{it+1|t}$ denotes the 1-step-ahead conditional mean, $\sigma^2_{it+1|t}$ 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 (3) is

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

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 $\{Y_{it}\}$ is then given by

$$Q_p(Y_{it+1}|\mathcal{I}_t) = \mu_{it+1|t} + \sqrt{\sigma^2_{it+1|t}}\, F^{-1}_{Z_i}(p), \tag{4}$$

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

$$Q_p(Y_{it+h}|\mathcal{I}_t) = F^{-1}_{\widetilde{Y}_{it+h|t}}(p), \tag{5}$$

where $\widetilde{Y}_{it+h|t}$ 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 (3); therefore, we label the quantile forecast in (5) as iterated. Section A.1 of the Online Appendix details the simulation algorithm for this computation.

To make the model in (3) operational, the conditional mean $\mu_{it+1|t}$, conditional variance $\sigma^2_{it+1|t}$ 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),

$$\sigma^2_{it+1|t} = \sigma^2_i(1 - \alpha_i - \beta_i) + \alpha_i \varepsilon^2_{it} + \beta_i \sigma^2_{it}, \tag{6}$$

where ε_{it} 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 Online 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 Online Appendix.

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 $i = 1, \ldots, n$,

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

where $Q_p(Y_{it+h}|\mathcal{I}_t)$ is set equal to (2) for quantile regression or (4) and (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 $i=1,\dots,n$ given by

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

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 $i=1,\dots,n$ given by

$$GaR^{J,BJPR}_{it+h|t} = \mu_{it+h|t} + d^1_p \sigma_{it+h|t}, \tag{7}$$

where $\mu_{it+h|t}$ is the h-step-ahead conditional mean, $\sigma_{it+h|t}$ 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 $(\widetilde{Z}_{1\,t+h|t},\dots,\widetilde{Z}_{n\,t+h|t})'$, with $\widetilde{Z}_{it+h|t} = (\widetilde{Y}_{it+h|t} - \mu_{it+h|t})/\sigma_{it+h|t}$ and $\widetilde{Y}_{it+h|t}$ defined as in Section 2.2.2. Typically, d^1_p is unknown and can be approximated by resampling. Section A.1 of the Online 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 $i=1,\dots,n$ given by

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

where $\mu_{it+h|t}$ is the h-step-ahead conditional mean, $\sigma_{it+h|t}$ 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 $(\widetilde{Z}_{1\,t+h|t},\dots,\widetilde{Z}_{n\,t+h|t})'$ and with $\widetilde{Z}_{it+h|t}$ defined as above.

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

$$\hat{C}^M = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{\{Y_{it} > GaR^M_{it|t-h}\}}\right) \text{ and } \hat{C}^J = \frac{1}{T}\sum_{t=1}^{T}\mathbb{1}_{\bigcup_{i=1}^{n}\{Y_{it} > GaR^J_{it|t-h}\}}.$$

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

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

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 $GaR^M_{it|t-h}$ and $GaR^J_{it|t-h}$ 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 $H^M_{it} = \mathbb{1}_{\{Y_{it} \le GaR^M_{it|t-h}\}} - p$, that is, a sequence of binary random variables that are 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 $H^J_t = \mathbb{1}_{\bigcup_{i=1}^{n}\{Y_{it} \le GaR^J_{it|t-h}\}} - p$. Let $W_{1\,t},\dots,W_{K\,t}$ denote a set of auxiliary predictors observed in period t, and

[8] In the empirical application, all the quantile regression models that we consider have at most 2 observations per country such that $GaR^M_{it|t-h} > \hat{Q}_{0.99}(Y_i)$ or $GaR^J_{it|t-h} > \hat{Q}_{0.99}(Y_i)$. 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.

consider the regression

$$H_{t+h} = c_0 + \sum_{k=1}^{K} c_k W_{kt} + u_{t+h}, \tag{8}$$

where H_t may denote either H_{it}^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 (8) are zero. Thus, the dynamic quantile test is based on testing the null $\mathrm{H}_0 : c_0 = \ldots = c_K = 0$ 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

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

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.

3. Empirical analysis

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 Online Appendix.

3.2. In-sample analysis

3.2.1. 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

$$Q_{0.05}(Y_{it+h}|\mathcal{I}_t) = \alpha_i^{0.05} + \beta_{i0}^{0.05} Y_{it} + \beta_{i1}^{0.05} X_{ikt}, \tag{9}$$

which is similar in spirit to the linear regression specifications used to test for Granger causality. The quantile regression in (9) based on the NFCI is used as a benchmark and is called QR-NFCI. To simplify comparisons, for each predictor X_{ikt}, the

[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 $j \geq m$.

[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).

Table 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 $Q_p(Y_{it+h}|\mathcal{I}_t) = \alpha_0^p + \beta_0^p Y_{it} + \beta_1^p X_{ikt}$, where X_{ikt} is the variable of interest. Significance is assessed on the basis of (block) bootstrap standard errors with blocks of length 4.

corresponding quantile regression given by (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 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 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.

3.2.2. 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 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 $p = .05$ (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 $p = .05$ 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 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.

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, 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 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 2.3. The regions depicted throughout this section are computed out-of-sample. A natural concern regarding joint prediction regions

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

[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 OA1 and OA2 in the Online Appendix display further marginal and joint GaR plots for the G7 countries.

Table 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.

Table 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 Online Appendix.

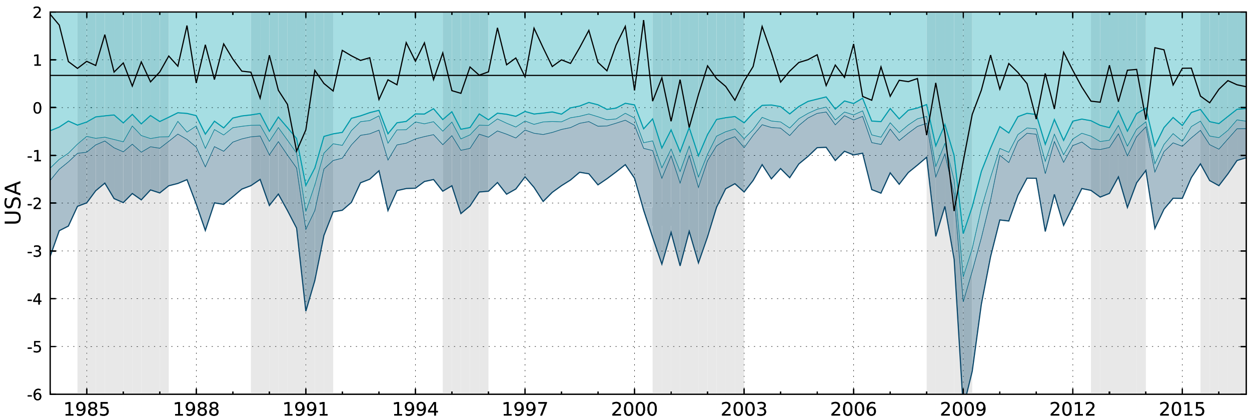

Fig. 1. GaR Prediction Regions Comparison for the USA This figure displays the 1-step-ahead 95% marginal GaR, 90%/95% joint GaR, 95%/95% joint GaR and 95% joint GaR for the USA. The lightest region represents the marginal GaR region, and the darkest region represents the 95% joint GaR for the 24 countries 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).

is their length. As shown in Figure 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.

Table 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.

We backtest marginal and joint GaR forecasts using the tools introduced in Section 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 (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 Online 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 2.3, with the conditional mean and variance forecasts replaced by their recursively estimated unconditional counterparts.

Table 5
Superior predictive ability test Pairwise comparison.

		Forward				*Cumulative*			
h		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.

3.3.1. Marginal GaR forecasting

Table 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 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 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 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.

[15] Table OA11 in the Online Appendix extends this finding to the set of bivariate quantile regressions including the lagged GDP and each potential predictor considered.

Table 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).
$^{***}p < .01$; $^{**}p < .05$; $^{*}p < .10$.

3.3.2. Joint GaR forecasting

Table 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.

Table 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.	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.	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.	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.	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).

Table 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
1	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**
	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**
2	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**
	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**
3	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
	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**
4	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
	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. $^{***}p < .01$; $^{**}p < .05$; $^{*}p < .10$.

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.

3.3.3. Robustness analysis

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

Table 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.

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 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.

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 3.3 with the h-step ahead growth rate replaced by the cumulative

[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.

h-step ahead GDP growth.[17] Table 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 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 OA1 (in the Online Appendix) and 8 contain the 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 OA2 of the Online 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 Online 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.

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 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 suggests caution against relying too heavily on this technique.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jmoneco.2020.11.003.

References

Adrian, T., Boyarchenko, N., Giannone, D., 2019. Vulnerable growth. American Economic Review 109 (4), 1263–1289.
Adrian, T., Brunnermeier, M.K., 2016. Covar. American Economic Review 106 (7), 1705–1741.
Adrian, T., Grinberg, F., Liang, N., Malik, S., 2018. The term structure of growth-at-riskWP.
Ahir, H., Bloom, N., Furceri, D., 2018. The world uncertainty indexWP.
Allen, L., Bali, T.G., Tang, Y., 2012. Does systemic risk in the financial sector predict future economic downturns? Review of Financial Studies 25, 3000–3036.
Amisano, G., Giacomini, R., 2007. Comparing density forecasts via weighted likelihood ratio tests. Journal of Business & Economics Statistics 25 (2), 177–190.
Azzalini, A., 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.
Baker, S.R., Bloom, N., Davis, S.J., 2016. Measuring economic policy uncertainty. Quarterly Journal Of Economics 131 (4), 1593–1636.
Beutel, J., 2019. Forecasting growth at riskWorking Paper.
Borio, C., Lowe, P., 2002. Asset prices, financial and monetary stability: Exploring the nexusBIS working paper 114.
Brownlees, C., 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., Kelly, B., 2011. A practical guide to volatility forecasting through calm and storm. Journal of Risk 14 (2), 3–22.
Caldara, D., Iacoviello, M., 2018. Measuring geopolitical riskInternational Finance Discussion Papers 1222.
Caldera Sanchéz, A., Röhn, 2016. How do policies influence gdp tail risks?WP.
Carriero, A., Clark, T.E., Marcellino, M.G., 2020. Capturing macroeconomic tail risks with bayesian vector autoregressionsFRB of Cleveland Working Paper no.20-02R.
Chavleishvili, S., Manganelli, S., 2019. Forecasting and stress testing with quantile vector autoregressionWP.
Christoffersen, P., 1998. Evaluating interval forecasts. Int. Econ. Rev. (Philadelphia) 39 (4), 841–862.

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

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

Claessens, S., Kose, A.M., Terrones, M.E., 2008. What happens during recessions, crunches and busts?IMF Working Paper.
De Nicolò, G., 2019. An early warning system for tail risksWorking Paper.
ECB, 2019. Financial Stability Review, May 2019. ECB.
Engle, R.F., Manganelli, S., 2004. Caviar: conditional autoregressive value at risk by regression quantiles. Journal of Business & Economics Statistics 22 (4), 367–381.
Estrella, A., 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., Zakrajsek, E., 2013. Credit spreads as predictors of real time economic activity. Review of Economics and Statistics 95 (5), 1501–1519.
Faust, J., Wright, J., 2013. Forecasting inflation. Handbook of Economic Forecasting 2A, 3–56.
Ghysels, E., Iania, L., Striaukas, J., 2018. Quantile-based inflation risk models. Technical Report. National Bank of Belgium.
Giacomini, R., Komunjer, I., 2005. Evaluation and combination of conditional quantile forecasts. Journal of Business & Economics Statistics 23 (4), 416–431.
IMF, 2017. Global Financial Stability Report: Is Growth at Risk?. In: Global Financial Stability Report, 3. International Monetary Fund, pp. 91–118.
Koenker, R.W., Basset Jr., G., 1978. Regression quantiles. Econometrica 46 (1), 33–50.
Pakel, C., Shephard, N., Sheppard, K., 2011. Nuisance parameters, composite likelihoods and a panel of GARCH models. Stat. Sin. 21 (1), 307–329.
Plagborg-Møller, M., Reichlin, L., Ricco, G., Hasenzagl, T., 2020. When is growth at risk? Working Paper.
Prasad, A., Elekdag, S., Jeasakul, P., Lafarguette, R., Alter, A., Feng Xiaochen, A., Wang, C., 2019. Growth at risk: Concept and application in imf country surveillance.WP.
Romano, J.P., Shaikh, A.M., Wolf, M., 2010. Hypothesis testing in econometrics. Annu. Rev. Econom. 2 (1), 75–104.
Romano, J.P., Wolf, M., 2007. Control of generalized error rates in multiple testing. Ann. Stat. 35 (4), 1378–1408.
Stock, J.H., Watson, M.W., 2006. Forecasting with many predictors. Handbook of Economic Forecasting 1, 516–554.
White, H., Tae-Hwan, K., Manganelli, S., 2015. Var for var: measuring tail dependence using multivariate regression quantiles. J. Econom. 187, 169–188.
Wolf, M., Wunderli, D., 2015. Bootstrap joint prediction regions. Journal of Time Series Analysis 36 (1), 352–376.