ERASMUS UNIVERSITEIT ROTTERDAM

The relevance of regime-switching models for macroeconomic downside risk prediction: A GaR forecasting analysis

Master Thesis Quantitative Finance

Jorian van der Kuilen - 427233

Supervisor: Prof. Dr. C. Zhou
Co-reader: B. van Os

January 29, 2022

ABSTRACT

This study investigates the relevance of regime-switching models for macroeconomic downside risk prediction. More specifically, the study investigates the one-quarter ahead to one-year ahead GaR forecasting performances of a unique implementation of the MS-GARCH model relative to the forecasting performances of the single-regime GARCH, EGARCH and GJR-GARCH models. The data consists of a balanced panel of 25 OECD countries with GDP growth rates ranging from 1961Q1 to 2020Q4. To tackle the scarcity problem of macroeconomic data, the study relies on the assumption of homogeneous model parameters across countries. The validity of this assumption in this study is demonstrated by the small impact of heterogeneity across countries on our findings. The results show the significant outperformance of the MS-GARCH model compared to single-regime models in forecasting one-quarter ahead GaR and the incapability of the MS-GARCH to improve upon single-regime models in forecasting GaR on longer forecasting horizons.

Contents

I. Introduction

Following the Global Financial Crisis, there has been a growing worldwide interest in the relationship between financial stability and business cycles. The failure to predict the Global Financial Crisis has motivated policymakers to increasingly focus on economic downside risk, causing a renewed development of downside risk measures based on key economic variables. Financial vulnerabilities contain valuable information about future economic growth. Therefore, tracking the evolution of financial vulnerabilities is of large value to economic policymakers at central banks and governments. In 2017, the International Monetary Fund (IMF) proposed the 95% Growth-at-Risk (GaR) as a measure to quantify the likelihood and extent of extreme macroeconomic events (IMF (2017)). GaR defines the minimum real economic growth at a certain confidence level and is considered as the macroeconomic counterpart of Value-at-Risk (VaR), which is extensively used in the field of risk management. Due to the different behavior of upside and downside risk and varying relationships between real economic growth and financial predictors across the distribution of GDP growth, the lower quantiles contain more information on downside risk than traditional point forecasts.

In a recent out-of-sample performance study, Brownlees & Souza (2021) extend GaR literature by proposing the use of GARCH-type models for downside risk prediction. Despite the popularity of quantile regressions in GaR literature, Brownlees & Souza (2021) find evidence that GARCH-type models yield more accurate GaR forecasts. The GARCH-type specifications implemented in Brownlees & Souza (2021) assume constant model parameters and therefore ignore different macroeconomic developments across phases of the business cycle. Over the past decades, the political importance of correctly identifying and predicting the business cycle phases of economies has motivated researchers to investigate the cyclical behavior of GDP growth rates. Since the introduction of the Markov-switching model by Hamilton (1989), a large variety of studies implements Markov-switching approaches to capture the cyclical behavior of GDP growth across phases of the business cycle. This body of literature provides broad evidence of regime-dependence of GDP growth rates. Since the popularization of GaR by IMF (2017), the out-of-sample performance of regime-switching approaches to model GaR is unexplored. By contrast, the more extensive VaR framework provides evidence for the strong ability of the regime-switching models to capture volatility dynamics of financial time series. The high GaR forecasting accuracy of GARCH-type models in Brownlees & Souza (2021), the evidence of cyclical and regime-dependent behavior of GDP growth rates, and the high performance of regime-switching GARCH models within the VaR framework raise expectations on the performance of regime-switching approaches in GaR applications. Based on this reasoning, it is the aim of this paper to investigate the GaR forecasting performance of a regime-switching GARCH model relative to single-regime GARCH models.

The study builds on earlier financial and macroeconomic studies that implement single-regime and regime-switching applications of volatility forecasting models. More specifically, we construct short-horizon (one-quarter ahead) and long-horizon (two- to four-quarter ahead) GaR forecasts using single-regime and regime-switching GARCH models. The data sample consists of a balanced panel of GDP growth rates of 25 OECD countries ranging from 1961Q1 to 2020Q4. The GDP growth rates are defined as the quarterly percentage change of seasonally adjusted real GDP, compared to the same quarter in the previous year. The high significance of the first-order autocorrelation and insignificance of most higher-order autocorrelations in the sample motivates our choice of conditional mean and conditional variance processes that include one lag, leaving the implementation of higher order autoregressive models for further research. In the single-regime and regime-switching models, the conditional means of GDP growth follow AR(1) and Markov-switching AR(1) processes, respectively. Moreover, to model the conditional variance of GDP growth in the single-regime framework, we implement standard GARCH, Exponentional GARCH (EGARCH), and Glosten-Jagannathan-Runkle GARCH (GJR) specifications. The EGARCH and GJR models extend the GARCH model by relaxing parameter restrictions and allowing for asymmetric responses to volatility shock, respectively. The implementation of multiple single-regime variance equations contributes to the breath of the conclusion on the performances of a regime-switching model relative to single-regime models. In the regime-switching framework, we confine ourselves to the two-state Markov-switching GARCH (MS-GARCH) model that allows for different GDP growth dynamics during expansionary and recessionary periods. The choice for two states is based on the advantage of more GDP growth observations per state and the vast majority of macroeconomic studies that investigate the number of regimes that GDP growth exhibits. To evaluate the GaR forecasting performances of the MS-GARCH model against the performances of the single-regime models, we rely on the comparative backtesting framework proposed by Nolde & Ziegel (2017). After the evaluation of the relative short-horizon and long-horizon GaR forecasting performances, we deepen the analysis by investigating the relative prediction power of the models during expansionary and recessionary periods to validate the economic significance of our findings. In addition to the main part of this research on marginal GaR modelling, we construct joint GaR forecasts in the extension of this paper to provide an indication on the relevance of regime-switching models to policymakers focusing on joint macroeconomic risk prediction.

Dissimilarities between macroeconomic and financial data call for a different GaR modelling approach compared to the VaR framework. Due to the quarterly frequency of macroeconomic observations, series on GDP growth are often shorter than financial data series. The scarcity of macroeconomic data poses a problem to the estimation of GARCH models, which require moderately large data series to obtain stable parameters. To avoid unstable parameter estimates, we assume homogeneous model parameters across countries to pool information across the panel rather than using information from individual countries only. Studies in macroeconomic literature debate on a continuous basis on the validity of homogeneous model parameters to model GDP growth. To evaluate the validity of the assumption of homogeneous GDP dynamics across countries in this study, we perform the in-sample analysis and the forecasting analysis on two samples. Besides the sample of 25 OECD countries, we propose a sample that omits countries with most aberrant GDP dynamics based on the aggregated log likelihood over an appropriate sample period. Throughout this study, the samples that comprise 25 and 20 OECD countries are denoted by the 100%-sample and the 80%-sample, respectively. If the analyses prove that omitting the countries that exhibit aberrant GDP dynamics has an insignificant impact on the GaR forecasting performances of the models, we consider the assumption of homogeneous model parameters across countries in this study as valid. Moreover, it provides evidence for the robustness of our findings with respect to the countries in the sample. Other modelling differences of GaR compared to VaR relate to the serial correlation of GDP growth, the specific error distribution of GDP growth rates, and the fixed confidence level of GaR.

We summarize the findings of this study in a three-fold conclusion. First, the short-horizon forecasting exercise shows the statistically significant outperformance of the MS-GARCH model relative to single-regime models in forecasting one-quarter ahead GaR. As the deepening analysis confirms the economic singificance of the finding, the outperformance of the MS-GARCH model in predicting one-quarter ahead macroeconomic downside risk reveals the relevance of regime-switching models to economists and policymakers. Second, on longer forecasting horizons, the MS-GARCH model is incapable of improving upon single-regime benchmarking models. Although the MS-GARCH model produces significantly more accurate one-year ahead GaR predictions, the outperformance is not considered as economically significant based on the high dependence of the outperfomance on phases of the business cycle and the unrealistic patterns of the forecasts during expansionary periods. The third component of our conclusion is the demonstrated validity of the assumption of homogeneous model parameters across countries in this study. The similar GaR forecasting performances of the models before and after excluding Finland, Greece, Iceland, South-Korea, and Luxembourg as countries that exhibit most aberrant GDP growth dynamics, indicates the small impact of heterogeneity across countries on the findings of this study. In addition to the main findings of this study, the extension on joint GaR suggests the potential relevance of regime-switching models to policymakers focusing on joint macroeconomic downside risk prediction, although significant evidence is not included in this study.

This paper is organized as follows. Section II provides a literature overview. The literature section describes the novelty of this paper by means of overviews of bodies of literature on GaR, regime-dependence of GDP growth and applications of the MS-GARCH model. Moreover, Section III presents the methodology of this paper. The section sequentially defines the GaR framework, discusses the construction of GaR forecasts, and describes the backtesting framework to evaluate relative GaR forecasting performances. Section IV describes the sample of GDP growth rates that this study utilizes. Section V presents the main results of this paper on the modelling of the marginal GaR. In addition, Section VI provides the extension on joint GaR. Finally, Section VII concludes and VIII provides the discussion and avenues for further research.

II. Literature

The increasing focus of policymakers for macroeconomic downside risk has motivated researchers to enlarge the variety of tools to assess the likelihood and extent of extreme economic events (Brownlees & Souza (2021)). GaR studies comprise various methods to produce forecasts of quantiles of the density of GDP growth rates. Quantile regressions allow for direct linkage of quantiles of GDP growth distributions to financial downside risk predictors. This property makes quantile regressions popular for downside risk prediction. For instance, Adrian & Brunnermeier (2016) construct systemic risk measures based on a quantile regression approach. Moreover, based on panel quantile regressions with country fixed effects, Adrian, Grinberg, Liang & Malik (2018) find that the predictive ability of financial conditions is higher in periods of high credit risk. The quantile regression implementation of Adrian, Boyarchenko & Giannone (2019) discovers that macroeconomic downside risk variability is higher than macroeconomic upside variability. Chavleishvili & Manganelli (2019), Carriero, Clark & Marcellino (2020), and Plagborg-Møller, Reichlin, Ricco & Hasenzagl (2020) propose structural quantile VAR (QVAR), Bayesian VAR (BVAR), and dynamic skew-t models, respectively, as alternative models to quantile regression for economic downside risk prediction. Despite specific advantages of the models, none of them consistently provides better GaR forecasts than the quantile regression approach as implemented by Adrian et al. (2019).

The recent study of Brownlees & Souza (2021) extends the GaR literature by proposing the use of GARCH-type models for macroeconomic downside risk prediction. The forecasting exercise of the study shows that GARCH-type specifications yield more accurate GDP growth density forecasts than quantile regressions. Brownlees & Souza (2021) provide two possible explanations for this finding. First, GARCH-type models are more robust for prediction. In fact, the relevance and predictive ability of downside risk predictors in quantile regressions vary over time, while GARCH-type models are agnostic about downside risk predictors. Second, the scarcity of macroeconomic time series impedes quantile regressions to capture the dynamics of extreme conditional quantiles.

Whereas the performance study by Brownlees & Souza (2021) assumes a constant distribution of GDP growth over time, a broad field of macroeconomic literature investigates the cyclical movements of GDP growth and of the central moments of its conditional distribution. Cette, Fernald & Mojon (2016) and Antolin-Diaz, Drechsel & Petrella (2017) conclude that the long-term means of GDP growth in, respectively, the US and Europe, declined prior to the Great Recession. Moreover, Jurado, Ludvigson & Ng (2015) investigate the volatility of GDP growth during economic turmoils and find stark counter-cyclicality of GDP growth rates. This implies that downside volatility is higher than upside volatility during recessionary periods. Giglio, Kelly & Pruitt (2016) perform a broad analysis on the skewness of the distribution of GDP growth rates of the US and multiple European economies. The study concludes that the skewness of the density of GDP growth rates sharply decreases as economies tumble into recessions. The implementation of a parametric Skew-t model in Delle Monache, De Polis & Petrella (2020), that allows for skewness shifts of GDP growth around business cycle points, supports the finding of Giglio et al. (2016) by delivering out-of-sample

US GDP growth density forecasts that improve upon standard benchmarks.

A popular approach to capture the cyclical behavior of GDP growth across phases of the business cycle is Markov-switching. Markov-switching models involve multiple structures that account for different behaviors within time series, while switching between these structures allows for capturing complex dynamic patterns. Since the introduction in Hamilton (1989), researchers use Markov-switching to model the cyclical behaviour of GDP growth across phases of the business cycle. The empirical application of Markov-switching in Hamilton (1989) suggests different behavior of US Gross National Product (GNP) growth rates in recessionary and expansionary states of the economy. The political importance of identifying business cycle phases has given rise to numerous implementations of Markov-switching models to capture the regime-dependence of GDP growth rates. For instance, Krolzig (1997) explores the use of the multivariate MS-VAR model to capture the business cycle dynamics of GDP growth in the Euro-zone. Furthermore, Clements & Krolzig (1998) show the superiority of the MS-AR model to linear single-regime models in modelling US GNP business cycle features. Other papers reveal the successfulness of Markov-switching methods in predicting business cycle turning points. Lahiri & Wang (1994) perform a two-state Markov-switching approach and manage to signal all US business cycle turning points between 1954 and 1992. Chauvet & Hamilton (2006) consider the unconditional distribution of GDP growth rates as a mixture of a so-called recession density and an expansion density and use this to construct well-performing turning point forecasts of US GDP growth rates. Moreover, motivated by the evidence of regime-dependence of the variability of GDP growth, as derived by McConnell & Perez-Quiros (2000), Bhar & Hamori (2003) propose a two-state MS-GARCH method to capture the volatility dynamics of GDP growth in Japan, UK and USA. Finally, Duprey, Klaus & Peltonen (2017) conclude that Markov-switching models outperform single-regime models in predicting European business cycle phases by implementing financial stress measures in a Markov-switching framework. This literature overview expounds evidence for regime-dependence of GDP growth and the conventionality of the assumption of two states of the economy. The transmission of a methodology that allows for regime-dependence to the GaR framework makes our research a novel contribution to the growing GaR literature.

The current lack of implementations of regime-switching approaches to model GaR is in stark contrast with the more extensive VaR literature. Financial risk management studies provide broad evidence of the outperformance of regime-switching VaR forecasting methods relative to single-regime approaches. More specifically, financial risk management literature comprises a large variety of studies that find MS-GARCH models to deliver higher VaR forecasting accuracies than single-regime GARCH models. The initial applications of Markov-switching autoregressive models reveal that MS-GARCH specifications suffer from lower tractability of likelihood estimations than MS-ARCH models (Cai (1994), Hamilton & Susmel (1994), and Gray (1996)). To solve the 'path-dependence' problem of the MS-GARCH model, Gray (1996), Dueker (1997), and Klaassen (2002) implement different approaches to aggregate past regime-specific conditional variances. Haas, Mittnik & Paolella

(2004) propose an alternative solution and rely on the assumption that regime-specific variances only depend on their own lagged values and past shocks, rather than on all previous regimes. Other approaches to tackle the path-dependence problem are mostly alternatives to traditional maximum likelihood estimation. For instance, Bayesian estimation methods offer the advantage of integrating parameter uncertainty through Markov Chain Monte Carlo (MCMC) techniques (e.g. Ardia et al. (2008), Geweke & Amisano (2010), and Bauwens, Dufays & Rombouts (2014)), while Augustyniak (2014) implements Monte Carlo approaches that rely on importance sampling. The various solutions to the path dependence problem allow for a large variety of studies that rely on MS-GARCH implementations to model the dynamics of financial time series with greater success than single-regime models. Marcucci (2005) concludes that MS-GARCH models generate better short-horizon S&P100 volatility forecasts than single-regime GARCH, EGARCH and GJR-GARCH specifications under both statistical and VaR-based risk management loss functions. Sajjad, Coakley & Nankervis (2008) confirm the outperformance of the MS-GARCH model based on an expansion of the study of Marcucci (2005) focusing on one-day-ahead VaR for both long and short positions on the FTSE100 and S&P 500. Moreover, BenSaïda, Boubaker, Nguyen & Slim (2018) find superiority of the MS-GJR-GARCH model to single-regime specifications in signaling and predicting unobservable market volatility shocks, as well as in constructing capital charges based on VaR forecasts. Finally, Ardia, Bluteau, Boudt & Catania (2018) include both frequentist and Bayesian estimation methods and compare the forecasting performances of 32 single-regime and MS-GARCH specifications from a risk management perspective. The large-scale empirical study on stock return, stock market indices, and exchange rates shows the outperformance of MS-GARCH specifications relative to their single-regime counterparts in terms of VaR, Expected Shortfall (ES) and left-tail distribution forecasts.

Opposed to the described studies that focus on financial applications, the modelling of GDP growth using regime-switching volatility forecasting requires methodological adjustments that are proposed in separate studies. Based on the uniqueness of a simultaneous implementation of the methodological adjustments, this paper forms a novel contribution to the framework on Markov-switching volatility forecasting models.

III. Methodology

This section outlines the methodology of our paper. First, Section III.A defines the GaR framework. Next, Section III.B describes the methodology to construct GaR forecasts using single-regime and regime-switching volatility forecasting models. Finally, Section III.C describes the backtesting framework of this study to evaluate the relative GaR forecasting performances of the models.

A. The Growth-at-Risk framework

The IMF defines GaR as the minimum growth in real economic activity at a certain conditional probability and horizon. More specifically, GaR denotes a predetermined quantile of the conditional distribution of GDP growth. Throughout this study, the GDP growth rate of country i at quarter t is denoted as $Y_{i,t}$, where subscripts i and t range from $i = 1, \ldots, n$ and $t = 1, \ldots, T$, respectively. Equation (1) formalizes the IMF (2017) definition of the H-quarter ahead marginal GaR at the conditional probability level c:

$$\Pr\big(Y_{i,t+H} \leq GaR_{i,t+H|t}(c)|\zeta_t\big) = c, \tag{1}$$

where ζ_t denotes the information set available at quarter t.

Despite the high degree of similarity between GaR and VaR by definition, dissimilarities between macroeconomic and financial contexts call for adjustments to model GaR. First, as discussed in the introduction of this paper, to solve the problem of scarce macroeconomic data, we assume that model parameters are homogeneous across countries, which ensures stability of parameter estimates. This assumption allows for pooling information across the panel rather than using information from individual countries only. Under the assumption, we follow Brownlees & Souza (2021) and implement Composite Quasi-Maximum Likelihood (CQML) estimation. Composite Likelihood is proposed by Pakel, Shephard & Sheppard (2011) as an estimation procedure that exploits cross-sectional commonality to obtain stable estimates of panel GARCH parameters. Composite Likelihood estimation allows for estimation of parameters of the conditional mean and variance processes on the full data sample, rather than on data series of individual countries. Pakel et al. (2011) show by means of Monte Carlo simulations the ability of Composite Likelihood to successfully estimate conditional volatility on panels with reasonably large lengths of time series. In correspondence with Brownlees & Souza (2021), we label the implemented CQML estimation procedure as naïve due to the ignorance of the serial dependence of the multi-step innovations. To keep this section on the GaR framework parsimonious, the mathematical outlook of the estimation procedure is detailed in Section III.B.3.

Second, financial literature routinely relies on the absence of serial correlation in financial error terms. In the majority of financial studies, this allows to model the conditional mean of financial variables as a random walk. By contrast, macroeconomic literature provides wide evidence for the autoregressive character of GDP growth rates (e.g. Baffigi, Golinelli & Parigi (2004), Marcellino, Stock & Watson (2003) and Brownlees & Souza (2021)). Moreover, Section IV provides empirical evidence for significant first-order autocorrelation in the sample of GDP growth rates of this study. This motivates the choice for first-order autoregressive processes to model the conditional mean of GDP growth. More specifically, this study relies on the AR(1) and MS-AR(1) processes to model the conditional mean in the single-regime and regime-switching models, respectively. Section III.B.2 elaborates on the mathematical expressions of the conditional mean processes.

Third, VaR studies assume a large variety of conditional distributions to apply single-regime and regime-switching GARCH models on financial data. However, the distribution of GDP growth rates deviates considerably from distributions of financial returns and other financial variables. Adrian et al. (2019) derive that the entire conditional distribution of GDP growth rates evolves over time, rather than just the central tendency. For instance, GDP growth tends to be less volatile during expansionary periods than during recessions. The ability of the MS-GARCH model to switch to different regimes in case of shocks to volatility, accounts for the time-varying characteristics of GDP growth rates. Moreover, the conditional distribution of GDP growth is widely found to exhibit fat tails. For instance, Fagiolo, Napoletano & Roventini (2008) find excess kurtosis in the conditional distribution of GDP growth of a majority of OECD countries and demonstrate that fat tails characterize the distribution even after omitting outliers. On the skewness of the conditional distribution of GDP growth, Brownlees & Souza (2021) show that the standardized GARCH residuals of GDP growth do not exhibit systematic patterns for negative or positive skewness. Section IV investigates the distributional characteristics of the GDP growth rates in our data sample. Based on macroeconomic literature and on the empirical findings in Section IV, this study assumes GDP growth rates to follow a Student's t-distribution.

Finally, VaR and GaR studies differ in terms of the conditional probability level c. To determine the conditional probability level, VaR literature often refers to the capital requirements of financial institutions that are regulated under the Basel Accords. In line with the quantitative standards that are prescribed by the Basel III Committee, many VaR studies focus on the 99% VaR (Sharma (2012)). However, in a macroeconomic context, the central tendency is to follow the IMF to employ the conditional probability level $c = 95\%$. Based on this tendency, the conditional probability level is fixed as c=95% level throughout this study.

B. Modelling Growth-at-Risk

This section details the modelling of GaR based on the following structure. First, Section III.B.1 describes the modelling of GDP growth using single-regime and regime-switching GARCH-type models. Next, Section III.B.2 provides an overview of the implemented processes of the conditional mean and the conditional variance of GDP growth. Third, Section III.B.3 describes the estimation procedure to obtain the parameter estimates of the models. Finally, Section III.B.4 discusses the construction of short-horizon and long-horizon GaR forecasts.

B.1. GDP growth modelling in volatility forecasting models

The modelling of GDP growth in the single-regime framework relies on

$$Y_{i,t} = \mu_{i,t} + \left(h_{i,t}\right)^{1/2} u_{i,t}, \quad u_{i,t} \sim \mathcal{D}_i \tag{2}$$

where $\mu_{i,t}$ and $h_{i,t}$ are the conditional mean and conditional variance of GDP growth, respectively. Moreover, the error terms, denoted by $u_{i,t}$, are independent and identically distributed according to conditional distributions $\mathcal{D}_i$. The regime-switching framework permits the parameters in Equation (2) to switch across regimes, such that

$$Y_{i,t} = \mu_{i,t}^{(k)} + \left(h_{i,t}^{(k)}\right)^{1/2} u_{i,t}^{(k)}, \quad u_{i,t}^{(k)} \sim \mathcal{D}_i^{(k)}, \tag{3}$$

where k indicates the respective state of the economy. As discussed in the introduction section, this study follows the majority of macroeconomic literature to allow for different GDP growth dynamics during recessionary and expansionary periods, such that the number of regimes in the regime-switching framework equals two. Therefore, Expression (3) implies that the regime-switching framework models the conditional distribution of GDP growth as a mixture of two densities,

$$Y_{i,t}|\zeta_{t-1} \sim \begin{cases} f^{(\mathcal{D})}(\mu_{i,t}^{(1)}, h_{i,t}^{(1)}, \nu_i^{(1)}) & \text{with probability } \pi_{i,t}^{(1)} \\ f^{(\mathcal{D})}(\mu_{i,t}^{(2)}, h_{i,t}^{(2)}, \nu_i^{(2)}) & \text{with probability } 1-\pi_{i,t}^{(1)} \end{cases}, \tag{4}$$

where $\pi_{i,t}^{(k)}$ denote the ex-ante probabilities that are clarified in the next paragraph. Expressions (2) and (4) imply that the innovation distribution $\mathcal{D}$ determines the shape of the conditional density of future GDP growth in the single-regime framework and the regime-switching framework. Based on the considerations in Section III.A, the conditional error terms $u_{i,t}$ are assumed to follow a Student's t-distribution. In the single-regime models, the Student's t-distribution of the error terms is characterized by a zero mean, unit variance and country-specific shape parameters ν_i. Additionally, the regime-switching framework allows for different shape parameter values $\nu_i^{(k)}$ across the regimes. To ensure existence of the second moment, the shape parameters are restricted to exceed two. Based on the shape parameters, the excess kurtosis of the Student's t-distribution is defined as $6/(\nu-4)$ if $\nu > 4$ and as infinitely large if $2 \leq \nu < 4$.

In general, the Markov-switching model assumes a latent variable $s_{i,t}$ for all sample countries i, $i = 1, \ldots, n$, that evolves according to an unobserved first-order Markov chain with transition probability matrix $\mathbf{P}$. The elements of the transition probability matrix denote the transition probabilities $p^{(l,k)} = \Pr(s_{i,t+1} = k|s_{i,t} = l)$. An essential ingredient to the implementation of the MS-GARCH model based on maximum likelihood estimation are the ex-ante probabilities, $\pi_{i,t}^{(k)}$. The ex-ante probabilities denote the probability of being in a regime at quarter t given the available information from the previous quarter, $t-1$. The ex-ante probability of regime k at quarter t is specified as $\pi_{i,t}^{(k)} = \sum_{l=1}^{2} p^{(l,k)} * \eta_{(i,t-1)}^{(l)}$, where $\eta_{(i,t-1)}^{(l)}$ denotes the filtered probability of regime l at quarter $t-1$. The filtered probabilities $\eta_{(i,t)}^{(k)}$ reflect the regime-probabilities at quarter t, conditional on the information set ζ_t, and are obtained through the Hamilton filter. We refer to Hamilton (1989) and Hamilton & Susmel (1994) for a detailed description of the Hamilton filter. Moreover, we rely on Kim's smoother, derived by Kim (1994), to smooth the filtered probabilities conditional on the full sample information set, ζ_T.

B.2. Conditional mean and conditional variance processes

To make the modelling of GDP growth rates in Equations (2) and (3) operational, we must specify the processes for the conditional mean and the conditional variance. The literary evidence of serial dependence of GDP growth rates, as discussed in Section II, and the empirical evidence in Section IV motivate the use of first-order autoregressive processes to model the conditional mean of GDP growth. Similar to Brownlees & Souza (2021), our study specifies the one-quarter ahead conditional mean of the growth of GDP of country i at quarter t, in single-regime models as

$$\mu_{i,t+1} = \phi_0 + \phi_1 Y_{i,t}, \tag{5}$$

where the autoregressive parameter is restricted as $|\phi_1| < 1$ to ensure stationarity of the process. In the single-regime framework, we define the difference between the realized GDP growth and the conditional mean as the innovation term, i.e. $\epsilon_{i,t} = Y_{i,t} - \mu_{i,t}$. To model the conditional variance in the single-regime framework, we implement three different models. First, we consider the GARCH model. The GARCH(1,1) variance equation calculates the one-quarter ahead conditional variance of the GDP growth of country i at quarter t as

$$h_{i,t+1} = \omega + \alpha \epsilon_{i,t}^2 + \beta h_{i,t}, \tag{6}$$

where we impose parameter restrictions $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ to ensure positive conditional variance and $\alpha + \beta < 1$ to ensure stationarity of the conditional variance. In the GARCH(1,1) model, the unconditional volatility and the persistence of shocks are calculated as $\sigma = (\omega/(1 - \alpha - \beta))^{1/2}$ and $\rho = \alpha + \beta$, respectively, to improve the interpretation of the conditional variance parameters.

The second single-regime model that we implement is the Exponential GARCH (EGARCH) model as proposed by Nelson (1991). Motivated by the argument that the non-negativity constraints of the GARCH model are too restrictive, Nelson & Cao (1992) propose the EGARCH(1,1) variance equation

$$\ln h_{i,t+1} = \omega + \alpha \frac{|\epsilon_{i,t}|}{|h_{i,t}|} + \gamma \frac{\epsilon_{i,t}}{h_{i,t}} + \beta \ln h_{i,t}, \tag{7}$$

where $|\beta| < 1$ for stationarity. As γ is generally negative, the EGARCH model allows negative shocks to have more impact on the conditional variance than positive shocks. Financial researches that study the EGARCH model call the tendency of bad news (negative error terms) to have a larger impact on volatility than good news (positive error terms) the leverage effect and denote γ as the leverage parameter.

Third, this study implements the Glosten-Jagannathan-Runkle GARCH (GJR) model. Glosten, Jagannathan & Runkle (1993) propose the GJR model as an alternative approach to allow for asymmetric responses of conditional variance to positive and negative shocks. Under the GJR-GARCH(1,1) specification, the conditional variance of GDP growth is modelled as

$$h_{i,t+1} = \omega + \alpha \epsilon_{i,t}^2 (1 - \mathbb{I}[\epsilon_{i,t} > 0]) + \gamma \epsilon_{i,t}^2 \mathbb{I}[\epsilon_{i,t} > 0] + \beta h_{i,t}, \tag{8}$$

where the indicator function $\mathbb{I}[\cdot]$ equals 1 if the condition between brackets holds and 0 otherwise. The parameter restrictions $\omega > 0$, $\alpha \geq 0$, $\gamma \geq 0$, $\beta \geq 0$ and $\beta + (\alpha + \gamma)/2 < 1$ ensure positivity and stationarity of the conditional variance.

Regime-switching models allow parameters of the conditional mean and conditional variance to differ across regimes. More specifically, this implies that in the regime-switching framework, the conditional mean $\mu_{i,t}^{(k)}$ is modelled as an AR(1) process in both regimes $k = \{1, 2\}$ with state-dependent parameters $\phi_0^{(k)}$ and $\phi_1^{(k)}$,

$$\mu_{i,t+1}^{(k)} = \phi_0^{(k)} + \phi_1^{(k)} Y_{i,t}, \tag{9}$$

where $|\phi_1^{(k)}| < 1$ for $k = \{1, 2\}$. Likewise, the MS-GARCH(1,1) conditional variance equation computes the conditional variance as a function of state-dependent parameters only,

$$h_{i,t+1}^{(k)} = \omega^{(k)} + \alpha^{(k)} \epsilon_{i,t}^2 + \beta^{(k)} h_{i,t}, \tag{10}$$

with non-negativity restrictions $\omega^{(k)} > 0$, $\alpha^{(k)} \geq 0$ and $\beta^{(k)} \geq 0$ and stationarity restriction $\alpha^{(k)} + \beta^{(k)} < 1$ for $k = \{1, 2\}$. The regime-dependence of the conditional mean requires reformulation of the innovation terms $\epsilon_{i,t}$, such that in the regime-switching framework $\epsilon_{i,t} = Y_{i,t} - \left[\pi_{i,t}^{(1)}\ \mu_{i,t}^{(1)} + (1 - \pi_{i,t}^{(1)})\ \mu_{i,t}^{(2)}\right]$.

As described in the literature section of this study, researchers have proposed various approaches to tackle the path-dependence problem of MS-GARCH models. Equation (10) demonstrates that the past variance term $h_{i,t}$ does not only depend on information set ζ_{t-1} and current state $s_{i,t}$, but also on all states prior to quarter t. Consequently, the conditional variance at time t is subject to past variance patterns across multiple regimes. This implies that the computation of the variance term $h_{i,t}$ as in Equation (10) requires integration over an exponentially increasing number of paths, which is defined as the path-dependence problem. In this paper, we adopt the solution approach proposed by Klaassen (2002) and integrate out past regimes. More specifically, Klaassen (2002) replaces $h_{i,t}$ in Equation (10) by

$$E_t\left[h_{i,t}^{(k)} | s_{i,t+1}\right] = \sum_{l=1}^{2} \tilde{p}_{i,t}^{(l,k)} \left[\left(\mu_{i,t}^{(l)}\right)^2 + h_{i,t}^{(l)}\right] - \left[\sum_{l=1}^{2} \tilde{p}_{i,t}^{(l,k)} \mu_{i,t}^{(l)}\right]^2, \tag{11}$$

where the probabilities $\tilde{p}_{i,t}^{(l,k)}$ are calculated as

$$\tilde{p}_{i,t}^{(l,k)} = \Pr(s_{i,t} = l | s_{i,t+1} = k) = \frac{p^{(l,k)} \Pr(s_{i,t} = l | \zeta_{t-1})}{\Pr(s_{i,t+1} = k | \zeta_{t-1})} = \frac{p^{(l,k)}\ \pi_{i,t}^{(l)}}{\pi_{i,t+1}^{(k)}}. \tag{12}$$

$p^{(l,k)}$ and $\pi_{i,t}^{(k)}$ denote the transition probabilities and ex-ante probabilities, respectively.

B.3. Model estimation

To obtain parameter estimates, we rely on Composite Quasi-Maximum Likelihood (CQML) estimation. In the in-sample analysis, the models are estimated on the full data sample consisting of $n * T$ observations. By contrast, the out-of-sample estimation procedure relies on recurrent splits of the data sample into estimation windows and evaluation windows. In fact, H-quarter ahead GaR forecasts are generated from models estimated on moving estimation windows of τ_1 observations per country. At forecasting quarter t_0, the estimation window contains observations

$Y_{i,t_0-\tau_1+1}, \ldots, Y_{i,t_0}$ of all countries $i = 1, \ldots, n$, such that the successive $(T - t_0)$ observations per country in the remainder of the sample constitute the evaluation window. Using this notation, the forecasting quarter t_0 ranges from $t_0 = 1$ to $t_0 = T - H$. In addition, we denote the size of the initial evaluation window per country as τ_2, such that for all forecasting horizons, $\tau_1 + \tau_2 + H - 1 = T$.

For the sake of parsimony of the mathematical expressions in this section, we denote Ψ as the parameter set that includes all unique model parameters for all countries $i = 1, \ldots, n$. To commence, Ψ comprises the set θ that collects the parameters of the conditional mean and conditional variance processes. That is, $\theta = \{\phi_0^k, \phi_1^k, \omega^{(k)}, \alpha^{(k)}, \beta^{(k)}\}$ for $\{k = 1, 2\}$ in the MS-GARCH model, $\theta = \{\phi_0, \phi_1, \omega, \alpha, \beta\}$ in the single-regime GARCH model and $\theta = \{\phi_0, \phi_1, \omega, \alpha, \gamma, \beta\}$ in the single-regime EGARCH and GJR models. Second, Ψ includes the shape parameters of the conditional GDP growth distribution ν_i for all countries $i = 1, \ldots, n$ in the single-regime framework or $\nu_i^{(k)}$ for all countries and both states $k = \{1, 2\}$ in the regime-switching framework. Finally, in the MS-GARCH model, the (2×2) transition probability matrix $\mathbf{P}$ complements the parameter set Ψ.

The declaration of the ex-ante probabilities in Section III.A enables maximum likelihood estimation to obtain estimates of the parameters of the single-regime and regime-switching models. Mathematically, using the notation described in this section, the estimation of the parameters of the MS-GARCH model in the in-sample analysis yields maximizing the log-likelihood as

$$\hat{\Psi} = \arg\max_{\Psi} \sum_{i=1}^{n} \sum_{t=1}^{T} \sum_{k=1}^{2} \left(\log \, \pi_{i,t}^{(k)} \; f^{(\mathcal{D})}(Y_{i,t} | s_{i,t} = k, \Psi) \right), \tag{13}$$

where $f^{(\mathcal{D})}$ denotes the probability density function of the Student's t-distributed error terms with zero mean, unit variance and shape parameter $\nu_i^{(k)}$. Moreover, the out-of-sample estimation of the MS-GARCH model at forecasting quarter t_0 is formalized as

$$\hat{\Psi}_{t_0} = \arg\max_{\Psi_{t_0}} \sum_{i=1}^{n} \sum_{t=t_0-\tau_1+1}^{t_0} \sum_{k=1}^{2} \left(\log \, \pi_{i,t}^{(k)} \; f^{(\mathcal{D})}(Y_{i,t} | s_{i,t} = k, \Psi) \right) . \tag{14}$$

It is important to remark that Equations (13) and (14) imply the estimations of the single-regime models. In fact, setting the number of regimes equal to one eliminates the summation over regimes, the ex-ante probabilities and the regime-dependence of the parameters. In correspondence with Brownlees & Souza (2021), we label the CQML estimation procedures as naïve, as Expressions (13) and (14) ignore the serial dependence of multi-step innovations.

B.4. GaR forecasting

In the forecasting analysis, we distinguishes between the short-horizon (one quarter ahead) and long-horizon (two- to four-quarter ahead) prediction power of the models. Forecasting one-quarter ahead GaR relies on a two-step approach based on the observed macroeconomic conditions at quarter t_0. First, to construct one-quarter ahead forecasts of the conditional mean and conditional variance, we plug the observed GDP growth rates, error terms and conditional variances into the model-specific processes of the conditional mean and conditional variance (Equations (5)-(10), Section III.B.2). Thereafter, in the single-regime models, the second step of the two-fold approach constructs one-quarter ahead GaR forecasts as

$$GaR_{i,t_0+1|t_0}(c) = \mu_{i,t_0+1} + (h_{i,t_0+1})^{1/2} F^{-1}_{\mathcal{D}_i}(c), \tag{15}$$

where $F^{-1}_{\mathcal{D}_i}$ denotes the inverse cumulative distribution function of $\mathcal{D}_i$. Moreover, in the MS-GARCH model, one-quarter ahead GaR forecasts are obtained as

$$GaR_{i,t_0+1|t}(c) = \sum_{k=1}^{2} \pi^{(k)}_{i,t_0+1} \Big[\mu^{(k)}_{i,t_0+1} + \big(h^{(k)}_{i,t_0+1}\big)^{1/2} F^{-1}_{\mathcal{D}^{(k)}_i}(c)\Big]. \tag{16}$$

The construction of GaR forecasts as probability-weighted averages of quantiles of conditional univariate distributions as in Equation 16 is an approximation of quantiles of the conditional mixture distribution of GDP growth rates. The approximation ignores the interdependencies between the distributional characteristics of GDP growth, which are assumed to be small in this study. The approximation is advantageous in terms of implementation time relative to the theoretically correct alternative of Monte Carlo simulation. Therefore, we employ the approximation as in Equation (16) in this study and leave the implementation of Monte Carlo approaches to model quantiles of the conditional mixture distribution of GDP growth rates for further research.

Long-horizon GaR forecasts

To forecast GaR on horizons that are larger than one quarter, the GaR predictions require unknown macroeconomic information between the forecasting quarter t_0 and the predicted quarter $t_0 + H$. To mimic the feasible intermediate macroeconomic circumstances, we rely on Monte Carlo simulations. The Monte Carlo approach determines the H-quarter ahead GaR of country i as the 5% empirical quantile of simulated GDP growth rates $\{\tilde{Y}^{\{w\}}_{i,t_0+H}\}^{W}_{w=1}$, where W denotes the size of the simulation. In correspondence with Brownlees & Souza (2021), the Monte Carlo approach in this study comprises $W = 1000$ simulations.

The simulation of H-quarter ahead GaR requires the implementation of a two-step procedure that recurs for $j = 1, \dots, H$. First, we construct j-quarter ahead conditional mean and conditional variance forecasts using the processes defined in Section III.B.2. If $j = 1$, the observed GDP growth and resulting error terms and conditional variance determine the forecasts μ_{i,t_0+1} and h_{i,t_0+1}. If $j > 1$, the forecasts $\mu^{\{w\}}_{i,t_0+j}$ and $h^{\{w\}}_{i,t+0+j}$ of simulation w rely on simulated GDP growth rates $\tilde{Y}^{\{w\}}_{i,t_0+j-1}$ and corresponding error terms $\epsilon^{\{w\}}_{i,t_0+j-1}$, and conditional variances $h^{\{w\}}_{i,t+j-1}$. The simulation of GDP growth rates constitutes the second step of the recurrent procedure, which consists of simulating error terms $u^{\{w\}}_{i,t_0+j} \overset{iid}{\sim} \mathcal{D}_i$ or $u^{(k)\{w\}}_{i,t_0+j} \overset{iid}{\sim} \mathcal{D}^{(k)}_i$ and computing simulated GDP growth rates based on the conditional mean forecasts, conditional variance forecasts and error terms as in Equations (2) and (3). In the regime-switching framework, the determination of the state of the economy of country i at quarter $t_0 + j$ is based on a uniformly distributed variable and the regime-probabilities $\pi^{(k)\{w\}}_{i,t+j}$, such that the simulated GDP growth rates are obtained from a mixture distribution. Algorithm 1 in Appendix A details the mathematical steps to construct long-horizon GaR forecasts.

Code description

For the estimation of the single-regime and regime-switching models, as well as for the forecasting of short-horizon GaR, this study takes Matlab code developed by Marcucci (2005) as a starting point. The initial reliance of modelling on the code provided by Marcucci (2005) causes high levels of similarity on specific parts of the code of Marcucci (2005) and the code of this thesis. Due to the different contexts of the studies, the following adjustments to the code of Marcucci (2005) are made to model and forecast GaR. First, we expand the methodology of Marcucci (2005) to handle panel data by implementing the CQML estimation procedure as defined by Equations (13) and (14). Second, the autoregressive character of GDP growth requires the replacement of the random walk, to model the conditional mean of GDP growth, by AR(1)-processes. Third, the absence of serial correlation in financial data enables Marcucci (2005) to aggregate multi-step volatility forecasts to forecast long-horizon VaR. By contrast, this study implements a Monte Carlo approach to construct multi-step GaR forecasts.

C. Backtesting framework

To evaluate the GaR forecasting performances of the single-regime and regime-switching models, we implement the two-fold backtesting framework proposed by Nolde & Ziegel (2017). The backtesting framework is designed to determine the relative performances of risk measure forecasting methods. First, the unconditional and conditional calibration backtests provide a unifying framework to determine the ability of models to produce correctly calibrated risk measure forecasts. The unconditional and conditional calibration tests, that formalize traditional backtests, contribute to our understanding of the GaR forecasting abilities of the models. Moreover, we consider it as a minimum requirement that the proposed model passes both calibration tests to formulate conclusions on the outperformance of the model. After the calibration tests, the comparative backtest determines the mutual superiority of risk measure forecasting models based on consistent scoring functions. In this study, we follow Nolde & Ziegel (2017) to perform the two-fold backtesting framework at the significance level $\kappa = 0.05$.

The unconditional and conditional calibration tests rely on identification functions, rather than on loss functions. Following Nolde & Ziegel (2017), we define the GaR identification function of the H-quarter ahead GaR as $V_{i,t+H} = 1 - \alpha - \mathbb{I}[Y_{i,t} < GaR_{i,t+H|t}(c)]$. Based on the GaR identification function, the unconditional test statistic is given by

$$L_1 = \sqrt{n\tau_2}\ \hat{s}_V\ \overline{V}_{n\tau_2}, \tag{17}$$

where $\overline{V}_{n\tau_2} = \frac{1}{n\tau_2} \sum_{i=1}^{n} \sum_{t=\tau_1+H}^{T} V_{i,t}$ and $\hat{s}_V^2$ denotes the heteroskedasticity and autocorrelation (HAC) estimator of the asymptotic variance term $s_V^2 = var(\sqrt{n\tau_2}\ V)$, where the $(n \times \tau_2)$-vector

V consists of elements $V_{i,t}$. The null hypothesis of the test implies that the expectation of the identification function equals zero. Under the null hypothesis and under suitable assumptions on the identification function and the data-generating process, the test-statistic L_1 converges to a standard normally distributed variable. We refer to Nolde & Ziegel (2017) and Giacomini & White (2006) for a more detailed derivation of the test statistic and for the specific conditions.

Conditional calibration of the GaR forecasts is a stronger requirement than unconditional calibration. To test for correct conditional calibration, Nolde & Ziegel (2017) derive the test-statistic

$$L_2 = n\tau_2 \Big(\frac{1}{n\tau_2} \sum_{i=1}^{n} \sum_{t=\tau_1+H}^{T} V_{i,t}\mathbf{v}_{i,t} \Big)' \hat{\Omega}_{n\tau_2}^{-1} \Big(\frac{1}{n\tau_2} \sum_{i=1}^{n} \sum_{t=\tau_1+H}^{T} V_{i,t}\mathbf{v}_{i,t} \Big), \tag{18}$$

where $\mathbf{v}_{i,t+H} = \big(1, GaR_{i,t+H|t}(c)\big)'$ and $\hat{\Omega}_{n\tau_2}$ is the consistent variance estimator

$$\hat{\Omega}_{n\tau_2} = \frac{1}{n\tau_2} \sum_{i=1}^{n} \sum_{t=\tau_1+H}^{T} \big(V_{i,t}\mathbf{v}_{i,t}\big)\big(V_{i,t}\mathbf{v}_{i,t}\big)'.$$

The null hypothesis of the conditional calibration test implies that, conditional on the available information at quarter t, the conditional expectation of the identification function equals 0. Under this null hypothesis, as derived in Giacomini & White (2006), L_2 converges in distribution to a $\chi^2(2)$ distributed variable.

The comparative backtest compares the mutual forecasting accuracies of the models. In this study, we compare the forecasting performances of the proposed MS-GARCH models versus the single-regime GARCH, EGARCH and GJR benchmark models. Additionally, Section V.B.1 contains backtests with the GJR model as the proposed model to determine the superior single-regime model. The comparative backtest relies on two null hypotheses:

$$\begin{aligned}
&H_0^- : \text{The GaR forecasts of the proposed model are at least as accurate as the benchmark model forecasts,}\\
&H_0^+ : \text{The GaR forecasts of the proposed model are at most as accurate as the benchmark model forecasts.}
\end{aligned}$$

H_0^- adapts the null hypothesis of a correct model and correct estimation procedure to a comparative setting. To prevent the type I error that the proposed model is preferred over an established benchmark method based on the rejection of H_0^-, the proposed model passes the comparative backtest only if H_0^+ is rejected. To formalize the comparative backtest, Nolde & Ziegel (2017) define

$$\lambda^* := \limsup_{\tau_2 \to \infty} \frac{1}{n\tau_2} \sum_{i=1}^{n} \sum_{t=\tau_1+H}^{T} E\big[S_{i,t}^{PM} - S_{i,t}^{BM}\big] \text{ and}$$
$$\lambda_* := \liminf_{\tau_2 \to \infty} \frac{1}{n\tau_2} \sum_{i=1}^{n} \sum_{t=\tau_1+H}^{T} E\big[S_{i,t}^{PM} - S_{i,t}^{BM}\big] \ , \tag{19}$$

where subscripts ^{PM} and ^{BM} of scoring function $S_{i,t}$ indicate the respective proposed model and benchmark model. The consistent GaR comparative scoring function $S_{i,t}$ is defined as

$$S_{i,t} = \big(\mathbb{I}[Y_{i,t} < GaR_{i,t|t-H}] + \alpha - 1\big)\big(GaR_{i,t|t-H}(c)\big) - \mathbb{I}[Y_{i,t} < GaR_{i,t|t-H}(c)]\big(Y_{i,t}\big). \tag{20}$$

If $\lambda^* \leq 0$, the forecasts of the proposed model are interpreted as, on average, at least as good as the benchmark forecasts. Moreover, if $\lambda_* \geq 0$, this indicates that the GaR forecasts of the proposed model are, on average, at most as good as the forecasts of the benchmark model. This implies that the null hypotheses can be formulated as $H_0^- : \lambda^* \leq 0$ and $H_0^+ : \lambda_* \geq 0$. To test for both null hypotheses, Nolde & Ziegel (2017) introduce test statistic

$$L_3 = \frac{\Delta \overline{S}_{n\tau_2}}{\hat{s}_S / \sqrt{n\tau_2}}, \tag{21}$$

where $\Delta \overline{S}_{n\tau_2}$ is defined as

$$\Delta \overline{S}_{n\tau_2} := \frac{1}{n\tau_2} \sum_{i=1}^{n} \sum_{t=\tau_1+H}^{T} \big(S_{i,t}^{PM} - S_{i,t}^{BM}\big) \tag{22}$$

and $\hat{s}_S^2$ denotes the HAC estimator of the asymptotic variance $s_S^2 = var\big(\sqrt{n\tau_2}\Delta S\big)$, where $(n \times \tau_2)$-vector ΔS contains elements $S_{i,t}^{PM} - S_{i,t}^{BM}$ [1]. Under Theorem 4 of Giacomini & White (2006), L_3 is asymptotically standard normally distributed. This implies that we reject H_0^- when $1 - \Phi(L_3) \leq \kappa$ and H_0^+ when $\Phi(L_3) \leq \kappa$. Here, Φ denotes the Cumulative Density Function (CDF) of the standard normal distribution.

Finally, in accordance with Nolde & Ziegel (2017), the comparative backtest divides the possible outcomes of the comparative backtest into three zones. When H_0^+ is rejected, the comparative backtest assigns the proposed model model to the green zone, indicating significant outperformance relative to the benchmark model. A rejection of H_0^- yields a placement of the proposed model into the red zone, which implies that the proposed model produces signficantly worse GaR forecasts than the benchmark model. Lastly, when neither H_0^+ nor H_0^- is rejected, the comparative backtest assigns the proposed model to the yellow zone, indicating that the comparative backtest is indecisive on the relative performance of the proposed model compared to the benchmark model.

[1] In the comparative backtests performed on series of expansionary and recessionary GaR forecasts in Section V.B.3, the HAC variance estimator $\hat{s}_S^2$ is replaced by ordinary least squares variance estimators due to the uneven space of the time series.

IV. Data

The data sample of this study consists of GDP growth rates that are defined as the quarterly percentage changes of seasonally adjusted real GDP growth, compared to the same quarter in the previous year. We obtain the seasonally adjusted real GDP growth rates from the Organisation for Economic Co-operation and Development (OECD) database. To obtain a balanced panel of GDP growth rates, we restrict the sample to only comprise countries with at least 200 GDP growth rate observations. This selection criterion leaves us with a total of 25 OECD countries with GDP growth rates ranging from1961Q1 to 2020Q4 [2]. Due to unavailable GDP growth rates of CAN in 1961, the total number of GDP growth rates in the sample amounts to 5996 observations. To perform the out-of-sample performance study, the estimation window size is set equal to $\tau_1 = 160$ observations per country except for CAN in the first four forecasting iterations. Based on the size of the estimation window, the GaR forecasting performances of the models are evaluated based on the remaining $\tau_2 = 80$ GDP growth rates per country.

Table I reports the overall summary statistics of the full sample of GPD growth rates of all 25 countries. In addition, the table anticipates the results of the in-sample analysis by presenting the summary statistics of the 80%-sample, that excludes FIN, GRC, ISL, KOR and LUX as countries that exhibit most aberrant GDP growth dynamics. In addition to the summary statistics of the 100%-sample and the 80%-sample in Table I, Table XIV in Appendix B provides descriptive statistics of the GDP growth rates per country. From 1961Q1 to 2020Q4, the real economic activity of the 25 OECD countries increased by, on average, 3.07% compared to the same quarter one year earlier. Moreover, the standard deviation of the GDP growth rates in the 100%-sample equals 3.51%. Additionally, Table I shows that the aberrant countries experienced relatively high and volatile economic growth over the sample period. Motivated by the country-specific descriptive analysis in Appendix B, we consider omitting KOR as the main reason for the decrease in the average economic growth rate, while the lower economic volatility in the 80%-sample is particularly due to the exclusion of GRC, ISL and KOR. The convergence of the empirical quantiles after the exclusion of FIN, GRC, ISL, KOR, and LUX, indicates the aberrant behavior of the GDP growth dynamics of the countries. Finally, the Jarque-Bera tests reject the null hypothesis of normally distributed errors in both samples.

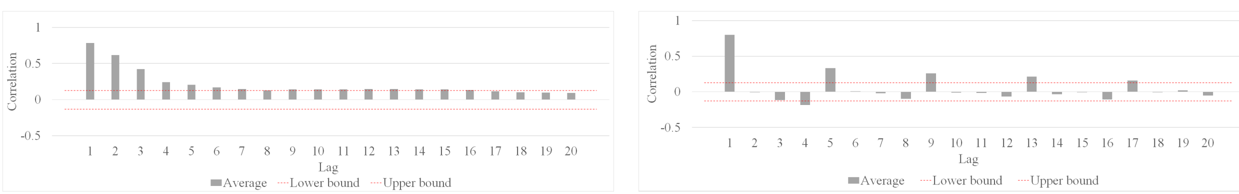
[2] The sample comprises the following countries: Australia (AUS), Austria (AUT), Belgium (BEL), Canada (CAN), Switzerland (CHE), Germany (DEU), Denmark (DNK), Spain (ESP), Finland (FIN), France (FRA), the U.K. (GBR), Greece (GRC), Ireland (IRL), Iceland (ISL), Italy (ITA), Japan (JPN), South Korea (KOR),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

Table I: **Summary statistics**

	Mean	Std. Dev.	5^{th} Perc.	25^{th} Perc.	Median	75^{th} Perc.	95^{th} Perc.	Normality test
100%-sample	3.07	3.51	-2.23	1.30	2.97	4.88	8.83	0.001***
80%-sampele	2.85	3.12	-1.85	1.30	2.85	4.47	7.62	0.001***

This table reports the summary statistics of the balanced panel of GDP growth rates of 25 OECD countries (100%-sample) and the panel of 20 OECD countries (80%-sample) after excluding the 20% most aberrant countries (FIN, GRC, ISL, KOR and LUX). The 100% and 80%-samples both range from 1961Q1 to 2020Q4 and consist of 5996 and 4796 observations, respectively. The GDP growth rates are in percentages. The column 'Normality test' reports the p-values of the Jarque-Bera test with a χ^2 distribution with 2 degrees of freedom and null hypothesis of normally distributed errors. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively.

Figure 1 shows the average autocorrelations and the average partial autocorrelations across the 25 OECD countries. Since the differences between the autocorrelations of the 100%-sample and the 80%-sample are negligible, we only report the average autocorrelations of the 100%-sample. Figure 1a indicates that, on average, current GDP growth rates are significantly autocorrelated with past GDP growth rates of at maximum three and half years ago. The partial autocorrelatios in Figure 1b indicate that removing the impact of intermediate correlations reveals the seasonality of the GDP growth rates, measured as percentage change compared to the same quarter in the previous year. The significant first-lag partial autocorrelation and the insignificant partial autocorrelations for most other lags motivate the choice for AR(1) processes to model the conditional mean of GDP growth. Figure 10 in Appendix B shows the autocorrelations and partial autocorrelations per country. The descriptive statistics and autocorrelations per country indicate that the GDP growth rates of most countries behave similarly. The GDP growth of GRC, JPN, ISL, IRL and KOR deviate from the other countries in some aspects.



(a) Average autocorrelations

(b) Average partial autocorrelations

Figure 1. This figure plots the average autocorrelations and average partial autocorrelations across all countries over the period 1961Q1-2020Q4. The dotted lines in the plots approximate the two standard error bounds and are calculated as $\pm\ 1.96/\sqrt{T}$. The GDP growth rates are measured at a quarterly frequency.

V. Results

We proceed by presenting our results, which are structured as follows. First, Section V.A discusses the results of the in-sample analyses of this research. More specifically, Section V.A presents estimations of the models on various samples to derive a set of countries that exhibit relatively homogeneous GDP growth dynamics. Throughout this study, the 100%-sample denotes the sample that includes all 25 OECD countries and the 80%-sample denotes the sample that includes 20 countries with relatively homogeneous GDP growth dynamics. Informed by the insights into the model dynamics from Section V.A, Section V.B discusses GaR forecasting performances of the single-regime and regime-switching models. In addition, based on comparisons of the relative forecasting performances of models estimated on the 100%-sample and 80%-sample, Section V.B concludes on the validity of the assumption of homogeneous model parameters. Throughout the results section, Appendices C - G provide the supplementary material to analyses performed on models estimated on the 100%-sample. Moreover, Appendix H gathers all supplementary material to the corresponding analyses performed on models estimated on the 80%-sample.

A. In-sample analysis

The results of the in-sample analysis are structured as follows. Section V.A.1 discusses the estimations of the single-regime and regime-switching models on the full sample. Based on the established dependence of the regime-switching framework on the extreme macroeconomic conditions during the outbreak of the Covid crisis, Section V.A.2 presents various model estimations to propose a set of countries that exhibit relatively similar dynamics in GDP growth. After determining the 80%-sample, Section V.A.2 provides insights into the dynamics of the models after excluding aberrant countries and the Covid crisis from the sample.

A.1. The full sample

This section discusses the estimations of the models based on the full sample comprising GDP growth rates of 25 OECD countries from 1961Q1 to 2020Q4. First, Table II presents the parameter estimates of the single-regime GARCH, EGARCH and GJR models. The estimates of ϕ_0 and ϕ_1 describe the AR(1) processes of the conditional mean. According to Table II, the parameter estimates of the conditional mean process are significant at the 1%-level in all single-regime models. Moreover, the high estimates of the autoregressive parameter ϕ_1 reflect the strong autocorrelation of GDP growth. Similar to the parameter estimates of the conditional mean processes, all estimates of the parameters of the conditional variance processes, ω, α, β and γ, are significant at the 1% level. To improve the interpretability of the parameters, we calculate the unconditional volatility and the persistence of shocks based on the parameter estimates of the GARCH model. From 1961 to 2020, the unconditional volatility of GDP growth across the countries equals 3.93%. Moreover, the persistence is close to one, indicating that the shocks to GDP growth impact the economic activity of countries for a long period into the future. The EGARCH and GJR models allow for asymmetric responses of the conditional variance to positive and negative GDP shocks. Since the estimate of the leverage parameter γ of the EGARCH model is significant and negative, a drop in GDP growth has a larger impact on the conditional variance than a rise of similar magnitude. In the GJR model, the estimates of the negative and positive shock parameters of the GJR model, α and γ, respectively, confirm the larger impact of negative macroeconomic news on conditional volatilities. To keep the tables in this study parsimonious, $\overline{\nu}$ denotes the average shape parameter of the conditional t-distributions of the GDP growth rates. Additionally, Table XV in Appendix C provides the conditional shape parameters per country. Except for NOR and FIN, all shape parameters ν_i of the single-regime models are significant at the 1% significance level. The estimates of ν_i indicate that the conditional distributions of the GDP growth rates of the countries exhibit excess kurtosis, which confirms the typical fat-tailed behavior of GDP growth. At last, Table II provides the Akaike information criterion (AIC) and Schwarz criterion (BIC) as in-sample goodness-of-fit statistics. Since the main focus of this research is the GaR forecasting ability of the models, we only report the in-sample goodness of fit statistics, but perform no formal tests.

Table II: In-sample estimations of single-regime models

	GARCH		EGARCH		GJR	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
ϕ_0	0.304***	(0.021)	0.273***	(0.021)	0.293***	(0.021)
ϕ_1	0.890***	(0.006)	0.897***	(0.006)	0.892***	(0.005)
ω	0.228***	(0.020)	−0.379***	(0.017)	0.223***	(0.020)
α	0.336***	(0.022)	0.638***	(0.027)	0.393***	(0.029)
β	0.606***	(0.018)	0.914***	(0.009)	0.609***	(0.018)
γ			−0.079***	(0.018)	0.274***	(0.028)
$\bar{\nu}$	5.225		4.692		5.347	
No. of Par.	30		31		31	
Log(L)	10544.515		10484.477		10537.867	
AIC	3.525		3.505		3.523	
BIC	3.558		3.400		3.558	

This table reports the maximum likelihood estimates of the single-regime GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models. $\bar{\nu}$ denotes the average shape parameter across the countries. Table XV in Appendix C provides the shape parameters for all individual countries. AIC is the Akaike information criterion calculated as $-2\log(L)/T + 2m/T$, where m denotes the number of parameters. BIC is the Schwarz criterion, calculated as $-2\log(L)/T + (m/T)\log(T)$. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The in-sample data of the 100%-sample consists of quarterly GDP growth rates from 1962Q1 to 2020Q4 for CAN and from 1961Q1 to 2020Q4 for the remaining 24 OECD countries.

Regime-switching model

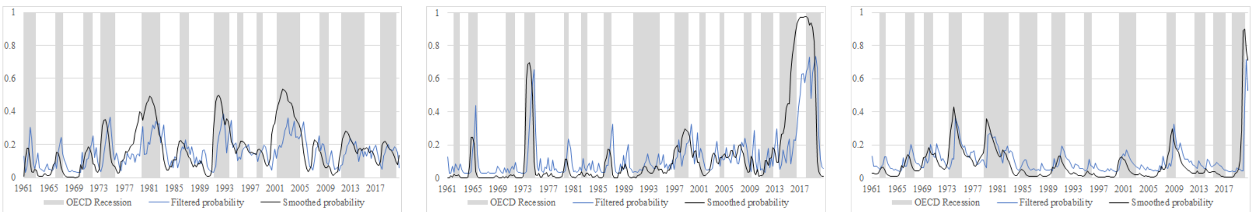
The regime-switching framework allows the model parameters to differ conditional on states of the economy. Table III presents the state-dependent parameter estimates of the MS-GARCH model. The estimates of the conditional mean parameters $\phi_0^{(k)}$ and $\phi_1^{(k)}$ are significant at the 1% significance level in both regimes $k = 1$ and $k = 2$. Moreover, the higher estimates of $\phi_0^{(1)}$ and $\phi_1^{(1)}$ indicate above average conditional GDP growth in the first regime. On the parameters of the conditional variance equation, Table III shows that all estimates of $\omega^{(k)}$, $\alpha^{(k)}$, $\beta^{(k)}$ and $\gamma^{(k)}$ are significant at the 10% level for both regimes. Moreover, apart from the estimates of $\alpha^{(2)}$ and $\beta^{(2)}$, all conditional variance parameters are significant at the 1% level. The estimates of the parameters confirm the existence of a low and a high volatility regime. In the first regime, the unconditional volatility of GDP growth rates amounts to 0.798%. The unconditional volatility in the second regime is higher and equals 4.630%. In addition, the higher shock persistence in the first regime indicates a longer-term influence of shocks on future GDP growth. Based on the lower conditional mean of GDP growth rates, the higher unconditional volatility and the lower persistence of shocks, we consider regime $k = 2$ as the recessionary state of the economy. Conversely, the expansionary regime $k = 1$ exhibits higher GDP growth rates, lower volatilities and higher shock persistence. The transition probabilities p and q of the MS-GARCH model are highly significant and specify that the expansionary regime is slightly more persistent than the recessionary regime. Moreover, the unconditional probability estimates of $\tilde{\pi}^{(k)}$ indicate that over the period 1961 until 2020, the unconditional probability that a country in the 100%-sample experiences a recession, equals 12.9%. Moreover, the more extreme behavior of GDP growth during recessions drives the higher excess kurtosis of the state-dependent conditional distribution, indicated by the lower average shape parameter $\overline{\nu}$. Table XVI in Appendix C provides the shape parameters of the conditional distributions of GDP growth per country.

Table III: In-sample estimation of the MS-GARCH model

	Expansionary regime ($k = 1$)		Recessionary regime ($k = 2$)	
	Coefficient	Std. Error	Coefficient	Std. Error
ϕ_0	0.287***	(0.022)	0.203***	(0.042)
ϕ_1	0.901***	(0.006)	0.783***	(0.020)
ω	0.058***	(0.012)	8.252***	(3.488)
α	0.300***	(0.024)	0.191*	(0.112)
β	0.609***	(0.020)	0.424**	(0.209)
p	0.972***	(0.004)		
q			0.811***	(0.034)
$\overline{\nu}$	16.413		4.483	
σ	0.798		4.630	
ρ	0.909		0.615	
$\tilde{\pi}$	0.871		0.129	
No. of Par.	62			
Log(L)	10460.500			
AIC	3.511			
BIC	3.581			

This table reports the maximum likelihood estimates of the MS-GARCH(1,1) model. $\overline{\nu}^{(k)}$ denote the average shape parameter across the countries per regime k. Table XVI in Appendix C provides the state-dependent shape parameters for the countries individually. $\sigma^{(k)} = (\omega^{(k)}/(1 - \alpha^{(k)} - \beta^{(k)}))^{1/2}$ denotes the unconditional volatility per regime. Moreover, $\rho^{(k)} = \alpha^{(k)} + \beta^{(k)}$ are the persistence of shocks in regime k. $\tilde{\pi}^{(k)}$ denote the unconditional probabilities of a country being in regime k and are given by $\tilde{\pi}^{(1)} = (1 - q)/(2 - p - q)$ and $\tilde{\pi}^{(2)} = (1 - p)/(2 - p - q)$ for states $k = 1$ and $k = 2$, respectively. AIC is the Akaike information criterion calculated as $-2\log(L)/T + 2m/T$, where m denotes the number of parameters. BIC is the Schwarz criterion, calculated as $-2\log(L)/T + (m/T)\log(T)$. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The in-sample data of the 100%-sample consists of quarterly GDP growth rates from 1962Q1 to 2020Q4 for CAN and from 1961Q1 to 2020Q4 for the remaining 24 OECD countries.

Figure 2 below plots the estimated filtered and smoothed probabilities of the recessionary regime from 1961Q1 to 2020Q4 for a selection of countries. More specifically, Figures 2a, 2b, and 2c present the filtered and smoothed recession probabilities for DEU, JPN, and USA, respectively. The recession probabilities for the remaining 22 OECD countries can be observed from Figure 11 in Appendix D. The concurrence of the increases in recession probabilities with the highlighted OECD recession periods in the figures confirm the recessionary character of the second regime. Furthermore, the relatively low recession probabilities in almost all countries prior to 2020 mark the high dependence of the recessionary regime on the recent Covid crisis. Although the recession probabilities rise during other economic crises, the filtered and smoothed recession probabilities remain inferior to the expansionary probabilities. For instance, during the oil crisis in 1973, Reagan's recession in the early 1980s and the Global Financial Crisis in 2008, the recession probabilities in the USA increase, but remain below 50%, as shown by Figure 2c. Due to the dependence of the recession regime on the Covid crisis, the MS-GARCH model estimated on the 100%-sample from 1961Q1 to 2020Q4 fails to capture some influential financial crises, such as the banking crisis in JPN in the early 1990s and the economic turmoil in GRC in 2009.



(a) DEU **(b)** JPN **(c)** USA

Figure 2. This figure plots the filtered (in blue) and smoothed (in black) recession probabilities of DEU, JPN, and USA. The filtered and smoothed recession probabilities range from 1961Q1 to 2020Q4. OECD recession periods are highlighted in gray.

Dependence of the model estimations on the Covid-crisis

Figure 3 below quantifies the impact of the extreme macroeconomic conditions in 2020 on the estimations of the single-regime and regime-switching models on the full sample. More specifically, Figure 3 plots the sum of the log likelihoods per year, aggregated over countries, for all models. The negative peaks of the aggregated log likelihoods in 2020 reflect the dependence of the single-regime and regime-switching models on the extreme macroeconomic circumstances during the Covid-outbreak in 2020. This dependence implies that discrepancies in the macroeconomic behavior of countries during the pandemic have large impacts on the model estimations.

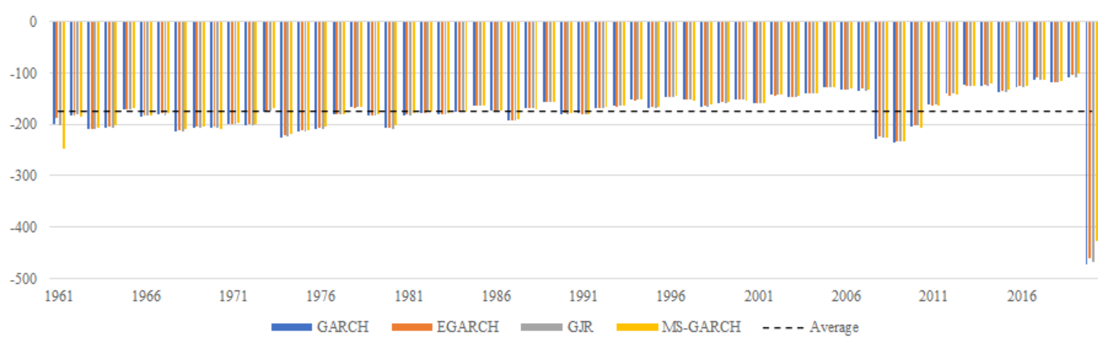


Figure 3. Sum of log likelihood over the 25 OECD countries per year. Due to missing GDP growth rates of CAN in 1961, the sum of likelihoods of 1961 are multiplied by a factor 25/24.

The estimations of the models on the full data sample provide insights into the dynamics of the single-regime and regime-switching models. More specifically, this subsection reveals the high impact of the extreme macroeconomic conditions in 2020 on the model estimations. Section V.A.2 proposes a sample of countries that exhibit relatively homogeneous GDP growth dynamics. By comparing the GaR forecasting performances of the models estimated on samples that include and exclude the countries with aberrant GDP growth dynamics, the forecasting analysis validates the assumption of homogeneous model parameters. The majority of forecasts in the forecasting analysis is generated from models estimated based on estimation windows prior 2020. To avoid that GDP dynamics during the Covid crisis determine the selection of aberrant countries, Section V.A.2 excludes the GDP growth data in 2020 to propose a sample of countries with relatively homogeneous GDP growth dynamics.

A.2. Excluding aberrant countries

To select countries that exhibit aberrant GDP growth dynamics, we re-estimate the single-regime and regime-switching models on a reduced sample period from 1961Q1 to 2019Q4. The GDP growth rates of CAN range from 1962Q1 to 2019Q4 due to missing observations in 1961. Tables XVII and XVIII in Appendix E report the estimations of the single-regime and regime-switching models. More importantly, Figure 4 below plots the log likelihood per country, aggregated over the sample period from 1961Q1 to 2019Q4. Based on the aggregated log likelihoods per country, the 80%-sample excludes KOR, GRC, ISL, FIN, and LUX as countries that exhibit most aberrant GDP growth dynamics. To validate the homogeneity of the GDP dynamics in the constructed 80%-sample, we recurrently omit the most aberrant country from the sample and compare the re-estimations of the models with the estimations based on the 80%-sample. Tables IV and V in the remainder of this section provide the estimations of the single-regime and regime-switching models, respectively, based on the 80%-sample, after excluding GDP growth data from 2020. Appendix E elaborates by means of estimations on samples of reduced numbers of countries on the impact of omitting more than five countries to the model estimations. The convergence of the parameter estimates after omitting KOR, GRC, ISL, FIN and LUX, as derived in Appendix E, proves the increased homogeneity of GDP growth dynamics of the countries in the 80%-sample.

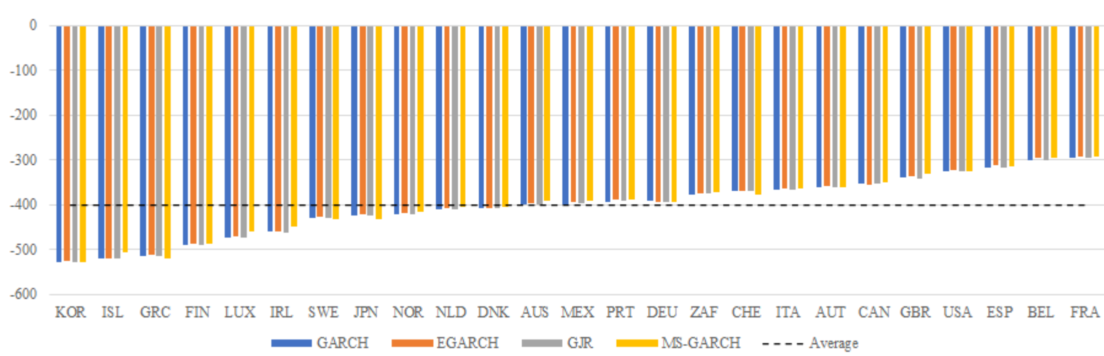


Figure 4. Sum of log likelihood over the period 1961Q1 - 2019Q4 per country. Due to missing GDP growth observations in 1961, the sum of the log likelihood of CAN is multiplied by a factor 240/236.

The remainder of this section discusses the dynamics of the single-regime and regime-switching models after omitting KOR, GRC, ISL, FIN, and LUX as countries with most aberrant GDP dynamics. In addition, since the majority of forecasts in the forecasting analysis is generated from models that are not estimated on GDP growth data from 2020, we consider it as informative to exclude the Covid pandemic from the estimations in this section, such that the estimation windows range from 1961Q1 to 2019Q4. First, Table IV below presents the parameter estimates of the single-regime GARCH, EGARCH and GJR models on the reduced sample. Similar to the estimations of the single-regime models on the full sample, in Table II, all parameters are significant at the 1% significance level. As observed in the descriptive analysis on the level of sample countries in Appendix B, the GDP growth in KOR is considerably higher than in the other countries. Therefore, we identify omitting KOR as one of the causes of the decreases of ϕ_0 in the single-regime model estimations. Moreover, the decreases of the constant variance terms ω in the GARCH and GJR models are driven by the exclusion of the volatile economies GRC, ISL, and KOR, as well as the relatively volatile macroeconomic circumstances in 2020. Another corollary of the exclusion of the Covid-pandemic from the estimation window is the more symmetric responsive behavior of GDP growth to positive and negative shocks, as indicated by the increase of the leverage parameter γ in the EGARCH model and the convergence of α and γ in the GJR model. Finally, due to the exclusion of countries with aberrant GDP dynamics and the Covid crisis, the average shape parameters decrease, indicating lower excess kurtosis. In addition to the average shape parameters of the single-regime models, Appendix Section H.A of Appendix H provides the shape parameters per country in the 80%-sample.

Table IV: **In-sample estimations of single-regime models**

	GARCH		EGARCH		GJR	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
ϕ_0	0.291***	(0.022)	0.259***	(0.023)	0.282***	(0.023)
ϕ_1	0.891***	(0.006)	0.899***	(0.006)	0.893***	(0.006)
ω	0.125***	(0.014)	−0.376***	(0.019)	0.123***	(0.013)
α	0.280***	(0.021)	0.523***	(0.028)	0.317***	(0.028)
β	0.651***	(0.018)	0.933***	(0.009)	0.654***	(0.018)
γ			−0.056***	(0.017)	0.241***	(0.025)
$\overline{\nu}$	27.822		36.104		38.247	
No. of Par.	25		26		26	
Log(L)	7523.538		7484.139		7520.550	
AIC	3.199		3.182		3.200	
BIC	3.233		3.218		3.233	

This table reports the maximum likelihood estimates of the single-regime GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models. $\overline{\nu}$ denotes the average shape parameter across the countries. Table XXI in Appendix Section H.A of Appendix E provides the shape parameters for all individual countries. AIC is the Akaike information criterion calculated as $-2\log(L)/T + 2m/T$, where m denotes the number of parameters. BIC is the Schwarz criterion, calculated as $-2\log(L)/T+(m/T)\log(T)$. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The in-sample data of the 100%-sample consists of quarterly GDP growth rates from 1962Q1 to 2019Q4 for CAN and from 1961Q1 to 2019Q4 for the remaining 19 OECD countries.

Regime-switching model

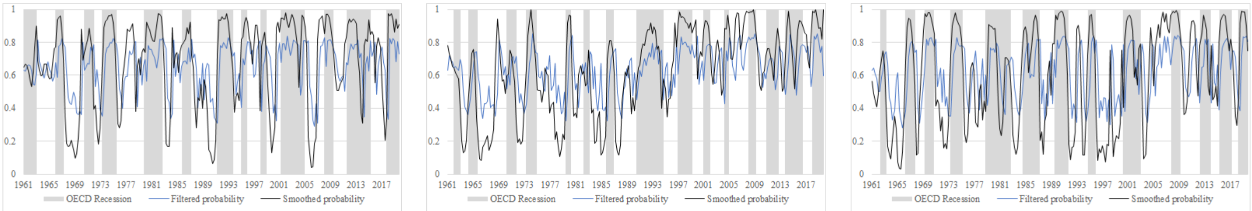
Below, Table V provides the estimation of the MS-GARCH model on the reduced sample of 20 OECD countries with data from 1961Q1 to 2019Q4. From the parameter estimates, we observe a considerable impact of the exclusion of the countries FIN, GRC, ISL, KOR and LUX and the GDP growth data from 2020. In particular, the exclusion of the extreme macroeconomic conditions during 2020 causes the recessionary regime ($k = 2$) to comprise a variety of financial crises. This is in contrast with the estimation of the MS-GARCH model on the full sample, in Table III, where the dynamics of the recessionary regime are mainly driven by the Covid crisis. The inclusion of more crises in the recessionary regime - and the simultaneous exclusion of these crises in the expansionary regime - results into a divergence of the constants of the conditional mean processes, $\phi_0^{(k)}$. In different words, the conditional mean of GDP growth in the expansionary regime increases, while the conditional mean of GDP growth in the recessionary regime decreases. On the contrary, the autoregressive character of the regimes, denoted by $\phi_1^{(k)}$, converges. As the recessionary regime comprises less volatile crises after the exclusion of the Covid crisis, the unconditional volatility of the recessionary regime decreases. Simultaneously, the unconditional volatility of the expansionary regime lowers due to the shifting of financial crises from the expansionary to the recessionary regime. Moreover, the reallocation of recessions results into higher (lower) persistence of the recessionary (expansionary) regime. Compared to the estimation of the MS-GARCH model on the full sample, the unconditional probability of a recession increases from 12.9% to 63.3%. As a final remark, the observation of smaller tails of the conditional distribution of GDP growth in the estimation of the single-regime models especially emerges in the recessionary regime of the MS-GARCH estimation. Table XXII in Appendix Section H.A of Appendix H provides an overview of shape parameters of the conditional distribution of GDP growth per regime and per county in the 80%-sample.

Table V: In-sample estimation of the MS-GARCH model

	Expansionary regime ($k = 1$)		Recessionary regime ($k = 2$)	
	Coefficient	Std. Error	Coefficient	Std. Error
ϕ_0	0.967***	(0.051)	0.148***	(0.029)
ϕ_1	0.851***	(0.011)	0.842***	(0.010)
ω	0.000	(0.022)	0.035**	(0.015)
α	0.371***	(0.055)	0.277***	(0.030)
β	0.543***	(0.046)	0.681***	(0.027)
p	0.748***	(0.028)		
q			0.854***	(0.020)
$\overline{\nu}$	16.230		9.455	
σ	0.000		0.913	
ρ	0.914		0.958	
$\tilde{\pi}$	0.367		0.633	
No. of Par.	52			
Log(L)	7447.976			
AIC	3.178			
BIC	3.249			

This table reports the maximum likelihood estimates of the MS-GARCH(1,1) model. $\overline{\nu}^{(k)}$ denote the average shape parameter across the countries per regime k. Table XXII in Appendix Section H.A of Appendix H provides the state-dependent shape parameters for the countries individually. $\sigma^{(k)} = (\omega^{(k)}/(1 - \alpha^{(k)} - \beta^{(k)}))^{1/2}$ denotes the unconditional volatility per regime. Moreover, $\rho^{(k)} = \alpha^{(k)} + \beta^{(k)}$ are the persistence of shocks in regime k. $\tilde{\pi}^{(k)}$ denote the unconditional probabilities of a country being in regime k and are given by $\tilde{\pi}^{(1)} = (1-q)/(2-p-q)$ and $\tilde{\pi}^{(2)} = (1-p)/(2-p-q)$ for states $k = 1$ and $k = 2$, respectively. AIC is the Akaike information criterion calculated as $-2\log(L)/T + 2m/T$, where m denotes the number of parameters. BIC is the Schwarz criterion, calculated as $-2\log(L)/T + (m/T)\log(T)$. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The in-sample data of the 100%-sample consists of quarterly GDP growth rates from 1962Q1 to 2019Q4 for CAN and from 1961Q1 to 2019Q4 for the remaining 19 OECD countries.

Below, Figure 15 provides insights in the capabilities of the MS-GARCH model that remain latent if the Covid-crisis is included in the estimation window. Figures 5a, 5b, and 5c plot the filtered and smoothed recession probabilities over the period 1961Q1 to 2019Q4 for DEU, JPN, and USA, respectively. In addition, Figure 15 in Appendix Section H.B of Appendix H provides the recession probabilities for the 17 remaining OECD countries in the 80%-sample. Compared to the recession probabilities observed in Section V.A.1, the less extreme characteristics of the recessionary regime allow for more frequent and larger movements of the state probabilities, indicating a higher awareness of the MS-GARCH model of deteriorating macroeconomic conditions. More specifically, the exclusion of the Covid crisis from the estimation window enables the recessionary regime to comprise a variety of financial crises prior to 2020. For instance, contrarily to the estimation of the MS-GARCH model in Section V.A.2, the recession probabilities in Figure 5b recognize the Japanese Banking Crisis in the 1990s. Moreover, the ability to distinguish between the European debt crisis in 2011-2012 and the European refugee crisis in 2014-2015 shows the increased precision of the MS-GARCH model. The high awareness of deteriorating macroeconomic conditions makes the MS-GARCH model more sensitive to consider periods of economic turmoil as recessions rather than the OECD. For instance, the filtered and smoothed recession probabilities of DEU exceed 80% in 1977, while the OECD neglects the East German coffee crisis as a recession.



(a) DEU **(b)** JPN **(c)** USA

Figure 5. This figure plots the filtered (in blue) and smoothed (in black) recession probabilities of DEU, JPN, and USA. The filtered and smoothed recession probabilities range from 1961Q1 to 2019Q4. OECD recession periods are highlighted in gray.

To validate the assumption of homogeneous model parameters in Section V.B, this section proposes the 80%-sample as a set of countries that exhibit relatively homogeneous GDP growth dynamics. Based on the aggregated log likelihood over the evaluation window, the 80%-sample excludes FIN, GRC, ISL, KOR, and LUX. The estimations of the single-regime and regime-switching models in Section V.A.1 demonstrate the high impact of the extreme macroeconomic conditions in 2020 on the estimations of the models. After excluding the Covid crisis from the data sample, the in-sample estimations of the single-regime and regime-switching models provide insights into the characteristics and capabilities of the models in the majority of the forecasting exercise. For instance, driven by a high awareness of deteriorating macroeconomic conditions, the MS-GARCH model is able to capture a large variety of financial crises.

B. Forecasting analysis

This section presents the GaR forecasting performances of the single-regime and regime-switching models. By comparing the prediction power of the models on various forecasting horizons, this section provides insights into the additional prediction power for downside macroeconomic risk that allowing for regime-dependence entails. The ou-of-sample window to evaluate the GaR forecasts ranges from 2001Q1 to 2020Q4. First, Section V.B.1 presents the short-horizon forecasting performances of the models based on one-quarter ahead GaR forecasts. Second, Section V.B.2 discusses the long-horizon GaR forecasting performances based on two-quarter ahead, three-quarter ahead, and one-year ahead forecasts. Additionally, Section V.B.2 evaluates the validity of the assumption of homogeneous model parameters in this study based on the observed short-horizon and long-horizon GaR forecasts. Third, to derive the economic significance of the findings in Sections V.B.1 and V.B.2, Section V.B.3 compares the GaR forecasting performances of the models during expansionary and recessionary periods. Sections V.B.1 and V.B.2 presents the unconditional coverage ratios of the GaR forecasts for interpretation purposes first, then discuss the results of the backtesting framework, and finally analyze the patterns of the GaR forecasts more deeply to explain the outcomes of the backtesting framework.

B.1. Short-horizon GaR forecasting performance

To obtain initial insights in the short-horizon GaR forecasting performances of the single-regime and regime-switching models, Table VI provides the unconditional coverage ratios of the one-quarter ahead 95% GaR forecasts of the models. Panels (a) and (b) represent estimations of the models on estimation windows from the 100%-sample and 80%-sample, respectively. The unconditional coverage ratios denote the fractions of GaR overshootings over the evaluation window, across the countries in the sample. As the unconditional coverage ratios of the GJR model are closest to 95% in both panels, the GJR model predicts one-quarter ahead GaR most accurately in terms of unconditional coverage. Conversely, on both samples, the EGARCH model produces least accurate one-quarter ahead GaR forecasts. The exclusion of the countries with most aberrant GDP growth dynamics has a slightly negative impact on the accuracy of the GARCH and MS-GARCH model. By contrast, the accuracy of the EGARCH model increases, while the accuracy of the GJR model remains similar. Since the unconditional coverage overlooks the magnitude of GaR overshootings, we report the unconditional coverage ratios, but draw no formal conclusions based on the measure.

Table VI: Unconditional coverage ratios of short-horizon GaR forecasts

	(a) 100%-sample	(b) 80%-sample
GARCH	0.946	0.940
EGARCH	0.896	0.914
GJR	0.949	0.949
MS-GARCH	0.959	0.961

This table reports the unconditional coverage ratios of the one-quarter ahead 95% GaR forecasts generated from the GARCH(1,1), EGARCH(1,1), GJR-GARCH(1,1) en MS-GARCH(1,1) models. The unconditional coverage ratio is calculated as $1 - \left(\sum_{i=1}^{n} \sum_{t=\tau_1+H}^{\tau_1+\tau_2} \mathbb{I}[Y_{i,t} < GaR_{i,t|t-1}(c)] \;/\; (n * (\tau_2 - H)) \right)$. The out-of-sample evaluation window ranges from 2001Q1 to 2020Q4.

To formalize the evaluation of the short-horizon forecasting performances of the models, we implement a two-fold backtesting framework. As a first component, Table VII provides the p-values of the unconditional and conditional calibration tests. Similar to Table VI, panels (a) and (b) of the table represent the GaR forecasts generated from the models estimated on estimation windows from the 100%-sample and the 80%-sample, respectively. At the significance level of 5%, neither the unconditional calibration test nor the conditional calibration test rejects the correctness of the GaR forecasts of the GARCH, GJR and MS-GARCH models. By contrast, both tests reject correct calibration of the GaR forecasts of the EGARCH model. At the significance level of 5%, the conclusions on the correctness of the GaR forecasts of the models are independent of the sample, i.e. the inclusion of 25 countries or 20 countries in the analysis. Nevertheless, the relatively large decrease of the p-values of the GARCH model indicates dependence of the GaR forecasting performance of the GARCH model on the inclusion of FIN, GRC, ISL KOR and LUX in the analysis.

Table VII: **Calibration tests on short-horizon GaR forecasts**

	(a) 100%-sample		(b) 80%-sample	
	Unconditional	Conditional	Unconditional	Conditional
GARCH	0.374	0.320	0.090	0.053
EGARCH	0.002	0.000	0.005	0.000
GJR	0.649	0.611	0.661	0.618
MS-GARCH	0.098	0.070	0.091	0.061

This table reports the p-values of the unconditional and conditional calibration tests. The calibration tests are performed on the one-quarter ahead 95% GaR forecasts generated from the GARCH(1,1), EGARCH(1,1), GJR-GARCH(1,1) and MS-GARCH(1,1) models. In panel (a), the models are estimated on moving estimation windows from the 100%-sample that comprises 25 OECD countries. In panel (b), the models are estimated on moving estimation windows from the 80%-sample that includes 20 OECD countries with relatively homogeneous GDP growth dynamics. The test statistics of the unconditional and conditional calibration tests are provided by Equations (17) and (18) on page 14. The null hypotheses of the unconditional and conditional calibration tests imply correct unconditional and correct conditional calibration of the GaR forecasts, respectively. Under the null hypotheses of the unconditional and conditional calibration tests, the test-statistics follow a standard normal and a chi-squared(1) distribution. The out-of-sample evaluation window ranges from 2001Q1 to 2020Q4.

To determine the relative GaR forecasting performances of the models, we perform comparative backtests to measure the one-quarter ahead GaR forecasting performances of the proposed GJR and MS-GARCH models relative to the benchmark GARCH, EGARCH and GJR models. Table VIII below presents the CDF of the test-statistics of the comparative backtest, $\Phi(L_3)$. Recall that the null hypothesis H_0^+ (H_0^-) that the GaR forecasts of the proposed models are at most (least) as good as the GaR forecasts of the benchmark models, is rejected if $1 - \Phi(L_3) \leq \eta$ ($\Phi(L_3) \leq \eta$). Although the outperformance of the GJR model relative to the GARCH model is not significant, the values below 0.5 indicate that the GJR model is the best performing single-regime model in forecasting short-horizon GaR in the analyses on both samples. Moreover, the comparative backtest provides statistical evidence for the significant outperformance of the MS-GARCH model in forecasting short-horizon GaR. Consequently, the backtesting framework assigns the MS-GARCH model to the green zone. The conclusion that the MS-GARCH model significantly outperforms single-regime models in forecasting one-quarter ahead GaR is independent of the inclusion of 20 or 25 OECD countries in the sample.

Besides the outperformance of the MS-GARCH model, the results in Tables VII and VIII indicate the inferior performance of the EGARCH model in forecasting one-quarter ahead GaR. The exponential form of the EGARCH model generates higher conditional volatilities that, in turn, result into underestimated GaR. Especially during periods of economic resilience, the distance between the realized GDP growth rates and the GaR forecasts of the EGARCH model increases sharply. Besides the frequency and magnitude of overshootings, the scoring functions in the backtesting framework take the extent of GaR underestimation into account, which explains the underperformance of the EGARCH model. For the analyses on the 100%-sample and the 80%-sample, respectively, Figure 12 in Appendix F and Figure 17 in Appendix Section H.D of Appendix H plot the extreme GaR forecasts of the EGARCH model alongside the GaR forecasts of the other models for DEU and USA.

Table VIII: Comparative backtest on short-horizon GaR forecasts

		(a) 100%-sample		(b) 80%-sample	
		GJR	MS-GARCH	GJR	MS-GARCH
Benchmark model	GARCH	0.356	0.016	0.346	0.010
	EGARCH	0.009	0.000	0.008	0.000
	GJR		0.020		0.020

This table reports the CDF of the test-statistic of the comparative backtest, $\Phi(L_3)$. The comparative backtest is performed on one-quarter ahead 95% GaR forecasts from the single-regime and regime-switching models. The null hypothesis that the GaR forecasts of the proposed models are at most (least) as good as the GaR forecasts of the benchmark models is rejected if $1-\Phi(L_3) \leq \eta$ ($\Phi(L_3) \leq \eta$). We test the GaR forecasting performances of the proposed GJR-GARCH(1,1) and MS-GARCH(1,1) models against the forecasting performances of the benchmark GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models. In panel (a), the models are estimated on moving estimation windows from the 80%-sample that comprises 25 OECD countries. In panel (b), the models are estimated on moving estimation windows from the 100%-sample that includes 20 OECD countries with relatively homogeneous GDP growth dynamics. The test statistic of the comparative calibration test $\Phi(L_3)$ is provided by Equation (21) on page 16 and follows a standard normal distribution. The out-of-sample evaluation window ranges from 2001Q1 to 2020Q4.

As the EGARCH model underperforms the other single-regime models, the remainder of this section forms an in-depth analysis on the relative performance of the MS-GARCH model compared to the GARCH and GJR models. Due to the high degree of similarity between GaR forecasts from models estimated on the 100%-sample and the 80%-sample, this section only reports figures based on models estimated on the 100%-sample. For completeness, Appendix Sections H.C and H.D in Appendix H provide the corresponding figures obtained from models estimated on the 80%-sample. Figure 6 below shows the course of the comparative scoring function, aggregated over the 25 sample countries, over the out-of-sample period. The comparative scoring function takes the frequency and magnitude of GaR overshootings into account, as well as the distance between realized GDP growth rates and GaR forecasts during quarters without overshootings. The graph shows that the scoring functions of the models peak simultaneously at major financial crises due to GaR overshootings. A deeper look on the figure reveals that the outperformance of the MS-GARCH model originates from the quarters 2008Q4, 2016Q1, 2020Q1 and 2020Q2, in particular. Conversely, the MS-GARCH model underperforms the single-regime GARCH and GJR in 2009Q1 and 2020Q3.

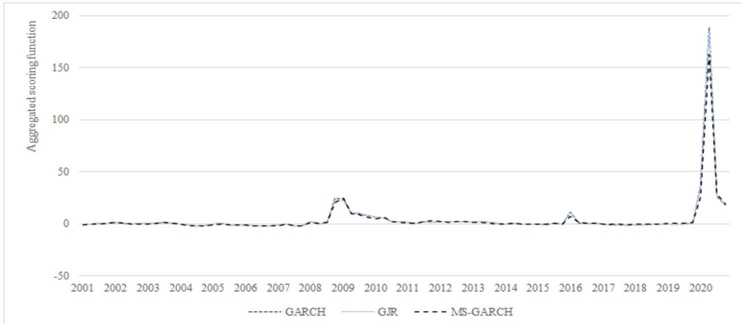


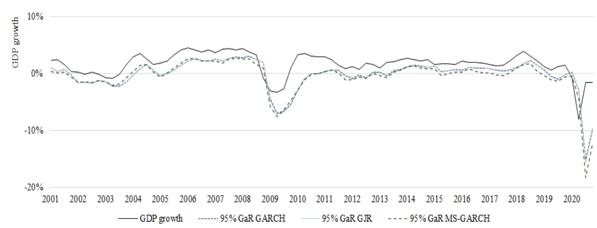
Figure 6. This figure plots the comparative scoring function of one-quarter ahead GaR forecasts, aggregated over the countries in the sample. The aggregated comparative scoring function is given by $\sum_{i=1}^{n} S_{i,t} = \sum_{i=1}^{n} \big(\mathbb{I}[Y_{i,t} < GaR_{i,t|t-1}(c)] + c - 1\big)\big(GaR_{i,t|t-1}(c)\big)\big) - \mathbb{I}[Y_{i,t|t-1} < GaR_{i,t|t-1}(c)]\big(Y_{i,t}\big)$, where the expression of $S_{i,t}$ is obtained from Equation (20). The dotted dark blue, solid light blue and dashed black lines in the figures represent the aggregated scoring functions of the GARCH, GJR and MS-GARCH models, respectively. The models are estimated on estimation windows from the 100%-sample. The out-of-sample period ranges from 2001Q1 to 2020Q4. The sums of the scoring functions over the out-of-sample period equal 382.252, 383.351 and 341.788 for the GARCH, GJR and MS-GARCH model, respectively.

The drivers of the outperformance and underperformance of the MS-GARCH model during specific quarters are further detailed by means of plots of the one-quarter ahead 95% GaR forecasts. Below, Figure 7 plots the realized GDP growth rates of CHE, DEU, IRL, and USA, alongside the one-quarter ahead GaR forecasts of the GARCH, GJR and MS-GARCH models. In addition to Figure 7, Appendix F provides the corresponding figures for the remaining 21 OECD countries in the 100%-sample. The patterns of GaR forecasts indicate two main reasons for the outperformance of the MS-GARCH model in forecasting one-quarter ahead GaR. First, the ability to switch to a recessionary regime when recession probabilities increase, enables the MS-GARCH model to anticipate turmoils of GDP growth earlier. As a result, the GaR forecasts of the MS-GARCH model decrease prior to the peak of the crisis, as does the magnitude of the successive GaR overshooting. Therefore, the early detection of signals of macroeconomic turmoils by the MS-GARCH model explains the outperformance of the model during the quarters prior to financial recessions in 2008 and 2020. For instance, the major decrease of the GaR forecasts of the MS-GARCH model for CHE during 2008Q3-2009Q1, prior to the height of the Great Financial Crisis in 2009, results into a smaller magnitude of the GaR overshooting in 2009Q2. Similar patterns occur in 2020, as the MS-GARCH model seems to adjust GaR forecasts more downwards than single-regime models based on falling GDP growth rates in 2020Q1.

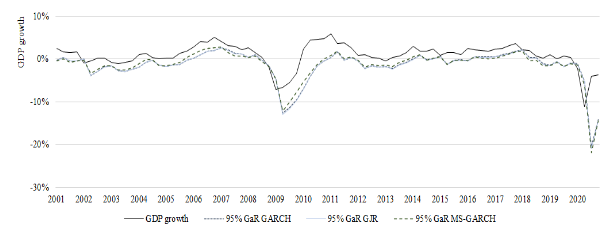
Second, the recessionary regime of the MS-GARCH model tempers GaR forecasts during extremely bullish periods of high macroeconomic volatility. From Figure 7c, we observe extraordinarily high and volatile economic growth in IRL from 2014Q1 to 2015Q4. Low corporate tax rates and subsequent relocations of large numbers of multinationals to IRL drive the expansion during this period. Due to high macroeconomic volatilities, the recession probabilities in the MS-GARCH model rise, which causes the GaR forecasts of the MS-GARCH model to decrease. In the single-regime models, the economic expansion results into rising GaR forecasts. Due to the lower GaR forecasts of the MS-GARCH model during the economic expansion in IRL, the GaR overshooting in

2016Q1 is of a substantially smaller magnitude. As the outperformance of the MS-GARCH model in 2016Q1 mainly originates from the GaR overshooting in IRL, the tempering of GaR forecasts during extremely bullish macroeconomic periods proves a notable advantage of the MS-GARCH model over the single-regime models.

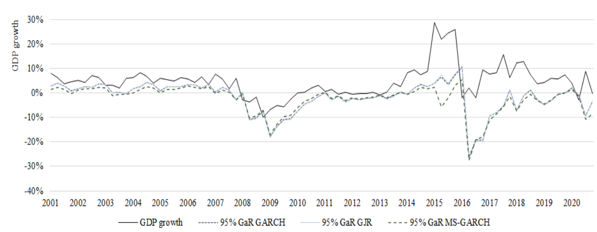
The overestimation of peaks of major financial crises by the MS-GARCH model partially compensates the relative outperformance. The MS-GARCH underperforms single-regime models during the quarters 2009Q1 and 2020Q3. Worldwide, 2009Q1 and 2020Q3 are considered as quarters following the beginning of major financial crises. Recession probabilities close to one cause the MS-GARCH model to produce lower GaR forecasts than single-regime models. If the macroeconomic outlook improves during these quarters, rather than deteriorates, the overestimation of the crisis by the MS-GARCH model enlarges the distances between the realized GDP growth rates and the GaR forecasts. For instance, the rapid recovery of economies in 2020Q3, following the Covid-outbreak in 2020Q1-Q2, explains the underperformance of the MS-GARCH model in this quarter, as observed for almost all countries in the sample.



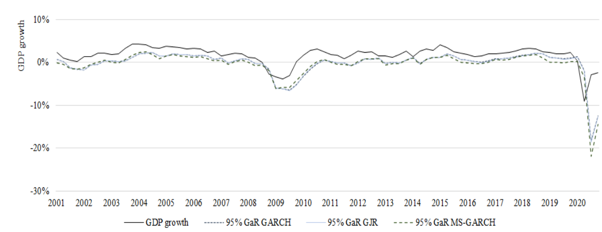
(a) CHE



(b) DEU



(c) IRL



(d) USA

Figure 7. This figure plots the realized GDP growth rates of CHE, DEU, IRL and USA alongside the one-quarter ahead 95% GaR forecasts of the GARCH(1,1), GJR-GARCH(1,1) and MS-GARCH(1,1) models. In the figures, the solid black lines plot the realized GDP growth rates of the countries. Moreover, the dotted dark blue, solid light blue and dashed black lines reflect the one-quarter ahead 95% GaR forecasts of the GARCH, GJR and MS-GARCH models, respectively. The models are estimated on estimation windows from the 100%-sample that comprises 25 OECD countries. The out-of-sample periods range from 2001Q1 to 2020Q4.

The short-horizon GaR forecasting exercise provides statistically significant evidence for the outperformance of the MS-GARCH in forecasting one-quarter ahead GaR. Moreover, the findings in this section are independent of the inclusion of 25 or 20 OECD countries in the analysis. The main drivers of the outperformance of the MS-GARCH model are the early detection of macroeconomic turmoils and the tempering of GaR forecasts during extremely bullish circumstances. Conversely, the overestimation of peaks of major financial crises in 2009Q1 and 2020Q3 partially compensates the outperformance of the MS-GARCH model. To obtain a deliberate opinion on the relative short-term and long-term capabilities of the MS-GARCH model, Section V.B.2 expands the forecasting horizon of the forecasting analysis.

B.2. Long-horizon GaR forecasting performance

In correspondence with Section V.B, this section first presents the unconditional coverage ratios of the GaR forecasts for interpretation purposes. Below, Table IX presents the unconditional coverage ratios of the two-quarter ahead (2Q-ahead), three-quarter ahead (3Q-ahead), and one-year ahead (1Y-ahead) GaR forecasts of the single-regime and regime-switching models. Similar to the tables in the previous section, panels (a) and (b) contain statistics obtained by estimating models on estimation windows from the 100%-sample and the 80%-sample, respectively. Compared to the unconditional coverage ratios of the short-horizon GaR forecasts in Table VI, the uncertainty of forecasting GaR over multiple quarters has a negative impact on the accuracy of the models. Moreover, the divergence of the unconditional coverage ratios from 95% as the forecasting horizon increases, demonstrates that the decreasing prediction accuracy of GaR forecasts of single-regime models. Conversely, the one-year ahead GaR forecasts of the MS-GARCH model are most accurate in terms of unconditional coverage, despite the higher extent of uncertainty on the macroeconomic circumstances between the forecasting quarter and the forecasted quarter. Furthermore, the exclusion of countries with most aberrant GDP growth dynamics in the 80%-sample has a negative impact on the long-horizon prediction accuracies of the GARCH and MS-GARCH models, a positive impact on the accuracy of the GJR model and an ambiguous impact on the accuracy of the EGARCH model.

Table IX: Unconditional coverage ratios of long-horizon GaR forecasts

	(a) 100%-sample			(b) 80%-sample		
	2Q-ahead	3Q-ahead	1Y-ahead	2Q-ahead	3Q-ahead	1Y-ahead
GARCH	0.933	0.926	0.917	0.930	0.922	0.914
EGARCH	0.885	0.879	0.801	0.890	0.879	0.790
GJR	0.933	0.926	0.919	0.936	0.928	0.920
MS-GARCH	0.932	0.934	0.937	0.932	0.932	0.934

This table reports the unconditional coverage ratios of the two-quarter ahead, three-quarter ahead and one-year ahead 95% GaR forecasts generated from the GARCH(1,1), EGARCH(1,1), GJR-GARCH(1,1) and MS-GARCH(1,1) models. The unconditional coverage ratio is calculated as $1 - \left(\sum_{i=1}^{n}\sum_{t=\tau_1+H}^{\tau_1+\tau_2}\mathbb{I}[Y_{i,t} < GaR_{i,t|t-H}(c)] \;/\; (n*(\tau_2 - H))\right)$. The out-of-sample periods of the two-quarter ahead (2Q-ahead), three-quarter ahead (3Q-ahead) and one-year ahead (1Y-ahead) GaR forecasts range from 2001Q2 to 2020Q4, 2001Q3 to 2020Q4 and 2001Q4 to 2020Q4, respectively.

To draw formal conclusions on the relative long-horizon GaR forecasting performances of the models, we perform the two-fold backtesting procedure on the two-quarter ahead, three-quarter ahead and one-year ahead GaR forecasts. First, Table X presents the p-values of the unconditional and conditional calibration tests. The outcomes of the unconditional and conditional calibration tests imply that the one-year ahead GaR forecasts of the MS-GARCH model comply with the requirements of correct unconditional and conditional calibration. By contrast, both calibration tests reject correct calibration of the one-year ahead GaR forecasts of the single-regime models and correct calibration of the two-quarter and three-quarter ahead GaR forecasts of all models. Although in line with the pattern of the unconditional coverage ratios, the increasing performance of the MS-GARCH model with the forecasting horizon requires further investigation into the forecasts to verify the economic significance of the statistically correct calibration of the one-year ahead GaR forecasts.

Table X: **Calibration tests on long-horizon GaR forecasts**

	Unconditional calibration test					
	(a) 100%-sample			(b) 80%-sample		
	2Q-ahead	3Q-ahead	1Y-ahead	2Q-ahead	3Q-ahead	1Y-ahead
GARCH	0.001	0.000	0.000	0.001	0.000	0.000
EGARCH	0.000	0.000	0.000	0.000	0.000	0.000
GJR	0.001	0.000	0.000	0.003	0.000	0.000
MS-GARCH	0.005	0.059	0.498	0.006	0.022	0.297
	Conditional calibration test					
GARCH	0.000	0.000	0.000	0.000	0.000	0.000
EGARCH	0.000	0.000	0.000	0.000	0.000	0.000
GJR	0.000	0.000	0.000	0.000	0.000	0.000
MS-GARCH	0.000	0.007	0.330	0.000	0.002	0.131

This table reports the p-values of the unconditional and conditional calibration tests on the long-horizon GaR forecasts. The calibration tests are performed on two-quarter ahead (2Q-ahead), three-quarter ahead (3Q-ahead) and one-year ahead (1Y-ahead) 95% GaR forecasts generated from the GARCH(1,1), EGARCH(1,1), GJR-GARCH(1,1) and MS-GARCH(1,1) models. In panel (a), the models are estimated on moving estimation windows from the 100%-sample that comprises 25 OECD countries. In panel (b), the models are estimated on moving estimation windows from the 80%-sample that includes 20 OECD countries with relatively homogeneous GDP growth dynamics. The test statistics of the unconditional and conditional calibration tests are provided by Equations (17) and (18) on page 14. The null hypotheses of the unconditional and conditional calibration tests imply correct unconditional and correct conditional calibration of the GaR forecasts, respectively. Under the null hypotheses of the unconditional and conditional calibration tests, the test-statistics follow a standard normal and a chi-squared(1) distribution. The out-of-sample evaluation window ranges from 2001Q1 to 2020Q4.

The comparative backtest measures the relative long-horizon GaR forecasting performances of the proposed MS-GARCH model against the single-regime benchmark models. Table XI provides the CDF of the test-statistic of the performed comparative backtests. The values below 0.05 in the 1Y-ahead columns of both panels indicate the statistically significant outperformance of the MS-GARCH model in forecasting one-year ahead GaR. On this forecasting horizon, the backtesting framework assigns the MS-GARCH model to the green zone. Moreover, on forecasting horizons of two and three quarters, the MS-GARCH is placed into the yellow zone based on the insignificance of the outperformance of the MS-GARCH model relative to the GARCH and GJR models. Similar to the short-horizon forecasting exercise, the EGARCH model is considered as the worst performing model based on increasingly unrealistic GaR forecasts during recessionary periods. Panels (a) and (b) express that performing the forecasting analysis on samples of 20 and 25 countries leads to similar conclusions on the relative long-horizon GaR forecasting performance of the MS-GARCH model.

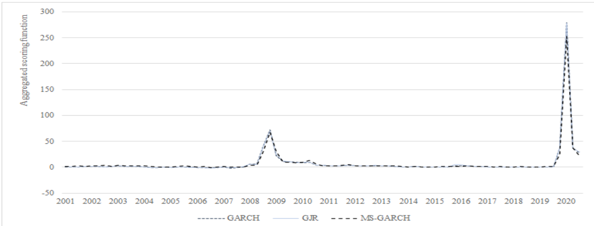
Table XI: Comparative backtest on long-horizon GaR forecasts

	(a) 100%-sample			(b) 80%-sample		
	2Q-ahead	3Q-ahead	1Y-ahead	2Q-ahead	3Q-ahead	1Y-ahead
GARCH	0.428	0.271	0.019	0.151	0.098	0.000
EGARCH	0.004	0.001	0.000	0.002	0.000	0.000
GJR	0.381	0.233	0.024	0.252	0.124	0.001

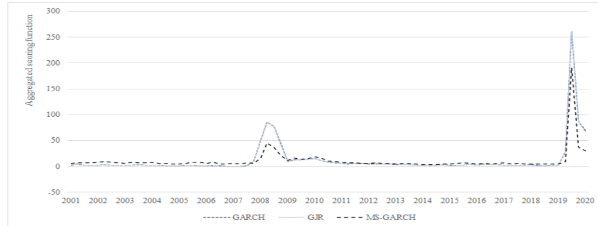
This table reports the CDF of the test-statistic of the comparative backtest, $\Phi(L_3)$. The comparative backtest is performed on two-quarter ahead (2Q-ahead), three-quarter ahead (3Q-ahead) and one-year ahead (1Y-ahead) 95% GaR forecasts from the single-regime and regime-switching models. The null hypothesis that the GaR forecasts of the proposed model are at most (least) as good as the GaR forecasts of the benchmark models is rejected if $1 - \Phi(L_3) \leq \eta$ ($\Phi(L_3) \leq \eta$). We test the GaR forecasting performances of the proposed MS-GARCH(1,1) models against the forecasting performances of the benchmark GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models. In panel (a), the models are estimated on moving estimation windows from the 100%-sample that comprises 25 OECD countries. In panel (b), the models are estimated on moving estimation windows from the 80%-sample that includes 20 OECD countries with relatively homogeneous GDP growth dynamics. The test statistic of the comparative calibration test $\Phi(L_3)$ is provided by Equation (21) on page 16 and follows a standard normal distribution. The out-of-sample evaluation windows of the two-quarter ahead, three-quarter ahead and one-year ahead GaR forecasts, range from 2001Q2 to 2020Q4, 2001Q3 to 2020Q4 and 2001Q4 to 2020Q4, respectively.

To identify the cause of the increasing forecasting performance of the MS-GARCH model with the forecasting horizon, we investigate the courses of aggregated scoring functions and the patterns of long-horizon GaR forecasts. First, Figure 8 plots the aggregated scoring functions over countries based on long-horizon GaR forecasts over the evaluation windows. The aggregated scoring functions in Figures 8a and 8b are based on the two-quarter and one-year GaR forecasts, respectively, generated from the GARCH, GJR and MS-GARCH models, estimated on estimation windows from the 100%-sample. In addition, Figure 19 in Appendix Section H.E of Appendix H provides the corresponding scoring functions of the analysis performed on the 80%-sample. Based on the underperformance of the EGARCH model, as indicated by the the backtesting framework results, and to increase the readability of the graphs, the figures omit the scoring functions of the EGARCH model.

Figures 8a and 8b demonstrate that the increase of the forecasting horizon has a two-fold impact on the relative GaR forecasting power of the MS-GARCH model. First, during periods of macroeconomic expansion, the increasing positive difference between the aggregated scoring function of the MS-GARCH model and the GARCH and GJR models indicates the augmenting underperformance of the MS-GARCH model during these quarters. Conversely, during macroeconomic recessions, the MS-GARCH model increasingly outperforms single-regime models with the forecasting horizon. Considerably lower peaks of the aggregated scoring functions of the MS-GARCH model during the Great Financial Recession and the Covid crisis in Figure 8b visualize the increasing outperformance of the MS-GARCH model during recessionary periods. Considering the outcomes of the comparative backtests, we conclude that the larger outperformance of the MS-GARCH model during recessionary periods than the underperformance during expansionary periods drives the overall outperformance of the MS-GARCH model in forecasting one-year ahead GaR.



(a) Based on two-quarter ahead GaR forecasts

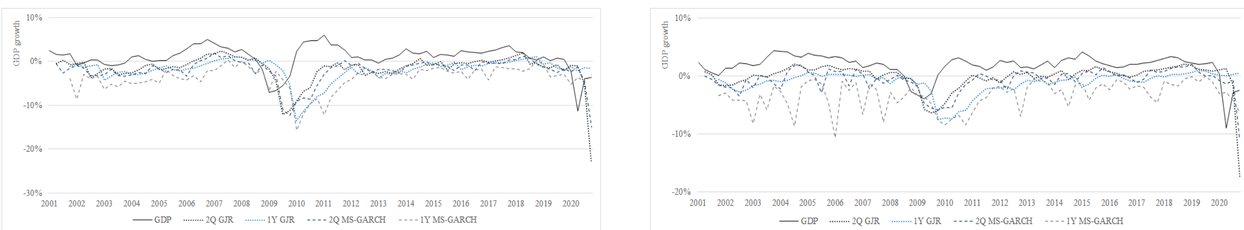


(b) Based on one-year ahead GaR forecasts

Figure 8. This figure plots the comparative scoring function of two-quarter ahead (Figure 8a) and one-year ahead (Figure 8b) GaR forecasts, aggregated over the countries in the sample. The aggregated comparative scoring function is given by $\sum_{i=1}^{n} S_{i,t} = \sum_{i=1}^{n} \big(\mathbb{I}[Y_{i,t} < GaR_{i,t|t-1}(c)] + c - 1\big)\big(GaR_{i,t|t-1}(c)\big) - \mathbb{I}[Y_{i,t|t-1} < GaR_{i,t|t-1}(c)]\big(Y_{i,t}\big)$, where the expression of $S_{i,t}$ is obtained from Equation (20). The dotted dark blue, solid light blue and dashed black lines in the figures represent the aggregated scoring functions of the GARCH, GJR and MS-GARCH models, respectively. The models are estimated on estimation windows from the 100%-sample. The out-of-sample periods of the two-quarter ahead and one-year ahead GaR forecasts range from 2001Q2 to 2020Q4 and from 2001Q4 to 2020Q4, respectively.

Analyzing the patterns of long-horizon GaR forecasts deepens the investigation into the drivers of the increasing relative performance of the MS-GARCH model with the forecasting horizon and the statistically significant outperformance of the MS-GARCH model in forecasting one-year ahead GaR. Below, Figure 9 plots the realized GDP growth rates of DEU and USA, alongside the two-quarter ahead and one-year ahead GaR forecasts of the GJR and MS-GARCH models. Both models are estimated on estimation windows from the 100%-sample. In addition to Figure 9, Figure 14 in Appendix G provides the plots of the remaining 23 OECD countries in the 100%-sample. Based on the high degree of similarity of long-horizon GaR forecasts of the GARCH and GJR models, and the superiority of the GJR model in terms of unconditional coverage, the figures omit the forecasts of the GARC model to increase the readability of the figures.

The patterns of long-horizon GaR forecasts visualize the reliance of H-quarter ahead GaR forecasts on the available macroeconomic information of H quarters earlier. For instance, the one-year ahead GaR forecasts of DEU and USA respond to the Great Financial Recession in 2010Q1 and 2009Q4, respectively, one year later after the beginning of the crisis in the countries. Similarly, the one-year ahead GaR forecasts in Figure 9 show that the final quarter of the out-of-sample period, 2020Q4, comes too early to reflect the Covid crisis. By comparing the forecasts of the models we observe that, due to a higher sensitivity to deteriorating macroeconomic circumstances, the MS-GARCH model produces spikier and, on average, lower long-horizon GaR forecasts than the single-regime models. We consider the high sensitivity to deteriorating economic conditions of the MS-GARCH model as key differentiator to the GaR forecasting performances of the models. On the one hand, the high sensitivity to deteriorating economic conditions results into the underestimation of GaR during periods of economic expansion, which drives the underperformance of the MS-GARCH model during periods of macroeconomic stability. For instance, in response to small decreases in GDP growth of USA during the second half of 2004, the two-quarter and one-year ahead GaR forecasts decrease to -2.98% and -10.69% in 2005Q3 and 2006Q1, respectively. However, due to the economic growth in USA in 2005 and 2006, the relatively low GaR forecasts of the MS-GARCH result into large distances between realized GDP growth rates and GaR forecasts. This illustrates the negative impact of the high sensitivity of deteriorating economic conditions on the feasibility of long-horizon GaR forecasts of the MS-GARCH model during periods of macroeconomic expansion. On the other hand, On the one hand, the more spiky and, on average, inferior GaR forecasts of the MS-GARCH limit the frequency and magnitude of GaR overshootings during recessionary periods. An illustration of this is the GDP growth of DEU in 2002Q2, that overshoots the one-year ahead GaR forecast of the GJR model but remains above the GaR forecast of the MS-GARCH model. The lower level of the one-year ahead GaR forecast, driven by the stagnation of the German economy in the first half of 2001, avoids a GaR overshooting in 2002Q2. The observed patterns of the long-horizon GaR forecasts of the single-regime and regime-switching models are in line with the underperformance of the MS-GARCH model during expansionary periods and the outperformance during recessionary periods, as observed from Figure 8. More specifically, the investigation into the patterns of long-horizon GaR forecasts shows that the high sensitivity of the MS-GARCH model to deteriorating macroeconomic conditions drives the statistically significant outperformance of the MS-GARCH model in forecasting one-year ahead GaR.



(a) DEU

(b) USA

Figure 9. This figure plots the realized GDP growth rates of DEU and USA alongside the two-quarter ahead and one-year ahead 95% GaR forecasts of the GJR-GARCH(1,1) and MS-GARCH(1,1) models. In the figures, the solid black lines plot the realized GDP growth rates of the countries. Moreover, the dotted black and dotted blue lines reflect the two-quarter ahead and one-year ahead 95% GaR forecasts of the GJR model, respectively, and the dashed black and dashed gray lines reflect the two-quarter ahead and one-year 95% GaR forecasts of the MS-GARCH model. The models are estimated on estimation windows from the 100%-sample that comprises 25 OECD countries. The out-of-sample periods of the two-quarter ahead and one-year ahead GaR forecasts range from 2001Q2 to 2020Q4 and 2001Q4 to 2020Q4, respectively.

In addition to the plots of long-horizon GaR forecasts in this section, from models estimated on estimation windows from the 100%-sample, Figure 20 in Appendix Section H.F of Appendix H provides the long-horizon GaR forecasts from the GJR and MS-GARCH models estimated on the 80%-sample. The exclusion of countries with most aberrant GDP growth dynamics has a slightly positive impact on the smoothness of the long-horizon GaR forecasts of the MS-GARCH model. Overall, the unconditional and conditional calibration tests and the comparative backtests in Sections V.B.1 and V.B.2 prove that the conclusions on the relative short-horizon and long-horizon GaR forecasting performances of the models are independent of the exclusion of countries with most aberrant GDP growth dynamics in the sample. We regard the significantly similar GaR forecasting performances of the models in both analyses as a proof of the small impact of the heterogeneity in GDP growth dynamics of the countries in the 100%-sample on our findings. Therefore, we consider the assumption of homogeneous model parameters across countries as valid in this study.

The main findings of this section are thee-fold. First, the long-horizon GaR forecasting exercise demonstrates the statistically significant outperformance of the MS-GARCH model in forecasting one-year ahead GaR. On the one hand, the high sensitivity of the MS-GARCH model to deteriorating macroeconomic circumstances has a negative impact on the relative performance of the MS-GARCH model during expansionary periods due to unrealistically low GaR forecasts. On the other hand, the high sensitivity has a larger positive impact on the relative performance of the MS-GARCH model during expansionary periods due to lower frequencies and magnitudes of GaR overshootings. Second, the MS-GARCH model is unable to significantly improve upon single-regime benchmark models in forecasting two-quarter ahead and three-quarter ahead GaR. Third, based on significantly similar GaR forecasting performances of the models in the analyses on the 100%-sample and the 80%-sample in Sections V.B.1 and V.B.2, this section concludes that the assumption of homogeneous model parameters across countries is valid in this study. Although the results of the backtesting framework in Sections V.B.1 and V.B.2 provide statistically significant evidence for the outperformance of the MS-GARCH model in forecasting one-quarter ahead and one-year ahead GaR, the distinguish in relative GaR forecasting performances during expansionary and recessionary periods and the infeasibility of one-year ahead GaR forecasts during expansionary periods require further investigation to validate the economic significance of the findings.

B.3. Forecasting performances during expansionary and recessionary periods

The validation of the economic significance of the findings on the relative short-horizon and long-horizon GaR forecasting performances of the MS-GARCH model relies on a deepening analysis on the relative GaR forecasting performances of the models during expansionary and recessionary periods. The similar patterns of short-horizon and long-horizon GaR forecasts in the analyses performed on the 100%-sample and the 80%-sample obviates the need to perform the deepening analysis on both samples. Therefore, in this section, we confine ourselves to models estimated on estimation windows from the 100%-sample that comprises all 25 OECD countries. The validation of relative GaR forecasting performances during expansionary and recessionary periods relies on OECD recession indicators to divide the evaluation windows into expansionary and recessionary quarters. Appendix I elaborates on the OECD recession indicators and the methodology of the OECD to identify quarters as recessions. Moreover, Table XXIII in Appendix I provides insights in the GDP growth of the 25 OECD countries during expansionary and recessionary quarters from 2001Q1 to 2020Q4. The OECD recession indicators enable us to construct series of expansionary GaR forecasts and recessionary forecasts. More specifically, the H-quarter ahead GaR forecast $GaR_{i,t+H|t}$ is labelled as expansionary (recessionary) if the economy of country i is in an expansionary (recessionary) state at quarter $t + H$.

To compare the relative forecasting performances of the models during expansionary and recessionary periods, we perform comparative backtests on series of expansionary and recessionary one-quarter ahead to one-year ahead GaR forecasts. Table XII provides the CDF of the test-statistics of the comparative backtests that measure the performance of the proposed MS-GARCH model relative to the single-regime benchmark models during expansionary (panel (a)) and recessionary (panel (b)) periods. First, Table XII demonstrates that the overall significant outperformance of the MS-GARCH model in forecasting one-quarter ahead GaR, as observed in Section V.B.1, is mainly driven by the significant outperformance of the MS-GARCH model during recessionary periods. During expansionary periods, the one-quarter ahead GaR forecasts of MS-GARCH model remain, although not significantly, more accurate than the corresponding GaR forecasts of the single-regime models. The higher performance of the MS-GARCH model in forecasting one-quarter ahead GaR during both expansionary and recessionary periods indicates the independence of the overall outperformance of future states of economies.

Second, opposed to the outperformance of the MS-GARCH model during recessionary periods in panel (b), panel (a) demonstrates the increasing underperformance of the MS-GARCH model in forecasting long-horizon GaR relative to the single-regime GARCH and GJR models. Therefore, the outcomes of the comparative backtests on series of long-horizon GaR forecasts during expansionary periods provide statistical evidence for the underperformance of the MS-GARCH model during periods of macroeconomic stability , as indicated in Section V.B.2. The decreases of the CDF-values based on the one-year ahead expansionary GaR forecasts, compared to the three-quarter ahead expansionary GaR forecasts, indicating a lower extent of underperformance of the MS-GARCH model, are explained by inconsistencies between the OECD recession indicators and the realized GDP growth rates. For instance, based on the methodology to classify troughs of recessions already as expansionary, the OECD classifies the economies of all countries, except KOR and GBR, as expansionary in 2020Q3 and 2020Q4, while only IRL and LUX achieve positive economic growth. As the outperformance of the MS-GARCH model during periods of low economic growth increases with the forecasting horizon, the inconsistencies between the OECD recession indicators and GDP growth rates particularly reduce the underperformance of the MS-GARCH during expansionary periods in forecasting one-year ahead GaR. Based on the results of the comparative backtests on series of expansionary and recessionary GaR forecasts, we conclude that the significant overall outperformance of the MS-GARCH model in forecasting one-year ahead GaR depends on future states of economies.

Table XII: **Comparative backtest on expansionary and recessionary GaR forecasts**

	(a) Expansionary periods				(b) Recessionary periods			
	1Q-ahead	2Q-ahead	3Q-ahead	1Y-ahead	1Q-ahead	2Q-ahead	3Q-ahead	1Y-ahead
GARCH	0.312	0.683	0.990	0.887	0.009	0.052	0.001	0.000
EGARCH	0.000	0.034	0.015	0.009	0.000	0.000	0.000	0.000
GJR	0.241	0.686	0.993	0.933	0.019	0.035	0.000	0.000

This table reports the CDF of the test-statistic of the comparative backtest, $\Phi(L_3)$. The comparative backtest is performed on series of expansionary and recessionary one-quarter ahead (1Q-ahead), two-quarter ahead (2Q-ahead), three-quarter ahead (3Q-ahead) and one-year ahead (1Y-ahead) 95% GaR forecasts. The null hypothesis that the GaR forecasts of the proposed model are at most (least) as good as the GaR forecasts of the benchmark models is rejected if $1 - \Phi(L_3) \leq \eta$ ($\Phi(L_3) \leq \eta$). We test the GaR forecasting performances of the proposed MS-GARCH(1,1) models against the forecasting performances of the benchmark GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models. Panel (a) provides the test results based on series of GaR forecasts during expansionary periods. The restults in panel (b) are based on series of GaR forecasts during recessionary periods. The test statistic of the comparative calibration test $\Phi(L_3)$ is provided by Equation (21) on page 16 and follows a standard normal distribution. The out-of-sample evaluation windows of the one-quarter ahead, two-quarter ahead, three-quarter ahead and one-year ahead GaR forecasts, range from 2001Qq to 2020Q4, 2001Q2 to 2020Q4, 2001Q3 to 2020Q4 and 2001Q4 to 2020Q4, respectively

Predicting the occurrence of recessions and expansions over long periods of time is extremely difficult. Therefore, the capability of a model to produce more accurate GaR forecasts than currently available models without regard for current and future states of the economy is of large value to economists and policymakers. In addition to the statistical significance of the outperformance of the MS-GARCH model in forecasting short-horizon GaR, as derived in Section V.B.1), we consider the outperformance as economically significant based on the independence of the outperformance of states of the economies. Conversely, the dependence of the outperformance of the MS-GARCH model in forecasting one-year ahead GaR on states of the economy implies the reliance of the quality of GaR forecasts on the occurrence and magnitude of recessions. Based on this reliance, as well as on the unrealistic patterns of one-year ahead GaR forecasts during expansionary periods, as observed in Section V.B.2, we reject the economic significance of the outperformance of the MS-GARCH in forecasting one-year ahead GaR.

VI. Joint GaR

Marginal GaR monitors the downside macroeconomic risk for countries individually. The ignorance of interdependencies between countries causes global policymakers to regard marginal GaR as a myopic global risk management tool. As financial crises spread contagiously, small groups of countries may jeopardize the global macrostability. For instance, the sovereign debt crisis in the Eurozone in 2011 originated from smaller periphery economies in Southern Europe. The risk of deteriorating macroeconomic circumstances in specific economies to threaten larger macroeconomic systems illustrates the value of tracking joint macroeconomic risk of countries to supranational economic policymakers and financial supervisory institutions. To monitor the risk of system-wide macroeconomic events in a volatility forecasting framework, Brownlees & Souza (2021) employ the use of joint GaR as prediction regions that contain the GDP growth of a minimum of countries in the sample at desired probability levels. The out-of-sample analysis of Brownlees & Souza (2021) proves the ability of the single-regime GARCH model to construct joint GaR forecasts that exhibit correct unconditional coverage. To provide insights into the relevance of regime-switching models for joint macroeconomic risk prediction, this study forecasts one-quarter ahead joint GaR based on single-regime and regime-switching models. Since joint GaR forecasting is not the main focus of this paper, but rather an avenue for further research, we perform no formal evaluation procedure on the joint GaR forecasts and confine ourselves to presenting the joint GaR forecasts and the unconditional coverage ratios. Moreover, based on the exposed incapability of the MS-GARCH model to achieve statistically and economically significant outperformance in long-horizon marginal GaR forecasting, we restrict this extension to one-quarter ahead joint GaR forecasting. In addition, based on the observed underperformance of the EGARCH model in short-horizon marginal GaR forecasting, the extension focuses on the GARCH, GJR and MS-GARCH models. Finally, due to the exposed similar marginal GaR forecasting performances of models estimated on samples of 25 and 20 OECD countries, the extension on joint GaR only concerns the full sample of 25 countries.

The one-quarter ahead joint GaR of country i at quarter t at coverage level r and conditional probability level c is defined as the prediction region that contains at least $(r * n)$ realized GDP growth rates of the countries in the sample with conditional probability c. To construct one-quarter ahead joint GaR forecasts in the single-regime framework, this study relies on the approach of Brownlees & Souza (2021). Moreover, to construct joint GaR forecasts from the MS-GARCH model, we extend the methodology of Brownlees & Souza (2021) to the regime-switching framework. The construction of joint GaR prediction regions in Brownlees & Souza (2021) relies on the Bootstrap Joint Prediction Regions (BJPR) approach of Wolf & Wunderli (2015). This approach models the one-quarter ahead prediction region that contains the GDP growth of at least $(r * n)$ countries at conditional probability $c = 95\%$, $GaR^J_{i,t+1|t}(r, c)$, as

$$GaR^J_{i,t+1|t}(r, c) = \hat{\mu}_{i,t+1|t} + d_c^{r*n}\left(\hat{h}_{i,t+1|t}\right)^{1/2}, \tag{23}$$

where $\hat{\mu}_{i,t+1|t}$ and $\hat{h}_{i,t+1|t}$ are estimated as the mean and variance of simulated GDP growth rates. Furthermore, d_c^{r*n} approximates the c-quantile of the joint conditional distribution of GDP growth rates of countries based on a bootstrapping approach. Algorithm 2 in Appendix J details the steps to construct one-quarter ahead joint GaR forecasts based on the BJPR approach. In the bootstrapping procedure, we follow Brownlees & Souza (2021) to choose $M = 1000$ as the number of bootstrap replications.

Results

Table XIII below provides the unconditional coverage ratios of the joint GaR forecasts. The joint unconditional coverage ratios denote the fractions of quarters in the evaluation window, for which the set of one-quarter ahead joint GaR forecasts contains at least $(r*n)$ realized GDP growth rates of the n countries in the sample. Similar to the analysis on one-quarter ahead marginal GaR forecasts, the evaluation window ranges from 2001Q1 to 2020Q4. The columns of Table XIII represent the coverage levels $r = 80\%$, $r = 88\%$ and $r = 96\%$, implying required numbers of countries covered by the joint GaR forecasts of $(r * n) = 20$, $(r * n) = 22$ and $(r * n) = 24$, respectively. The results in Table XIII show that the joint unconditional coverage ratios converge to 95% as the coverage level r approaches unity. The decreases of the unconditional coverage with lower coverage levels indicate that the GDP growth rates of the countries are not as granular as assumed by the empirical quantile approximation d_c^{r*n}. Overall, the joint unconditional coverage ratios of the MS-GARCH model are closest to 95%, which indicates the superior joint GaR prediction accuracy of the MS-GARCH model in terms of unconditional coverage. In addition to the unconditional coverage ratios, Appendix K provides plots of the joint GaR forecasts. Moreover, Appendix K elaborates on the impact of the coverage level r on the height and accuracy of the joint GaR forecasts and a suspected reason for the higher joint GaR prediction accuracy of the MS-GARCH model. As we perform no formal procedure to evaluate the joint GaR forecasting performances of the models, it is premature to conclude that the MS-GARCH model outperforms single-regime models in joint GaR forecasting. Nevertheless, the superior prediction accuracy of the MS-GARCH model and the observations from the plots of the joint GaR forecasts indicate the potential relevance of regime-switching models to policymakers focusing on joint macroeconomic downside risk prediction.

Table XIII: **Unconditional coverage ratios of joint GaR forecasts**

	$r = 80\%$	$r = 88\%$	$r = 96\%$
GARCH	0.914	0.926	0.951
GJR	0.901	0.926	0.951
MS-GARCH	0.926	0.951	0.951

This table reports the unconditional coverage ratios of the one-quarter ahead joint GaR forecasts generated from the GARCH(1,1), GJR-GARCH(1,1) en MS-GARCH(1,1) models. The columns represent the coverage levels $r = 80\%$, $r = 88\%$ and $r = 96\%$ of the joint GaR forecasts, implying, respectively, coverages of at least $r * n = 20$, $r * n = 22$ and $r * n = 24$ countries of the 25 sample countries. respectively, with a probability of $c = 95\%$. The joint unconditional coverage rate is calculated as $1 - \big(\sum_{t=\tau_1+1}^{\tau_1+\tau_2} \mathbb{I}\big[\sum_{i=1}^{n} \mathbb{I}[Y_{i,t} < GaR_{i,t|t-1}^{J}(r,c)] > r * n\big] \;/\; ((\tau_2 - 1) * n)\big)$. The evaluation window ranges from 2001Q1 to 2020Q4.

VII. Conclusion

This study investigates the potential relevance of regime-switching models for macroeconomic downside risk prediction. To form a broad conclusion, the study compares short-horizon and long-horizon GaR forecasting performances of the MS-GARCH model relative to the performances of the single-regime GARCH, EGARCH and GJR models. First, the introduction of regime-switching to GaR literature is an extension to the available models of policymakers and economists to quantify and predict macroeconomic downside risk. Second, the unique implementation of the MS-GARCH model contributes to the existing literary framework on regime-switching models, and provides a starting point for future researches to implement more sophisticated GaR forecasting models that rely on regime-switching.

The conclusion of this study is three-fold. First, the MS-GARCH model significantly outperforms single-regime models to forecast short-horizon GaR. The independence of the outperformances of states of future states of the economy confirms the economic significance of the results in the forecasting exercise. The capability of the MS-GARCH model to anticipate macroeconomic turmoils earlier and the tempering of GaR forecasts during extreme economic expansions are the main drivers of the outperformance. Moreover, the overestimation of peaks of major financial crises merely compensates the outperformance of the MS-GARCH model. Second, the MS-GARCH model is incapable of improving upon single-regime benchmark models in forecasting long-horizon GaR. Despite the statistically significant outperformance on the forecasting horizon of one year, we reject the economical significance of the finding based on the reliance of the outperformance on phases of the business cycle and the unrealistic patterns of GaR forecasts during expansionary periods. The third component of the conclusion is the validity of the assumption of homogeneous model parameters across countries in this study. To derive this, the in-sample analysis proposes a sample that excludes FIN, GRC, ISL, KOR, and LUX, as countries that exhibit aberrant GDP growth dynamics. The similar GaR forecasting performances of the models before and after the exclusion of the countries indicate the small impact of the heterogeneity of GDP growth dynamics in the sample on our findings, and confirm the validity of the assumption. In addition to the main focus of the study, the extension on joint GaR forecasting indicates the potential relevance of regime-switching models to policymakers focusing on joint macroeconomic downside risk prediction, although statistical evidence is not provided by this study.

VIII. Discussion and Further Research

Although we conclude that the MS-GARCH model significantly outperforms single-regime models in short-horizon GaR forecasting, some remarks must be made regarding our research approach. First, in the MS-GARCH model, the modelling of GaR relies on the approximation of mixture quantiles as probability-weighted averages of univariate quantiles. This approach ignores the interdependencies between the distributional characteristics of GDP growth in the expansionary and recessionary regimes. Therefore, the theoretically correct alternative of Monte Carlo simulation potentially provides different results. Second, the estimation procedure of this study ignores the high cross-correlations of GDP growth dynamics of countries. The high degrees of co-movement between GDP growth rates around major financial crises, such as the Great Financial Crisis and the Covid crisis, illustrate the high cross-correlation in our sample. Our choice for the Composite Likelihood estimation procedure, that solves the scarcity problem of macroeconomic data, has the disadvantage that it overlooks the cross-correlation in the sample. Third, this study focuses on a fixed split of the data sample into estimation windows, ranging from 1961 to 2000, and evaluation windows, ranging from 2001 to 2020. The choice of the sizes of the estimation and evaluation windows is restricted by the availability of macroeconomic data, which poses a trade-off between using the information of observations for more accurate model estimations or forecast evaluations. Nevertheless, the execution of forecast exercises based on different sizes of the estimation windows and evaluation windows has the potential to result into different findings on the relative GaR forecasting performances of the models.

Finally, we suggest a few avenues for future research. First, to gain more insights in the additional GaR prediction power of regime-switching models, we suggest the implementation of different models that rely on regime-switching to forecast GaR. Alternative tail modelling approaches, such as Extreme Value Theory (EVT), benefit from advantages that tackle the demerits of our research approach. For instance, approaches based on copulae and block maximum estimation are proposed to account for cross-correlation in underlying data samples. Second, to expand the conclusions on the relative outperformance of the MS-GARCH model in particular, we suggest to extend the implementation of the MS-GARCH model of this study. Avenues for extended MS-GARCH implementations are the incorporation of more states of the economy, the inclusion of more than one lag in the conditional mean and variance processes, and the implementation of Markov-switching variants of the EGARCH and GJR models. As a final avenue for further research, we suggest to extend the joint GaR forecasting analysis to obtain more insights into the relative capabilities of the models to predict joint macroeconomic downside risk. The implementation of a formal procedure to evaluate the joint GaR forecasting performances and the consideration of longer forecasting horizons might contain valuable information to economists and policymakers focusing on the risk of supranational macroeconomic turmoils.

REFERENCES

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

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-Risk*, IMF Working Paper WP/18/180. Washington: International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/ 2018/08/02/The-Term-Structure-of-Growth-at-Risk-46150.

Antolin-Diaz, J., Drechsel, T. & Petrella, I. (2017), Tracking the slowdown in long-run GDP growth, *Review of Economics and Statistics* **99**(2), 343–356.

Ardia, D., Bluteau, K., Boudt, K. & Catania, L. (2018), Forecasting risk with Markov-switching GARCH models: A large-scale performance study, *International Journal of Forecasting* **34**(4), 733–747.

Ardia, D. et al. (2008), *Financial risk management with Bayesian estimation of GARCH models*, Vol. 612, Lecture Notes in Economics and Mathematical Systems.

Augustyniak, M. (2014), Maximum likelihood estimation of the Markov-switching GARCH model, *Computational Statistics & Data Analysis* **76**(2), 61–75.

Baffigi, A., Golinelli, R. & Parigi, G. (2004), Bridge models to forecast the Euro area GDP, *International Journal of Forecasting* **20**(3), 447–460.

Bauwens, L., Dufays, A. & Rombouts, J. V. K. (2014), Marginal likelihood for Markov-switching and change-point GARCH models, *Journal of Econometrics* **178**(3), 508–522.

BenSaïda, A., Boubaker, S., Nguyen, D. K. & Slim, S. (2018), Value-at-Risk under market shifts through highly flexible models, *Journal of Forecasting* **37**(8), 790–804.

Bhar, R. & Hamori, S. (2003), Alternative characterization of the volatility in the growth rate of real GDP, *Japan and the World Economy* **15**(2), 223–231.

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

Bry, G. & Boschan, C. (1971), Front matter to "Cyclical analysis of time series: Selected procedures and computer programs", *in* 'Cyclical Analysis of Time Series: Selected Procedures and Computer Programs', National Bureau of Economic Research, pp. 2–13.

Cai, J. (1994), A Markov model of switching-regime ARCH, *Journal of Business & Economic Statistics* **12**(3), 309–316.

Carriero, A., Clark, T. E. & Marcellino, M. G. (2020), Nowcasting tail risks to economic activity with many indicators, *FRB Cleveland Working Paper* .

Cette, G., Fernald, J. & Mojon, B. (2016), The pre-great recession slowdown in productivity, *European Economic Review* **88**(3), 3–20.

Chauvet, M. & Hamilton, J. D. (2006), Dating business cycle turning points, *Contributions to Economic Analysis* **276**(1), 1–54.

Chavleishvili, S. & Manganelli, S. (2019), Forecasting and stress testing with quantile vector autoregression, *ECB Working Paper Series 2330* .

Clements, M. P. & Krolzig, H.-M. (1998), A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP, *The Econometrics Journal* **1**(1), 47–75.

Delle Monache, D., De Polis, A. & Petrella, I. (2020), Modeling and forecasting macroeconomic downside risk, *Bank of Italy Termi di discussione (Economic working paper) 1324* .

Dueker, M. J. (1997), Markov switching in GARCH processes and mean-reverting stock-market volatility, *Journal of Business & Economic Statistics* **15**(1), 26–34.

Duprey, T., Klaus, B. & Peltonen, T. (2017), Dating systemic financial stress episodes in the EU countries, *Journal of Financial Stability* **32**(4), 30–56.

Fagiolo, G., Napoletano, M. & Roventini, A. (2008), Are output growth-rate distributions fat-tailed? some evidence from OECD countries, *Journal of Applied Econometrics* **23**(5), 639–669.

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

Giacomini, R. & White, H. (2006), Tests of conditional predictive ability, *Econometrica* **74**(6), 1545–1578.

Giglio, S., Kelly, B. & Pruitt, S. (2016), Systemic risk and the macroeconomy: An empirical evaluation, *Journal of Financial Economics* **119**(3), 457–471.

Glosten, L. R., Jagannathan, R. & Runkle, D. E. (1993), On the relation between the expected value and the volatility of the nominal excess return on stocks, *The Journal of Finance* **48**(5), 1779–1801.

Gray, S. F. (1996), Modeling the conditional distribution of interest rates as a regime-switching process, *Journal of Financial Economics* **42**(1), 27–62.

Haas, M., Mittnik, S. & Paolella, M. S. (2004), A new approach to Markov-switching GARCH models, *Journal of Financial Econometrics* **2**(4), 493–530.

Hamilton, J. D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**(2), 357–384.

Hamilton, J. D. & Susmel, R. (1994), Autoregressive conditional heteroskedasticity and changes in regime, *Journal of Econometrics* **64**(1-2), 307–333.

IMF (2017), Global financial stability report: Is growth at risk?, *Global Financial Stability Report*.

Jurado, K., Ludvigson, S. C. & Ng, S. (2015), Measuring uncertainty, *American Economic Review* **105**(3), 1177–1216.

Kim, C.-J. (1994), Dynamic linear models with Markov-switching, *Journal of Econometrics* **60**(1-2), 1–22.

Klaassen, F. (2002), Improving GARCH volatility forecasts with regime-switching GARCH, *Empirical Economics* **27**(2), 363–394.

Krolzig, H.-M. (1997), The Markov-switching vector autoregressive model, *in* 'Markov-switching vector autoregressions', Vol. 454, Springer, Berlin, pp. 6–28.

Lahiri, K. & Wang, J. G. (1994), Predicting cyclical turning points with leading index in a Markov switching model, *Journal of Forecasting* **13**(3), 245–263.

Marcellino, M., Stock, J. H. & Watson, M. W. (2003), Macroeconomic forecasting in the Euro area: Country specific versus area-wide information, *European Economic Review* **47**(1), 1–18.

Marcucci, J. (2005), 'Forecasting stock market volatility with regime-switching GARCH models', *Studies in Nonlinear Dynamics & Econometrics* **9**(4), 1–55.

McConnell, M. M. & Perez-Quiros, G. (2000), Output fluctuations in the United States: What has changed since the early 1980's?, *American Economic Review* **90**(5), 1464–1476.

Nelson, D. B. (1991), Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* **59**(2), 347–370.

Nelson, D. B. & Cao, C. Q. (1992), Inequality constraints in the univariate GARCH model, *Journal of Business & Economic Statistics* **10**(2), 229–235.

Nolde, N. & Ziegel, J. F. (2017), Elicitability and backtesting: Perspectives for banking regulation, *The annals of applied statistics* **11**(4), 1833–1874.

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

Plagborg-Møller, M., Reichlin, L., Ricco, G. & Hasenzagl, T. (2020), When is growth at risk?, *Brookings Papers on Economic Activity* **2020**(1), 167–229.

Sajjad, R., Coakley, J. & Nankervis, J. C. (2008), Markov-switching GARCH modelling of Value-at-Risk, *Studies in Nonlinear dynamics and Econometrics* **12**(3), 1–31.

Sharma, M. (2012), Evaluation of Basel III revision of quantitative standards for implementation of internal models for market risk, *IIMB Management Review* **24**(4), 234–244.

Wolf, M. & Wunderli, D. (2015), Bootstrap joint prediction regions, *Journal of Time Series Analysis* **36**(3), 352–376.

Appendices

Appendix A Construction of long-horizon GaR forecasts

This appendix details the construction of long-horizon GaR forecasts, as described in Section III.B.4. The pseudo-code in Algorithm 1 provides all steps to construct long-horizon GaR forecasts using both single-regime and regime-switching models. More specifically, the algorithm details the construction of the H-quarter ahead GaR forecast of country i, forecasted at forecasting quarter t.

Result: The H-quarter ahead GaR forecast of country i, $GaR_{i,t+H|t}(c)$
INPUTS ;
Require: $\{Y_{i,t}\}_{i=1}^{n}$: GDP growth rates of n countries at quarter t ;
Require: $\{\mu_{i,t}\}_{i=1}^{n}$ or $\{\mu_{i,t}^{(k)}\}_{i=1}^{n}$: state-independent or state-dependent conditional means of GDP growth of n countries at quarter t ;
Require: $\{h_{i,t}\}_{i=1}^{n}$ or $\{h_{i,t}^{(k)}\}_{i=1}^{n}$: state-independent or state-dependent conditional variances of GDP growth of n countries at quarter t ;
Require: $\{\pi_{i,t}^{(k)}\}_{i=1}^{n}$: only in the MS-GARCH model: ex-ante regime probabilities of n countries at quarter t ;
Require: Ψ: set of model parameters ;
Require: c: quantile of interest ;
Require: W: number of simulations ;

PROCEDURE ;
for j *in* $1,\ldots,H$ **do**
  for w *in* $1,\ldots,W$ **do**
    if $j=1$ **then**
      if *single-regime model* **then**
        $\epsilon_{i,t} = Y_{i,t} - \mu_{i,t}$;
        $\mu^{\{w\}}_{i,t+1} = \phi_0 + \phi_1 Y_{i,t}$;
        if *GARCH* **then**
          $h^{\{w\}}_{i,t+1} = \omega_i + \alpha_i \epsilon^2_{i,t} + \beta_i h_{i,t}$;
        else if *EGARCH* **then**
          $h^{\{w\}}_{i,t+1} = \exp\{\omega + \alpha \frac{|\epsilon_{i,t}|}{|h_{i,t}|} + \gamma \frac{\epsilon_{i,t}}{h_{i,t}} + \beta \ln h_{i,t}\}$;
        else if *GJR* **then**
          $h^{\{w\}}_{i,t+1} = \omega + \alpha \epsilon^2_{i,t}(1 - \mathbb{I}[\epsilon_{i,t} > 0]) + \gamma \epsilon^2_{i,t} \mathbb{I}[\epsilon_{i,t} > 0] + \beta h_{i,t}$;
        Simulate $u^{\{w\}}_{i,t+1} \overset{iid}{\sim} \mathcal{D}_i$;
        $\tilde{Y}^{\{w\}}_{i,t+1} = \mu^{\{w\}}_{i,t+1} + (h^{\{w\}}_{i,t+1})^{1/2} u^{\{w\}}_{i,t+1}$;
      else if *regime-switching model* **then**
        $\epsilon_{i,t} = Y_{i,t} - \left[\pi^{(1)}_{i,t}\mu^{(1)}_{i,t} + (1-\pi^{(1)}_{i,t})\mu^{(2)}_{i,t}\right]$;
        for $k=1,2$ **do**
          $\mu^{(k)\{w\}}_{i,t+1} = \phi^{(k)}_0 + \phi^{(k)}_1 Y_{i,t}$;
          $h^{(k)\{w\}}_{i,t+1} = \omega^{(k)} + \alpha^{(k)} \epsilon^2_{i,t} + \beta^{(k)} h_{i,t}$,
            where $h_{i,t}$ is computed as the Klaassen (2002) expectation term in Equation (11)
        end
        Simulate $U^{\{w\}}_{i,t+1} \sim \mathrm{Unif}(0,1)$;
        if $U^{\{w\}}_{i,t+1} \leq \pi^{(1)\{w\}}_{i,t+1}$ **then**
          Simulate $u^{(1)\{w\}}_{i,t+1} \overset{iid}{\sim} \mathcal{D}^{(1)}_i$;
          $\tilde{Y}^{\{w\}}_{i,t+1} = \mu^{(1)\{w\}}_{i,t+1} + \left(h^{(1)\{w\}}_{i,t+1}\right)^{1/2} u^{(1)\{w\}}_{i,t+1}$
        else if $U^{\{w\}}_{i,t+1} > \pi^{(1)\{w\}}_{i,t+1}$ **then**
          Simulate $u^{(2)\{w\}}_{i,t+1} \overset{iid}{\sim} \mathcal{D}^{(2)}_i$;
          $\tilde{Y}^{\{w\}}_{i,t+1} = \mu^{(2)\{w\}}_{i,t+1} + \left(h^{(2)\{w\}}_{i,t+1}\right)^{1/2} u^{(2)\{w\}}_{i,t+1}$
        where $\pi^{(1)\{w\}}_{i,t+1}$ are obtained as described in Section III.B.1
    else if $j>1$ **then**
      if *Single-regime model* **then**
        $\epsilon^{\{w\}}_{i,t+j-1} = Y^{\{w\}}_{i,t+j-1} - \mu^{\{w\}}_{i,t+j-1}$;
        $\mu^{\{w\}}_{i,t+j} = \phi_0 + \phi_1 Y^{\{w\}}_{i,t+j-1}$;
        if *GARCH* **then**
          $h^{\{w\}}_{i,t+j} = \omega_i + \alpha_i (\epsilon^{\{w\}}_{i,t+j-1})^2 + \beta_i h^{\{w\}}_{i,t+j-1}$;
        else if *EGARCH* **then**
          $h^{\{w\}}_{i,t+j} = \exp\{\omega + \alpha \frac{|\epsilon^{\{w\}}_{i,t+j-1}|}{|h^{\{w\}}_{i,t+j-1}|} + \gamma \frac{\epsilon^{\{w\}}_{i,t+j-1}}{h^{\{w\}}_{i,t+j-1}} + \beta \ln h^{\{w\}}_{i,t+j-1}\}$;
        else if *GJR* **then**
          $h^{\{w\}}_{i,t+j} = \omega + \alpha (\epsilon^{\{w\}}_{i,t+j-1})^2 (1 - \mathbb{I}[\epsilon^{\{w\}}_{i,t+j-1} > 0]) + \gamma (\epsilon^{\{w\}}_{i,t+j-1})^2 \mathbb{I}[\epsilon^{\{w\}}_{i,t+j-1} > 0] + \beta h^{\{w\}}_{i,t+j-1}$;
        Simulate $u^{\{w\}}_{i,t+j} \overset{iid}{\sim} \mathcal{D}_i$;
        $\tilde{Y}^{\{w\}}_{i,t+j} = \mu^{\{w\}}_{i,t+j} + (h^{\{w\}}_{i,t+j})^{1/2} u^{\{w\}}_{i,t+j}$;
      else if *Regime-switching model* **then**
        $\epsilon^{\{w\}}_{i,t+j-1} = Y^{\{w\}}_{i,t+j-1} - \left[\pi^{(1)\{w\}}_{i,t+j-1}\mu^{(1)\{w\}}_{i,t+j-1} + (1-\pi^{(1)\{w\}}_{i,t+j-1})\mu^{(2)\{w\}}_{i,t+j-1}\right]$;
        for $k=1,2$ **do**
          $\mu^{(k)\{w\}}_{i,t+j} = \phi^{(k)}_0 + \phi^{(k)}_1 Y^{\{w\}}_{i,t+j-1}$;
          $h^{(k)\{w\}}_{i,t+j} = \omega^{(k)} + \alpha^{(k)} (\epsilon^{\{w\}}_{i,t+j-1})^2 + \beta^{(k)} h^{\{w\}}_{i,t+j-1}$,
            where $h^{\{w\}}_{i,t+j-1}$ is computed as the Klaassen (2002) expectation term in Equation (11)
        end
        Simulate $U^{\{w\}}_{i,t+j} \sim \mathrm{Unif}(0,1)$;
        if $U^{\{w\}}_{i,t+j} \leq \pi^{(1)\{w\}}_{i,t+j}$ **then**
          Simulate $u^{(1)\{w\}}_{i,t+j} \overset{iid}{\sim} \mathcal{D}^{(1)}_i$;
          $\tilde{Y}^{\{w\}}_{i,t+j} = \mu^{(1)\{w\}}_{i,t+j} + \left(h^{(1)\{w\}}_{i,t+j}\right)^{1/2} u^{(1)\{w\}}_{i,t+j}$
        else if $U^{\{w\}}_{i,t+j} > \pi^{(1)\{w\}}_{i,t+j}$ **then**
          Simulate $u^{(2)\{w\}}_{i,t+j} \overset{iid}{\sim} \mathcal{D}^{(2)}_i$;
          $\tilde{Y}^{\{w\}}_{i,t+j} = \mu^{(2)\{w\}}_{i,t+j} + \left(h^{(2)\{w\}}_{i,t+j}\right)^{1/2} u^{(2)\{w\}}_{i,t+j}$
        where $\pi^{(1)\{w\}}_{i,t+j}$ are obtained as described in Section III.B.1
  end
end
The H-quarter ahead GaR forecast of country i, $GaR_{i,t+H|t}(c)$ is the 5% quantile of the simulated $\{Y^{\{w\}}_{i,t+H}\}^W_{w=1}$

Algorithm 1: Long-horizon GaR forecasting algorithm

Appendix B Country level descriptive analysis

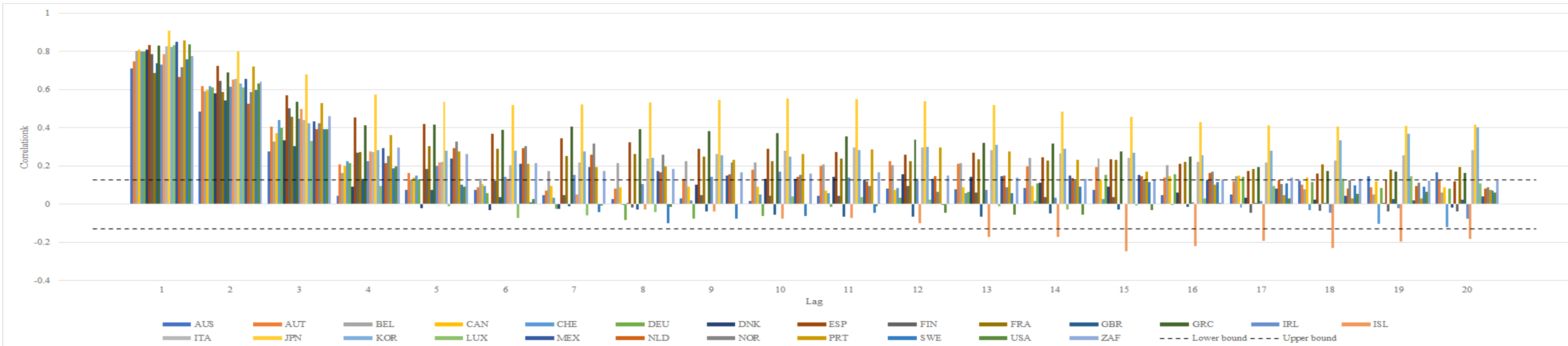
In addition to the information on the overall data sample in Section IV, this appendix provides country-specific information on the GDP growth rates. More specifically, this section provides descriptive statistics and (partial) autocorrelations per country. Below, Table XIV summarizes the descriptive statistics of the GDP growth rates for all 25 OECD countries individually. From this table, we observe some notable differences between economies. For instance, the average growth in economic activity from 1961 to 2020 was considerably higher in KOR and IRL than in other countries in the sample. Moreover, we observe that the GDP growth rates of GRC, IRL, ISL, JPN and KOR have been relatively volatile during the past 60 years. Finally, the 5^{th} percentiles indicate that the economies of GRC, ISL and MEX have experienced quarters with large decreases in GDP growth since 1961.

Table XIV: Descriptive statistics per country

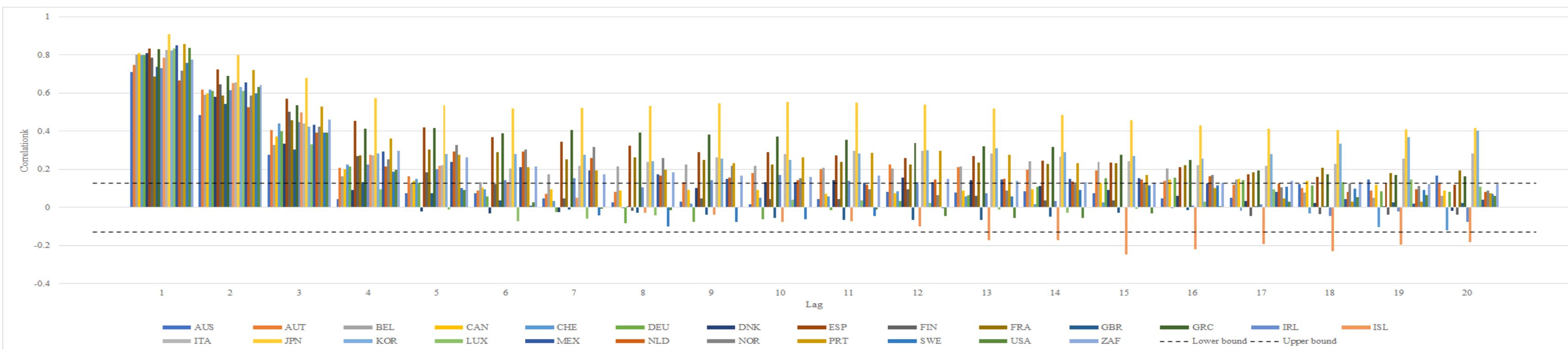
Country	Data Period	Mean	Std. Dev.	5^{th} Perc.	25^{th} Perc.	Median	75^{th} Perc.	95^{th} Perc.
AUS	$1961{-}Q1 : 2020{-}Q4$	3.33	2.26	−0.73	2.09	3.24	4.75	6.93
AUT	$1961{-}Q1 : 2020{-}Q4$	2.60	2.51	−1.17	1.54	2.66	3.91	6.35
BEL	$1961{-}Q1 : 2020{-}Q4$	2.50	2.42	−0.81	1.34	2.40	3.93	6.24
CAN	$1962{-}Q1 : 2020{-}Q4$	2.99	2.64	−1.73	1.79	3.12	4.50	6.67
CHE	$1961{-}Q1 : 2020{-}Q4$	2.16	2.67	−2.32	0.84	2.19	3.81	6.18
DEU	$1961{-}Q1 : 2020{-}Q4$	2.27	2.62	−1.86	0.77	2.27	3.87	6.12
DNK	$1961{-}Q1 : 2020{-}Q4$	2.27	2.61	−1.89	0.66	2.38	3.56	6.72
ESP	$1961{-}Q1 : 2020{-}Q4$	3.14	3.70	−2.53	1.66	3.11	4.66	9.40
FIN	$1961{-}Q1 : 2020{-}Q4$	2.73	3.45	−2.74	0.97	2.90	4.99	7.95
FRA	$1961{-}Q1 : 2020{-}Q4$	2.53	2.72	−0.78	1.22	2.30	4.17	6.27
GBR	$1961{-}Q1 : 2020{-}Q4$	2.18	2.87	−2.74	1.40	2.36	3.63	5.69
GRC	$1961{-}Q1 : 2020{-}Q4$	2.45	5.15	−8.15	0.13	2.61	5.55	10.32
IRL	$1961{-}Q1 : 2020{-}Q4$	4.85	4.52	−0.90	2.54	4.46	6.69	11.47
ISL	$1961{-}Q1 : 2020{-}Q4$	3.66	4.58	−4.13	0.75	3.98	6.63	10.37
ITA	$1961{-}Q1 : 2020{-}Q4$	2.19	3.31	−2.91	0.51	1.79	4.04	7.58
JPN	$1961{-}Q1 : 2020{-}Q4$	3.60	4.13	−1.44	0.95	2.83	5.70	11.38
KOR	$1961{-}Q1 : 2020{-}Q4$	7.25	4.65	−0.61	3.82	7.22	10.63	14.35
LUX	$1961{-}Q1 : 2020{-}Q4$	3.60	3.60	−1.26	1.45	3.39	6.08	9.50
MEX	$1961{-}Q1 : 2020{-}Q4$	3.56	3.92	−4.41	1.84	3.76	6.04	9.10
NLD	$1961{-}Q1 : 2020{-}Q4$	2.66	2.68	−1.67	1.19	2.88	4.28	6.85
NOR	$1961{-}Q1 : 2020{-}Q4$	3.02	2.25	−0.72	1.51	3.10	4.63	6.57
PRT	$1961{-}Q1 : 2020{-}Q4$	3.01	3.76	−3.77	1.25	2.98	5.64	8.79
SWE	$1961{-}Q1 : 2020{-}Q4$	2.48	2.56	−2.04	1.07	2.85	4.07	5.95
USA	$1961{-}Q1 : 2020{-}Q4$	2.96	2.41	−1.63	1.78	3.07	4.33	6.52
ZAF	$1961{-}Q1 : 2020{-}Q4$	2.78	3.03	−1.75	0.99	2.95	4.94	7.23

This table reports the summary statistics of the balanced panel of GDP growth rates for 25 OECD countries. Except for CAN, the sample contains 240 GDP growth observations per country that span from the first quarter of 1961Q1 to 2020Q4. The full sample consists of 5996 observations. The GDP growth rates are in percentages and measured at a quarterly frequency.

Figure 10 on the next page visualizes the autocorrelations and partial autocorrelations of the GDP growth rates per country. Figure 10a shows the decreasing pattern of autocorrelation for increasing lags, as also remarked on the aggregated level in Section IV. Moreover, it shows that the GDP growth rates of especially JPN and ISL exhibit divergent autocorrelations. First, GDP growth rates in JPN are significantly positively autocorrelated with all past growth rates in the past 5 years. Moreover, in ISL, GDP growth is significantly negatively correlated with past GDP growth of more than three years ago. These observations might indicate that the GDP growth dynamics of JPN and ISL deviate from the GDP growth dynamics of the other countries in the sample. The partial autocorrelations in Figure 10b confirm that the GDP growth rates of all countries are to some extent subject to seasonality.



(a) Autocorrelations per country



(b) Partial autocorrelations per country

Figure 10. This figure plots the average autocorrelations and average partial autocorrelations across all countries over the period 1961Q1-2020Q4. The dotted lines in the plots approximate the two standard error bounds and are calculated as $\pm\ 1/96/\sqrt{T}$. The GDP growth rates are measured at a quarterly frequency.

Appendix C Shape parameters

In addition to the parameter estimates of the single and regime-switching models in, respectively, Tables II and III in Section V.A.1, this appendix provides the shape parameters for the 25 individual OECD countries in the 100%-sample. Table XV provides the shape parameters of the Student's t conditional distribution of GDP growth per country in the three single-regime models.

Table XV: Estimates of shape parameters per country in single-regime models

	GARCH		EGARCH		GJR	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
AUS	5.825***	(1.164)	4.822***	(0.888)	6.011***	(1.226)
AUT	4.155***	(0.792)	3.989***	(0.714)	4.279***	(0.868)
BEL	3.171***	(0.313)	3.202***	(0.328)	3.187***	(0.319)
CAN	4.289***	(0.701)	3.637***	(0.476)	4.376***	(0.746)
CHE	4.566***	(0.942)	3.990***	(0.666)	4.560***	(0.950)
DEU	4.155***	(0.714)	3.681***	(0.503)	4.198***	(0.737)
DNK	6.663***	(2.125)	5.140***	(1.367)	6.671***	(2.238)
ESP	3.122***	(0.292)	3.277***	(0.346)	3.149***	(0.302)
FIN	7.396***	(2.821)	6.341***	(2.365)	7.902**	(3.331)
FRA	2.892***	(0.224)	2.968***	(0.251)	2.894***	(0.224)
GBR	3.261***	(0.349)	3.257***	(0.360)	3.266***	(0.352)
GRC	5.571***	(0.782)	5.327***	(0.836)	5.761***	(0.820)
IRL	5.467***	(1.022)	5.193***	(0.995)	5.498***	(0.974)
ISL	7.191***	(2.134)	6.926***	(2.380)	7.225***	(2.088)
ITA	4.459***	(0.846)	3.968***	(0.628)	4.495***	(0.868)
JPN	5.646***	(1.383)	4.657***	(1.043)	5.780***	(1.467)
KOR	5.714***	(1.407)	6.004***	(1.781)	6.039***	(1.573)
LUX	8.615***	(1.942)	7.448***	(2.123)	9.162***	(2.149)
MEX	4.330***	(0.605)	3.952***	(0.543)	4.456***	(0.640)
NLD	4.070***	(0.672)	3.950***	(0.643)	4.076***	(0.681)
NOR	9.990*	(5.670)	7.970*	(4.106)	10.301*	(5.958)
PRT	4.987***	(1.023)	4.419***	(0.728)	5.133***	(1.091)
SWE	6.415***	(1.469)	5.270***	(1.271)	6.448***	(1.520)
USA	3.571***	(0.458)	3.509***	(0.444)	3.632***	(0.487)
ZAF	5.111***	(0.638)	4.412***	(0.582)	5.164***	(0.652)

This table reports the estimates of the shape parameters of the conditional Student's t-distribution of GDP growth in the single-regime GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models. The table contains all country-specific shape parameters ν_i for all countries $i = 1, \ldots, 25$ in the 100%-sample. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The in-sample data in the 100%-sample consists of quarterly GDP growth rates from 1962Q1 to 2020Q4 for CAN and from 1961Q1 to 2020Q4 for the remaining 24 OECD countries.

The regime-switching framework allows the shape parameters of the conditional distributions of GDP growth to change across states of the economy. Below, Table XVI provides the shape parameters of the conditional t-distribution per regime and per country.

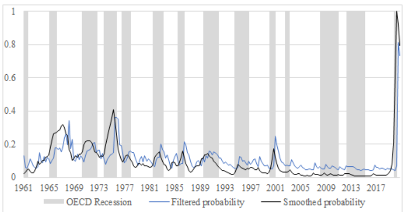
Table XVI: Estimates of shape parameters per country in the MS-GARCH model

	Expansionary regime ($k = 1$)		Recessionary regime ($k = 2$)	
	Coefficient	Std. Error	Coefficient	Std. Error
AUS	36.795	(125.784)	2.313***	(0.270)
AUT	6.744*	(3.485)	2.353***	(0.260)
BEL	4.856***	(1.635)	2.811***	(0.994)
CAN	26.384	(59.479)	2.297***	(0.171)
CHE	13.346	(12.052)	5.118	(6.274)
DEU	3.571***	(0.476)	2.046***	(0.028)
DNK	8.838	(5.877)	2.307***	(0.178)
ESP	4.697***	(1.530)	3.035**	(1.535)
FIN	9.721	(6.448)	2.501***	(0.365)
FRA	5.318***	(1.758)	16.321	(72.192)
GBR	4.497***	(1.264)	2.377***	(0.285)
GRC	10.886	(7.082)	2.598***	(0.458)
IRL	27.589	(53.217)	3.317**	(1.509)
ISL	34.396	(86.894)	16.442	(83.486)
ITA	10.651	(7.796)	2.005***	(0.002)
JPN	7.575***	(2.843)	2.028***	(0.019)
KOR	13.444	(12.091)	22.404	(158.152)
LUX	45.676	(182.305)	2.471***	(0.429)
MEX	24.798	(47.153)	2.550***	(0.391)
NLD	9.164	(5.894)	2.902***	(0.823)
NOR	21.510	(28.697)	2.308***	(0.209)
PRT	10.046	(6.510)	2.641***	(0.615)
SWE	22.070	(31.797)	2.225***	(0.145)
USA	5.178***	(1.805)	2.221***	(0.140)
ZAF	42.583	(177.919)	2.473***	(0.486)

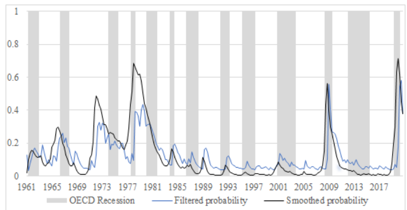
This table reports the estimates of the shape parameters of the conditional Student's t-distribution of GDP growth in the MS-GARCH(1,1) model. The table contains all country-specific shape parameters $\nu_i^{(k)}$ for both regimes $k = 1, 2$ and for all countries $i = 1, \ldots, 25$ in the 100%-sample. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The in-sample data in the 100%-sample consists of quarterly GDP growth rates from 1962Q1 to 2020Q4 for CAN and from 1961Q1 to 2020Q4 for the remaining 24 OECD countries.

Appendix D Recession probabilities

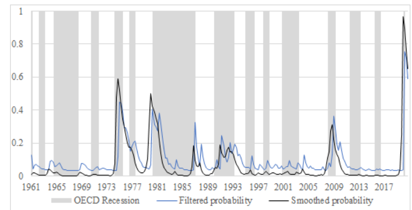
This appendix complements the recession probabilities of DEU, JPN, and USA in Figure 2 in Section V.A.1. Below, Figure 11 contains the filtered and smoothed recession probabilities for the remaining 22 OECD countries in the 100%-sample over the period 1961Q1-2020Q4.



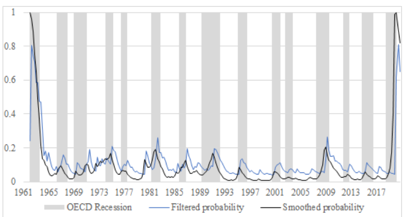
(a) AUS



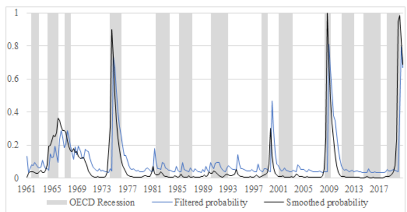
(b) AUT



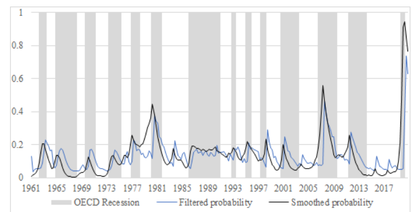
(c) BEL



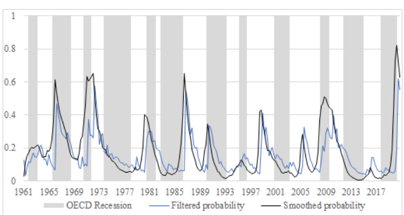
(d) CAN



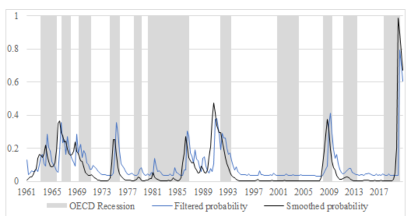
(e) CHE



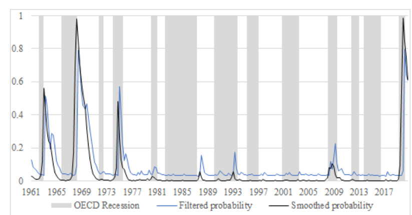
(f) DNK



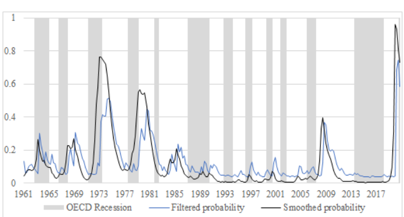
(g) FIN



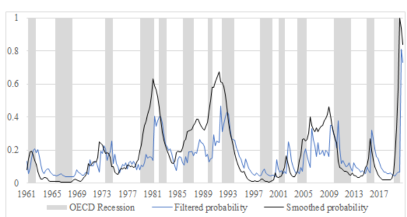
(h) ESP



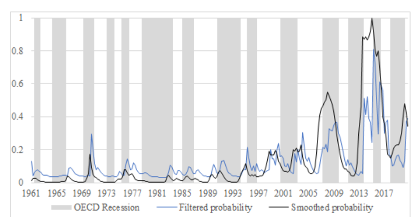
(i) FRA



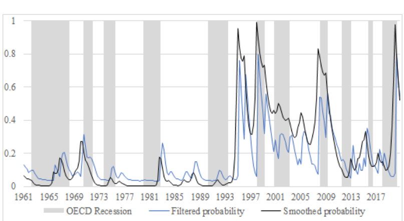
(j) GBR



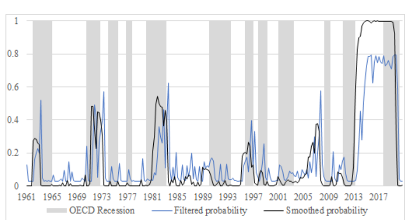
(k) GRC



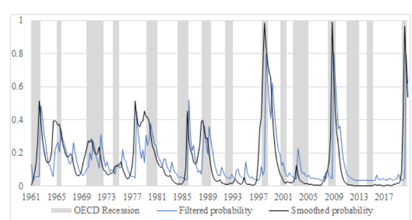
(l) IRL



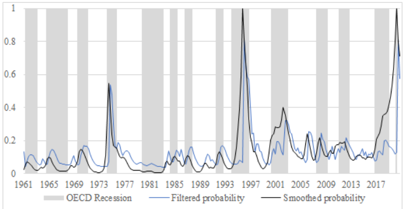
(m) ISL



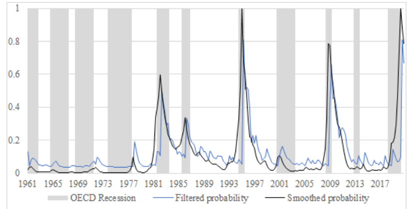
(n) ITA



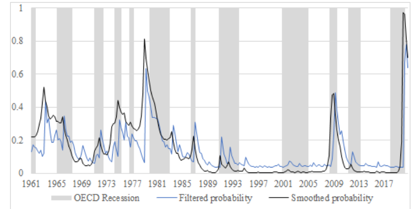
(o) KOR



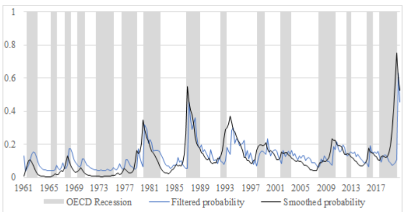
(p) LUX



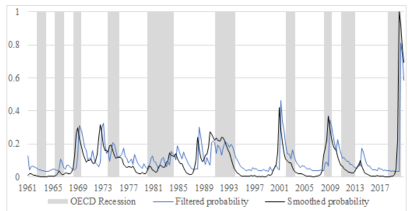
(q) MEX



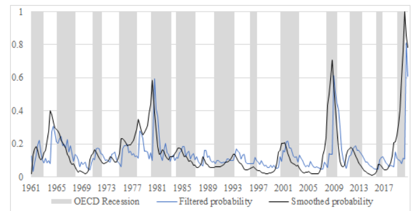
(r) NLD



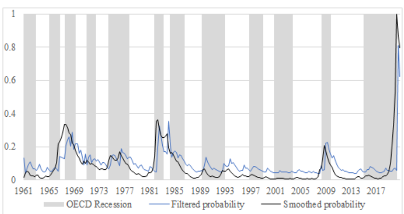
(s) NOR



(t) PRT



(u) SWE



(v) ZAF

Figure 11. This figure plots the filtered (in blue) and smoothed (in black) recession probabilities of the 25 countries in the 100%-sample, except DEU, JPN, and USA. The filtered and smoothed recession probabilities range from 1962Q1 to 2020Q4 for CAN and from 1961Q1 to 2020Q4 for the remaining OECD countries. OECD recession periods are highlighted in gray.

Appendix E In-sample estimations on various samples

To propose a sample that consists of countries with relatively homogeneous GDP dynamics, we perform model estimations based on samples of various numbers of countries. In addition to the estimations of the models on the full sample in Section V.A.1 and the reduced sample in Section V.A.2, this section provides estimations of the models on samples of 25, 18 and 17 OECD countries, with GDP growth rates that range from 1961Q1 to 2019Q4. First, Tables XVII and XVIII contain the estimates of the single-regime models and regime-switching model, respectively, based on a sample comprising GDP growth rates of 25 OECD countries, ranging from 1961Q1 to 2019Q4. These estimations are the basis for the choice of the countries with most aberrant GDP growth dynamics in Figure 4. Table XVII shows the parameter estimates of the single-regime GARCH, EGARCH and GJR models.

Table XVII: **Estimations of single-regime models on the sample of 25 countries**

	GARCH		EGARCH		GJR	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
ϕ_0	0.296***	(0.021)	0.270***	(0.021)	0.288***	(0.021)
ϕ_1	0.893***	(0.006)	0.899***	(0.006)	0.895***	(0.006)
ω	0.136***	(0.013)	−0.391***	(0.018)	0.135***	(0.013)
α	0.290***	(0.019)	0.563***	(0.026)	0.322***	(0.025)
β	0.651***	(0.016)	0.937***	(0.007)	0.651***	(0.016)
γ			−0.048***	(0.015)	0.258***	(0.023)
$\overline{\nu}$	31.167		27.706		32.086	
No. of Par.	30		31		31	
Log(L)	10050.614		10000.845		10047.907	
AIC	3.417		3.1401		3.417	
BIC	3.451		3.436		3.452	

This table reports the maximum likelihood estimates of the single-regime GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models. $\overline{\nu}$ denotes the average shape parameter across the countries. AIC is the Akaike information criterion calculated as $-2\log(L)/T + 2m/T$, where m denotes the number of parameters. BIC is the Schwarz criterion, calculated as $-2\log(L)/T + (m/T)\log(T)$. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The estimation data consists of quarterly GDP growth rates of 25 OECD countries ranging from 1962Q1 to 2019Q4 for CAN and from 1961Q1 to 2019Q4 for the remaining countries.

Furthermore, Table XVIII below provides the estimation of the MS-GARCH model based on the sample that consists of GDP growth rates from all 25 OECD countries, ranging from 1961Q1 to 2019Q4.

Table XVIII: **Estimations of the MS-GARCH model on the sample of 25 countries**

	Expansionary regime ($k = 1$)		Recessionary regime ($k = 2$)	
	Coefficient	Std. Error	Coefficient	Std. Error
ϕ_0	0.525***	(0.047)	0.200***	(0.024)
ϕ_1	0.839***	(0.011)	0.921***	(0.007)
ω	0.365***	(0.057)	0.133***	(0.016)
α	0.196***	(0.028)	0.503***	(0.039)
β	0.730***	(0.028)	0.401***	(0.032)
p	0.993***	(0.002)		
q			0.994***	(0.002)
$\overline{\nu}$	17.879		26.700	
σ	2.221		1.177	
ρ	0.926		0.904	
$\tilde{\pi}$	0.462		0.538	
No. of Par.	62			
Log(L)	9976.945			
AIC	3.404			
BIC	3.474			

This table reports the maximum likelihood estimates of the MS-GARCH(1,1) model. $\overline{\nu}^{(k)}$ denote the average shape parameter across the countries per regime k. $\sigma^{(k)} = (\omega^{(k)}/(1-\alpha^{(k)} - \beta^{(k)}))^{1/2}$ denotes the unconditional volatility per regime. Moreover, $\rho^{(k)} = \alpha^{(k)} + \beta^{(k)}$ are the persistence of shocks in regime k. $\tilde{\pi}^{(k)}$ denote the unconditional probabilities of a country being in regime k and are given by $\tilde{\pi}^{(1)} = (1-q)/(2-p-q)$ and $\tilde{\pi}^{(2)} = (1-p)/(2-p-q)$ for states $k = 1$ and $k = 2$, respectively. AIC is the Akaike information criterion calculated as $-2\log(L)/T + 2m/T$, where m denotes the number of parameters. BIC is the Schwarz criterion, calculated as $-2\log(L)/T + (m/T)\log(T)$. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The estimation data consists of quarterly GDP growth rates of 25 OECD countries ranging from 1962Q1 to 2019Q4 for CAN and from 1961Q1 to 2019Q4 for the remaining countries.

After determining the five most aberrant countries in terms of GDP growth dynamics, we verify the relative homogeneity of the countries in the 80%-sample by recurrently omitting the next country in the remaining sample with most aberrant GDP dynamics and re-estimating the models. In this procedure, the aggregated log likelihoods in Figure 4 indicate the order of the exclusion of the countries. In practice, this order implies that the countries that are iteratively excluded from the 80%-sample are IRL, SWE, JPN, NOR, and so forth. The estimations of the models based on samples that comprise less than 20 OECD countries show that the parameter estimates are robust to the exclusion of more than five countries from the full sample. To illustrate this, Tables XIX and XX provide the estimations of the single-regime models and regime-switching model, respectively, estimated on samples of 18 and 17 OECD countries. In both tables, panels (a) represent the estimations of the models based on the 72%-sample, excluding KOR, GRC, ISL, FIN, LUX, IRL and SWE to contain 18 countries. Moreover, panels (b) provide the estimations of the models based on the 68%-sample, excluding KOR, GRC, ISL, FIN, LUX, IRL, SWE, and JPN. In accordance with the estimations of the models based on the 80%-sample in Section V.A.2, the GDP growth rates of the remaining 18 and 17 countries in these estimations range from 1961Q1 to 2019Q4.

Table XIX below presents the estimations of the single-regime models based on the 72%-sample and the 68%-sample. Compared to the parameter estimates of the GARCH model on the 80%-sample, in Table IV on page 24, the parameters estimated on the 72%-sample change in absolute terms at most by 0.015 (ω). In the EGARCH and GJR models, the largest absolute parameter differences are 0.022 (α) and 0.020 (γ), respectively. Moreover, comparing the parameter estimates based on data from 17 countries (68%-sample) against the parameter estimates based on GDP growth rates of 20 countries (80%-sample) shows maximum absolute changes of 0.019 (ω), 0.014 (α) and 0.020 (ω) in the GARCH, EGARCH and GJR models, respectively.

Table XIX: **Estimations of single-regime models on samples of 18 and 17 countries**

	(a) 72%-sample (n=18)			(b) 68%-sample (n=17)		
	GARCH	EGARCH	GJR	GARCH	EGARCH	GJR
ϕ_0	0.282***	0.247***	0.270***	0.289***	0.257***	0.277***
ϕ_1	0.893***	0.902***	0.896***	0.891***	0.898***	0.893***
ω	0.110***	−0.367***	0.108***	0.106***	−0.375***	0.103***
α	0.269***	0.501***	0.312***	0.274***	0.509***	0.318***
β	0.664***	0.936***	0.668***	0.660***	0.936***	0.665***
γ		−0.062***	0.221***		−0.062***	0.222***
$\overline{\nu}$	38.870	40.707	42.50	38.963	38.391	41.697
No.of Par.	23	24	24	24	25	25
Log(L)	6633.436	6596.729	6629.332	6209.040	6174.758	6204.732
AIC	3.134	3.117	3.133	3.106	3.090	3.105
BIC	3.168	3.153	3.168	3.140	3.126	3.141

This table reports the maximum likelihood estimates of the single-regime GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models estimated on the 72%-sample (panel (a)) and the 68%-sample (panel (b)). $\overline{\nu}$ denotes the average shape parameter across the countries. AIC is the Akaike information criterion calculated as $-2\log(L)/T + 2m/T$, where m denotes the number of parameters. BIC is the Schwarz criterion, calculated as $-2\log(L)/T + (m/T)\log(T)$. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The 72%-sample (68%-sample) consists of quarterly GDP growth rates of 18 (17) OECD countries ranging from 1962Q1 to 2019Q4 for CAN and from 1961Q1 to 2019Q4 for the other countries.

Table XX reports the parameter estimates of the MS-GARCH model based on the 72%-sample and the 68%-sample. In the expansionary regime ($k = 1$), reducing the number of sample countries from 20 to 18 (17) results into a decrease of $\phi_0^{(1)}$ by 0.037 (0.000) and an increase of $\phi_1^{(1)}$ of 0.005 (-0.005). Moreover, the conditional variance parameters $\omega^{(1)}$, $\alpha^{(1)}$, $\beta^{(1)}$ increase by 0.000 (0.000), -0.020 (-0.016), and 0.013 (0.000). In the recessionary regime ($k = 2$) the changes of the parameter estimates in Table XX are of similar magnitude compared to the estimation of the MS-GARCH model based on the 80%-sample in Table V.

Table XX: Estimation of the MS-GARCH model on samples of 18 and 17 countries

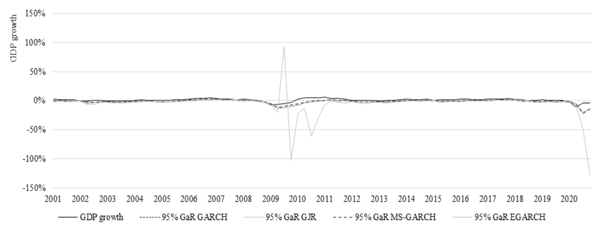
	(a) 72%-sample (n=18)		(b) 68%-sample (n=17)	
	Expansionary regime ($k = 1$)	Recessionary regeime ($k = 2$)	Expansionary regime ($k = 1$)	Recessionary regeime ($k = 2$)
ϕ_0	0.930***	0.157***	0.967***	0.164***
ϕ_1	0.856***	0.840***	0.846***	0.839***
ω	0.000	0.029**	0.000	0.027
α	0.351***	0.270***	0.355***	0.274***
β	0.556***	0.689***	0.543***	0.685***
p	0.747***		0.748***	
q		0.853***		0.857***
$\overline{\nu}$	18.885	10.952	24.133	17.144)
σ	0.000	0.841	0.000	0.812
ρ	0.907	0.959	0.898	0.959
$\tilde{\pi}$	0.367	0.633	0.362	0.638
No. of Par.	48		46	
Log(L)	6597.929		6140.300	
AIC	3.116		3.084	
BIC	3.225		3.156	

This table reports the maximum likelihood estimates of the MS-GARCH(1,1) model estimated on the 72%-sample (panel (a)) and the 68%-sample (panel (b)). $\overline{\nu}^{(k)}$ denote the average shape parameter across the countries per regime k. $\sigma^{(k)} = (\omega^{(k)}/(1-\alpha^{(k)}-\beta^{(k)}))^{1/2}$ denotes the unconditional volatility per regime. Moreover, $\rho^{(k)} = \alpha^{(k)} + \beta^{(k)}$ are the persistence of shocks in regime k. $\tilde{\pi}^{(k)}$ denote the unconditional probabilities of a country being in regime k and are given by $\tilde{\pi}^{(1)} = (1-q)/(2-p-q)$ and $\tilde{\pi}^{(2)} = (1-p)/(2-p-q)$ for states $k = 1$ and $k = 2$, respectively. AIC is the Akaike information criterion calculated as $-2\log(L)/T + 2m/T$, where m denotes the number of parameters. BIC is the Schwarz criterion, calculated as $-2\log(L)/T + (m/T)\log(T)$. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The 72%-sample (68%-sample) consists of quarterly GDP growth rates of 18 (17) OECD countries ranging from 1962Q1 to 2019Q4 for CAN and from 1961Q1 to 2019Q4 for the other countries.

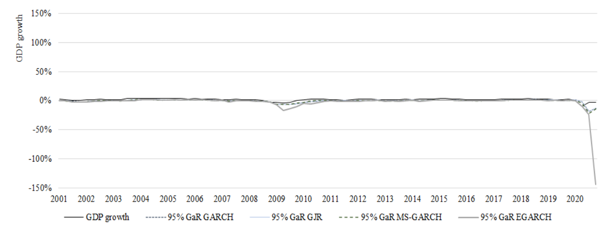
The estimations of the single-regime models and the regime-switching model in Tables XIX and XX show the robustness of the parameter estimates after excluding KOR, GRC, ISL, FIN, and LUX from the data sample. Based on the observed convergence of the parameter estimates, we infer that the 20 countries in the 80%-sample exhibit relatively homogeneous GDP growth dynamics. Therefore, we consider the 80%-sample an appropriate to validate the assumption of homogeneous model parameters in this study by comparing the GaR forecasting performances of the single-regime and regime-switching models estimated on estimation windows from the 100%- and 80%-samples.

Appendix F Short-horizon GaR forecasts

To complement Section V.B.1, this appendix provides plots of the one-quarter ahead GaR forecasts. First, to illustrate the extremity of GaR forecasts generated from the EGARCH model, Figure 12 below shows the course of the realized GDP growth rates of DEU and USA alongside the one-quarter ahead GaR forecasts of the MS-GARCH model and all single-regime models. The models are estimated based on moving estimation windows from the 100%-sample that comprises 25 OECD countries.



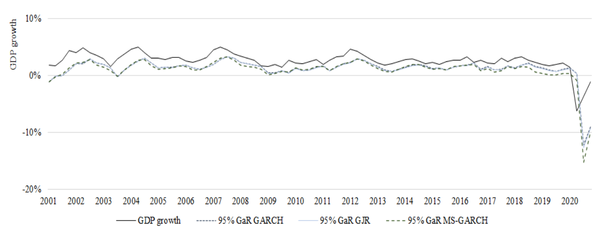
(a) DEU



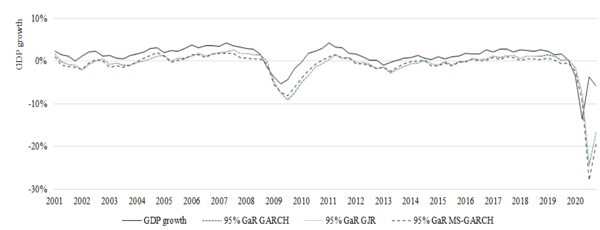
(b) USA

Figure 12. This figure plots the realized GDP growth rates of DEU and USA alongside the one-quarter ahead 95% GaR forecasts of the GARCH(1,1), EGARCH(1,1), GJR-GARCH(1,1) and MS-GARCH(1,1) models. The solid black, dotted dark blue, solid light blue, and dashed black lines reflect the realized GDP growth rates and the forecasts of the GARCH, GJR and MS-GARCH models, respectively. The models are estimated on estimation windows from the 100%-sample that comprises 25 OECD countries. The out-of-sample evaluation window ranges from 2001Q1 to 2020Q4.

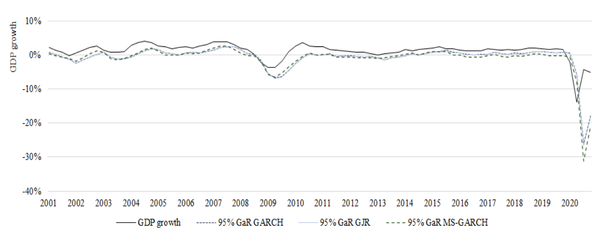
In addition to Figure 7, Figure 13 provides the realized GDP growth rates and GaR forecasts for the remaining 21 OECD countries in the 100%-sample. Due to the observed underperfomance of the EGARCH model, Figure 13 only provides the one-quarter ahead GaR forecasts of the GARCH, GJR and MS-GARCH models.



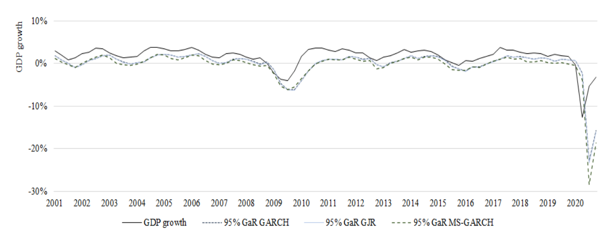
(a) AUS



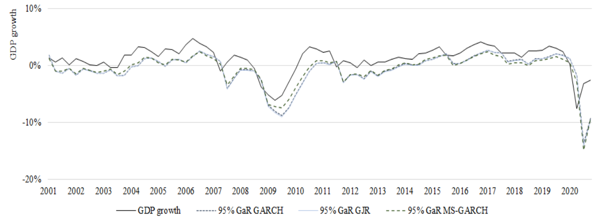
(b) AUT



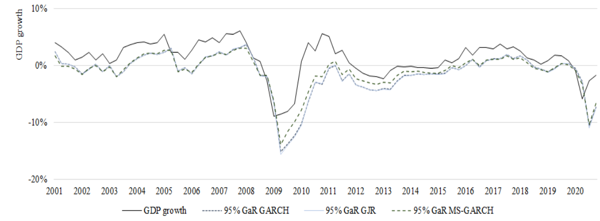
(c) BEL



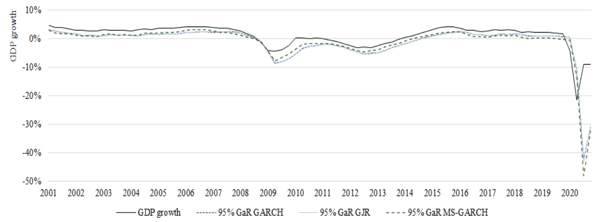
(d) CAN



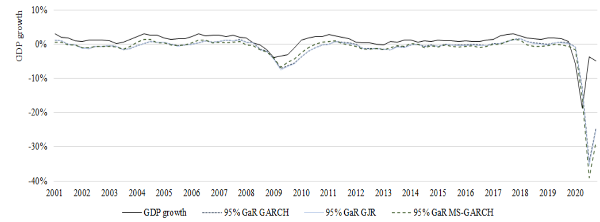
(e) DNK



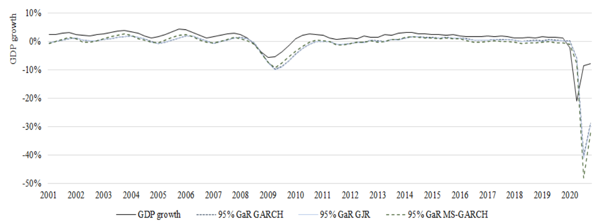
(f) FIN



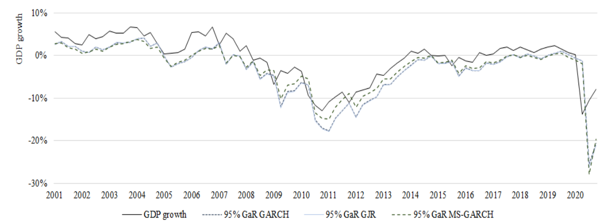
(g) ESP



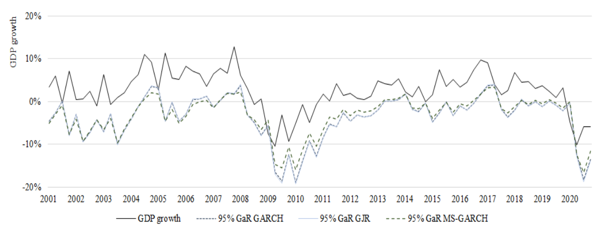
(h) FRA



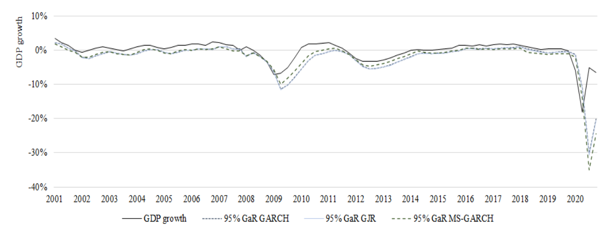
(i) GBR



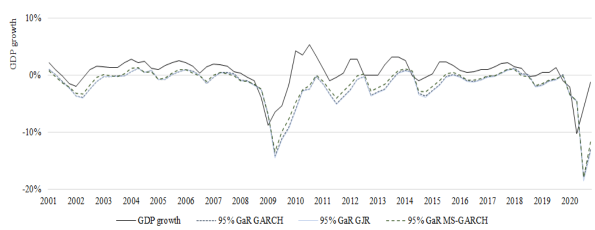
(j) GRC



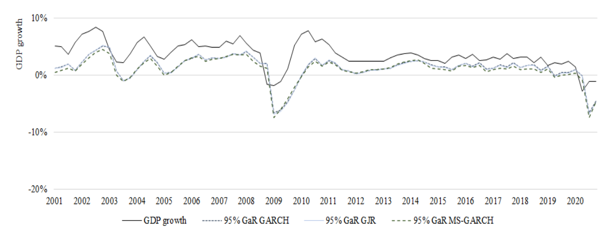
(k) ISL



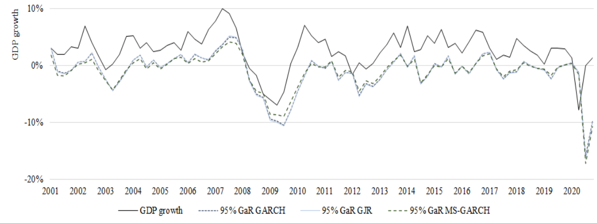
(l) ITA



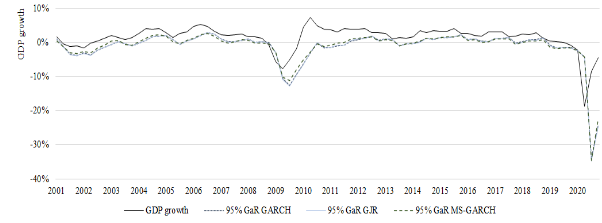
(m) JPN



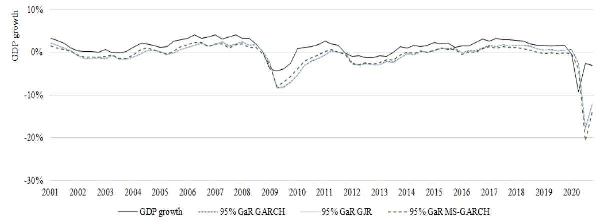
(n) KOR



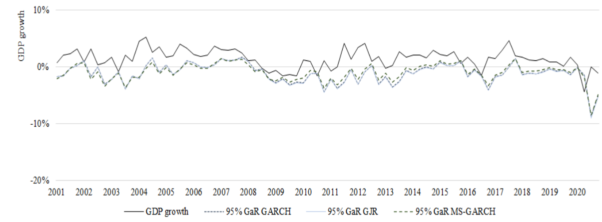
(o) LUX



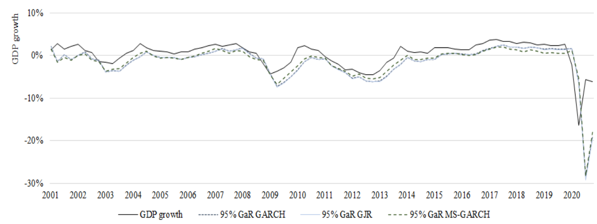
(p) MEX



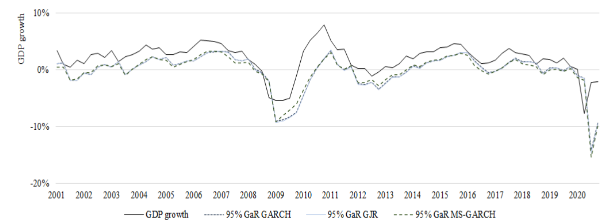
(q) NLD



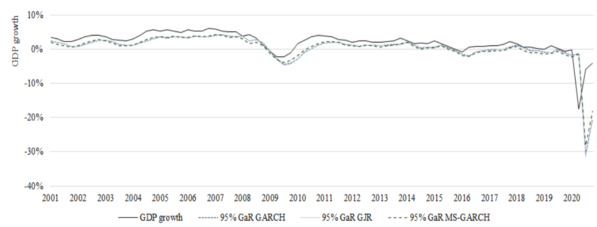
(r) NOR



(s) PRT



(t) SWE

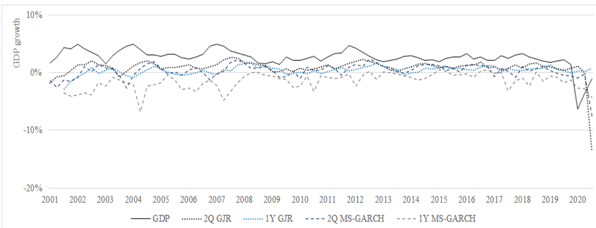


(u) ZAF

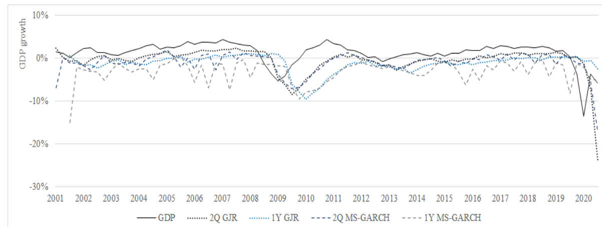
Figure 13. This figure plots the realized GDP growth rates alongside the one-quarter ahead 95% GaR forecasts of the GARCH(1,1), GJR-GARCH(1,1) and MS-GARCH(1,1) models. Subfigures (13a)-(13u) provide the graphs for all countries in the 100%-sample, except CHE, DEU, IRL and USA. In the figures, the solid black lines plot the realized GDP growth rates of the countries. Moreover, the dotted dark blue, solid light blue and dashed black lines reflect the one-quarter ahead 95% GaR forecasts of the GARCH, GJR and MS-GARCH models, respectively. The models are estimated on estimation windows from the 100%-sample that comprises 25 OECD countries. The out-of-sample periods range from 2001Q1 to 2020Q4.

Appendix G Long-horizon GaR forecasts

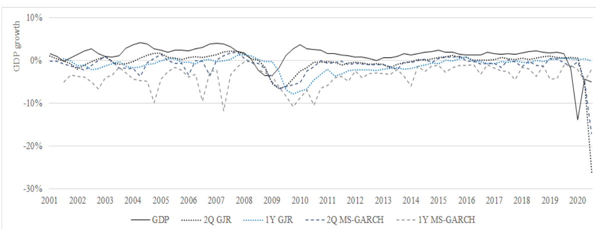
In addition to Figure 9 in Section V.B.2, this appendix provides the long-horizon GaR forecasts of the countries in the 100%-sample, except DEU and USA. Below, Figure 14 plots the realized GDP growth rates alongside the two-quarter and one-year ahead GaR forecasts of the GJR model and the MS-GARCH model. The models are estimated based on moving estimation windows from the 100%-sample that comprises 25 OECD countries.



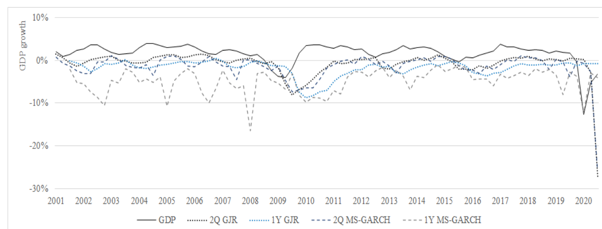
(a) AUS



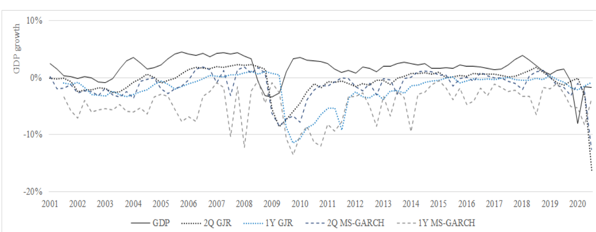
(b) AUT



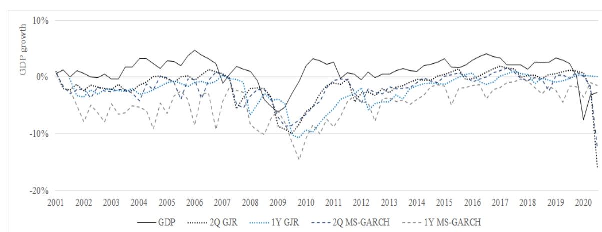
(c) BEL



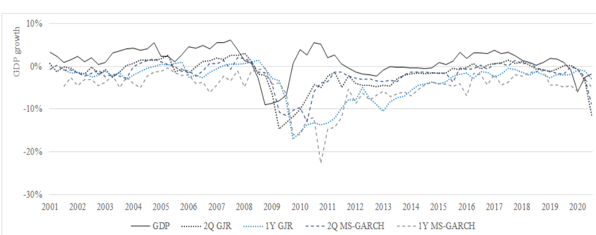
(d) CAN



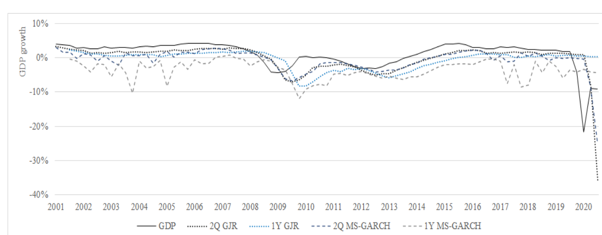
(e) CHE



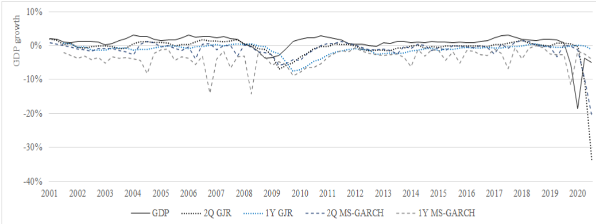
(f) DNK



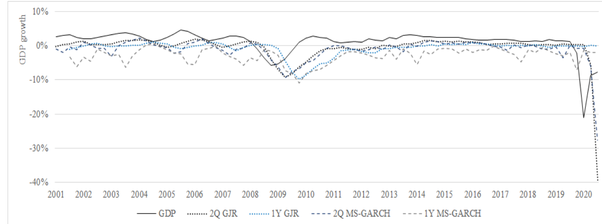
(g) FIN



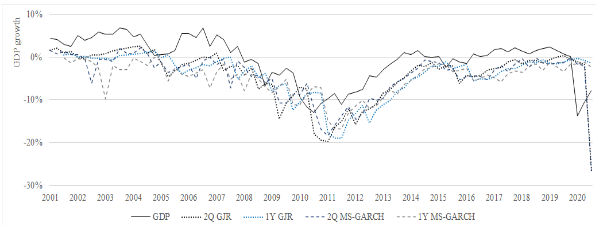
(h) ESP



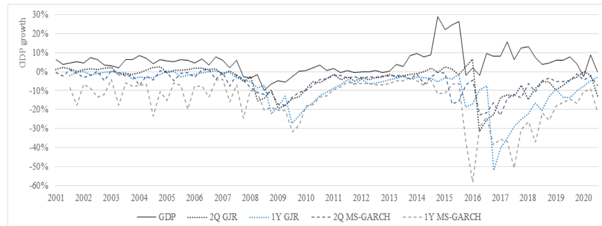
(i) FRA



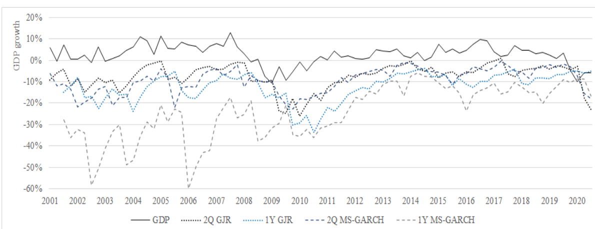
(j) GBR



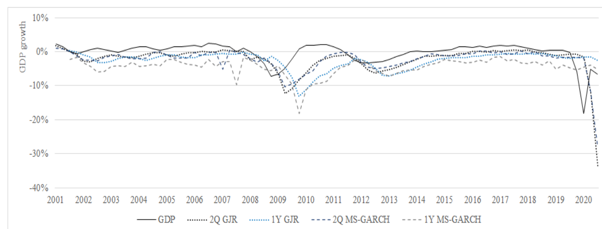
(k) GRC



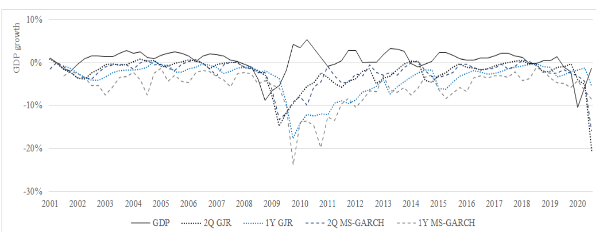
(l) IRL



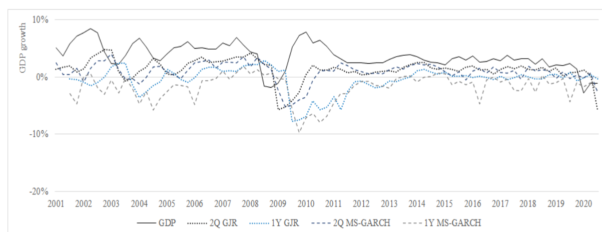
(m) ISL



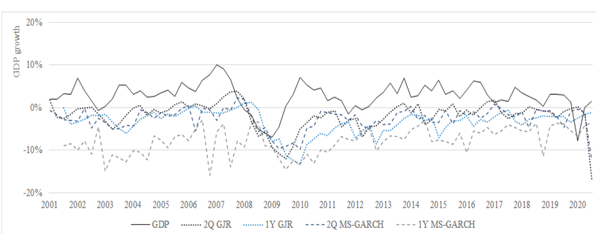
(n) ITA



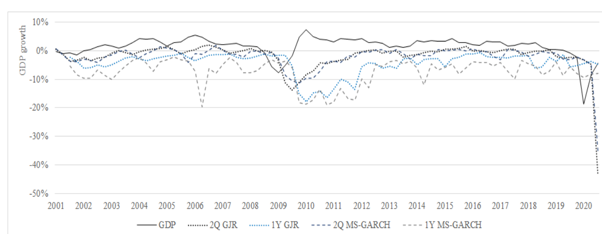
(o) JPN



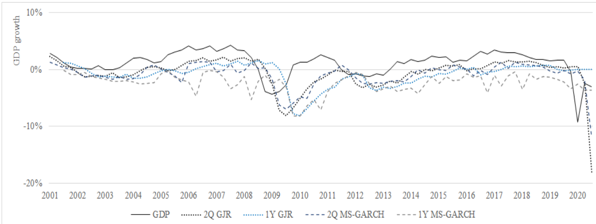
(p) KOR



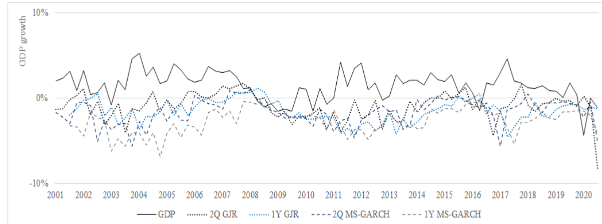
(q) LUX



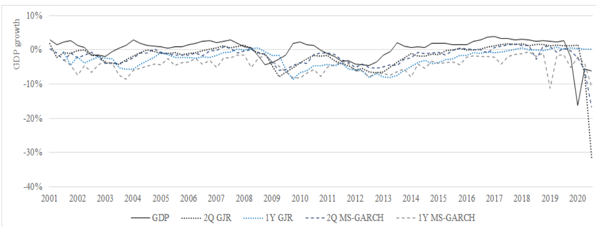
(r) MEX



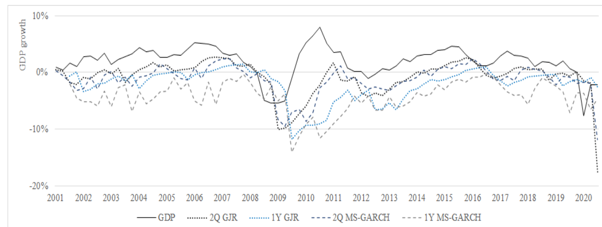
(s) NLD



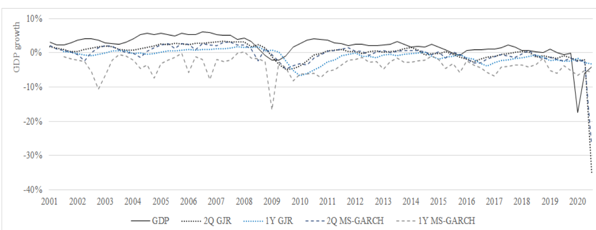
(t) NOR



(u) PRT



(v) SWE



(w) ZAF

Figure 14. This figure plots the realized GDP growth rates alongside the two-quarter ahead and one-year ahead 95% GaR forecasts of the GJR-GARCH(1,1) and MS-GARCH(1,1) models. Subfigures (14a)-(14w) provide the graphs for all countries in the 100%-sample, except DEU and USA. In the figures, the solid black lines plot the realized GDP growth rates of the countries. Moreover, the dotted black and dotted blue lines reflect the two-quarter ahead and one-year ahead 95% GaR forecasts of the GJR model, respectively, and the dashed black and dashed gray lines reflect the two-quarter ahead and one-year 95% GaR forecasts of the MS-GARCH model. The models are estimated on estimation windows from the 100%-sample that comprises 25 OECD countries. The out-of-sample periods of the two-quarter ahead and one-year ahead GaR forecasts range from 2001Q2 to 2020Q4 and 2001Q4 to 2020Q4, respectively.

Appendix H Analyses on the 80%-sample

To improve the structure of the appendix section of this study, this appendix comprises all supplementary results to the analyses performed on models estimated on estimation windows from the 80%-sample. The appendix sections of this appendix supplement various parts of the in-sample and out-of-sample analyses of this study. First, Appendix Section H.A provides the shape parameters per country of the model estimations in the in-sample analysis. Next, Appendix Section H.B contains the recession probabilities of the MS-GARCH model. To supplement the out-of-sample analysis, Appendix Section H.C plots the scoring functions based on short-horizon GaR forecasts, aggregated over the countries. Moreover, Appendix Section H.D contains plots of short-horizon GaR forecasts generated from models estimated on estimation windows from the 80%-sample. Appendix Section H.E provides the aggregated scoring functions over countries based on long-horizon GaR forecasts. Finally, Appendix Section H.F contains plots of the long-horizon GaR forecasts.

A Shape parameters

First, Tables IV and V below provide the shape parameters per country of the in-sample estimations of the single-regime and regime-switching models, respectively, in Section V.A.2.

Table XXI: Estimates of shape parameters per country in single-regime models

	GARCH		EGARCH		GJR	
	Coefficient	Std. Error	Coefficient	Std. Error	Coefficient	Std. Error
AUS	83.401	(467.654)	247.565	(4448.807)	290.907	(5559.813)
AUT	5.826***	(1.883)	5.854***	(1.957)	6.059***	(2.081)
BEL	4.231***	(0.825)	4.331***	(0.897)	4.278***	(0.849)
CAN	8.109***	(3.039)	5.551***	(1.173)	8.528**	(3.343)
CHE	6.537***	(1.889)	5.828***	(1.504)	6.465***	(1.852)
DEU	5.239***	(1.339)	4.898***	(1.149)	5.252***	(1.339)
DNK	13.521	(10.688)	14.981	(15.965)	12.984	(10.032)
ESP	4.075***	(0.749)	4.872***	(1.271)	4.145***	(0.789)
FRA	3.425***	(0.420)	3.591***	(0.485)	3.421***	(0.416)
GBR	4.134***	(0.775)	4.309***	(0.890)	4.124***	(0.767)
IRL	6.057***	(1.110)	5.990***	(1.104)	5.996***	(1.053)
ITA	9.205	(5.641)	7.541**	(3.585)	9.117*	(5.500)
JPN	7.122***	(2.092)	6.654***	(2.208)	7.194***	(2.093)
MEX	6.342***	(1.535)	5.996***	(1.374)	6.517***	(1.545)
NLD	5.847***	(1.824)	6.143***	(2.049)	5.817***	(1.795)
NOR	15.616	(12.594)	19.208	(22.286)	15.653	(12.105)
PRT	12.302	(8.918)	12.528	(9.532)	12.934	(9.528)
SWE	10.773**	(4.574)	10.927*	(5.742)	10.772**	(4.564)
USA	5.380***	(1.579)	5.764***	(1.904)	5.546***	(1.707)
ZAF	339.301	(8734.366)	339.541	(10167.754)	339.221	(8796.684)

This table reports the estimates of the shape parameters of the conditional Student's t-distribution of GDP growth in the single-regime GARCH(1,1), EGARCH(1,1) and GJR-GARCH(1,1) models. The table contains all country-specific shape parameters ν_i for all countries $i = 1, \ldots, 20$ in the 80%-sample. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The in-sample data in the 80%-sample consists of quarterly GDP growth rates from 1962Q1 to 2019Q4 for CAN and from 1961Q1 to 2019Q4 for the remaining 19 OECD countries.

The multi-state framework allows the shape parameters of the conditional GDP growth distribution to differ across the states of the economy. Below, Table XXII provides the shape parameters per regime and per country of the estimation of the MS-GARCH model in Table V.

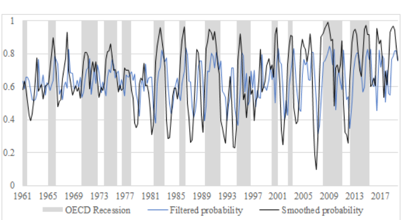
Table XXII: Estimates of shape parameters per country in the MS-GARCH model

	Expansionary regime ($k = 1$)		Recessionary regime ($k = 2$)	
	Coefficient	Std. Error	Coefficient	Std. Error
AUS	61.281	(717.127)	9.354	(8.851)
AUT	2.817***	(0.494)	5.313***	(1.927)
BEL	3.660***	(1.149)	3.769***	(0.829)
CAN	3.201***	(0.742)	10.756	(10.138)
CHE	5.177	(3.344)	5.105***	(1.593)
DEU	6.924	(6.969)	3.971***	(0.889)
DNK	38.084	(280.619)	11.532	(11.396)
ESP	5.508	(4.417)	4.945**	(1.921)
FRA	2.850***	(0.461)	2.944***	(0.338)
GBR	3.391***	(0.907)	3.456***	(0.620)
IRL	55.227	(353.475)	3.783***	(0.737)
ITA	13.139	(30.352)	5.447**	(2.149)
JPN	61.080	(591.35)	4.731***	(1.460)
MEX	6.749	(5.162)	4.941***	(1.470)
NLD	4.853	(3.007)	4.593***	(1.489)
NOR	4.521**	(1.960)	17.357	(30.457)
PRT	4.162***	(1.580)	22.441	(44.517)
SWE	8.098	(10.241)	9.214*	(5.106)
USA	8.763	(10.501)	3.508***	(0.653)
ZAF	25.117	(119.424)	51.941	(314.323)

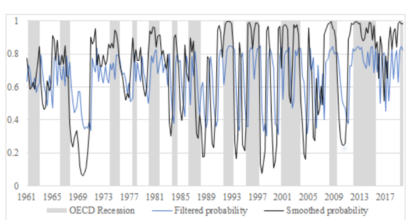
This table reports the estimates of the shape parameters of the conditional Student's t-distribution of GDP growth in the MS-GARCH(1,1) model. The table contains all country-specific shape parameters $\nu_i^{(k)}$ for both regimes $k = 1, 2$ and for all countries $i = 1, \ldots, 20$ in the 80%-sample. Asymptotic standard errors are in parentheses. '***', '**' and '*' denote significance at the 0.01, 0.05, and 0.10 level, respectively. The in-sample data in the 80%-sample consists of quarterly GDP growth rates from 1962Q1 to 2019Q4 for CAN and from 1961Q1 to 2019Q4 for the remaining 19 OECD countries.

B Recession probabilities

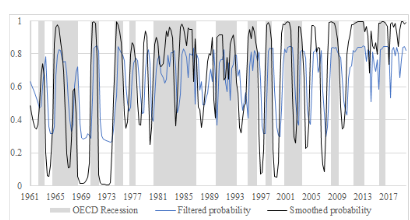
Second, this section, Appendix Section H.B of Appendix H, provides the filtered and smoothed recession probabilities for 17 OECD countries in the 80%-sample. Figure 15 supplements the recession probabilities of DEU, JPN and USA in Figure 5 in Section V.A.2.



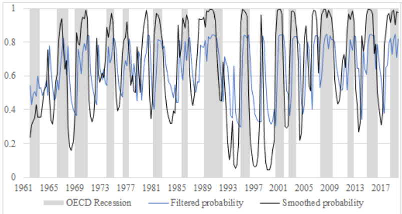
(a) AUS



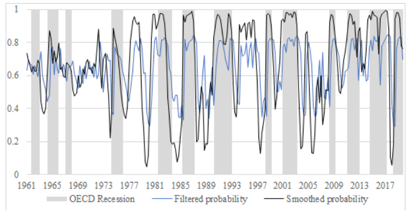
(b) AUT



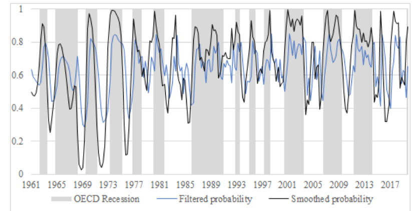
(c) BEL



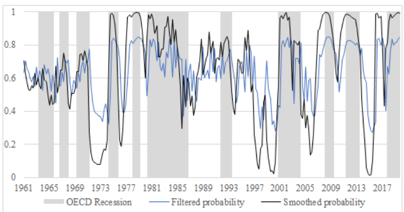
(d) CAN



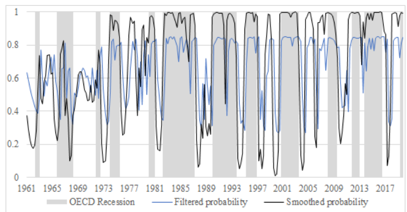
(e) CHE



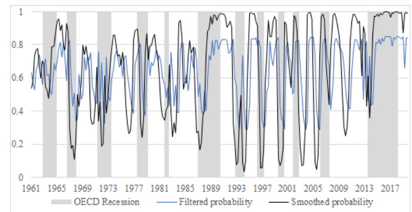
(f) DNK



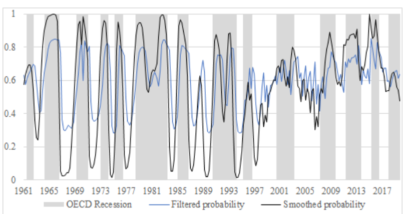
(g) ESP



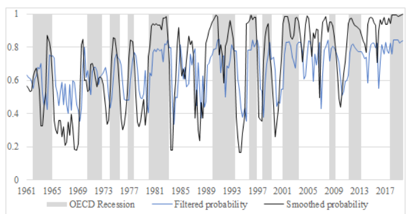
(h) FRA



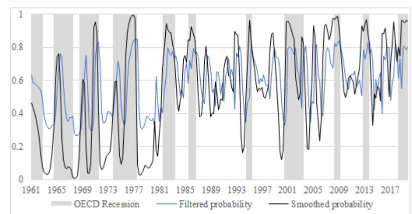
(i) GBR



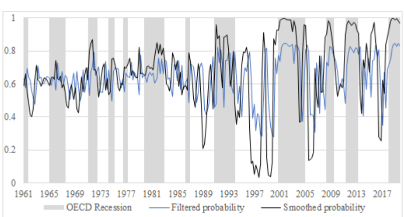
(j) IRL



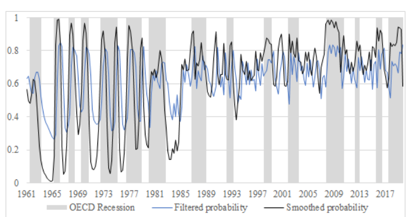
(k) ITA



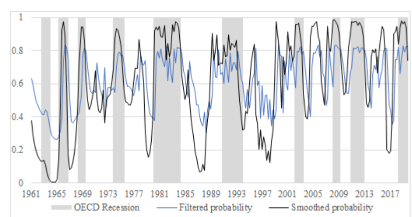
(l) MEX



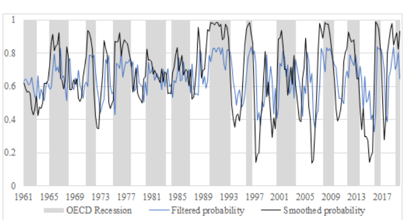
(m) NLD



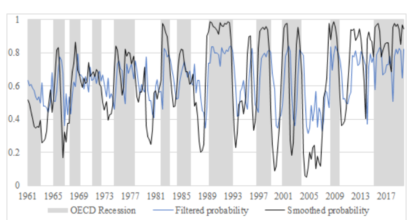
(n) NOR



(o) PRT



(p) SWE



(q) ZAF

Figure 15. This figure plots the filtered (in blue) and smoothed (in black) recession probabilities of the 20 countries in the 80%-sample, except DEU, JPN, and USA. The filtered and smoothed recession probabilities range from 1962Q1 to 2019Q4 for CAN and from 1961Q1 to 2019Q4 for the remaining OECD countries. OECD recession periods are highlighted in gray.

C Scoring functions based on short-horizon GaR forecasts

Third, this section, Appendix Section H.C of Appendix H, provides the plots of the scoring functions, aggregated over countries, over the evaluation window. The scoring functions are based on the one-quarter ahead GaR forecasts from the GARCH, GJR and MS-GARCH models, estimated on moving estimation windows from the 80%-sample. Based on the underperformance of the EGARCH model, as discussed in Section V.B.1, Figure 16 below omits the scoring function of the EGARCH model.

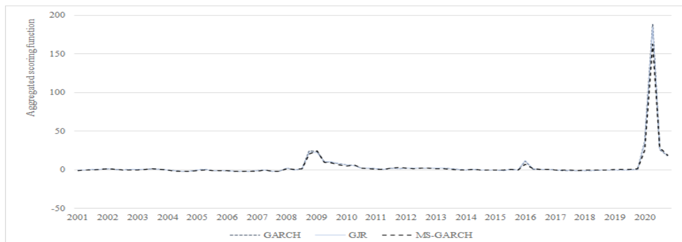
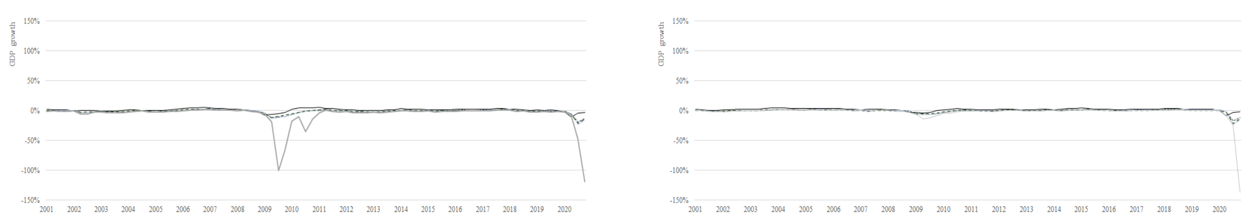


Figure 16. This figure plots the comparative scoring function of one-quarter ahead GaR forecasts, aggregated over the countries in the sample. The aggregated comparative scoring function is given by $\sum_{i=1}^{n} S_{i,t} = \sum_{i=1}^{n} \big(\mathbb{I}[Y_{i,t} < GaR_{i,t|t-1}(c)] + c - 1\big)\big(GaR_{i,t|t-1}(c)\big)\big) - \mathbb{I}[Y_{i,t|t-1} < GaR_{i,t|t-1}(c)]\big(Y_{i,t}\big)$, where the expression of $S_{i,t}$ is obtained from Equation (20). The dotted dark blue, solid light blue and dashed black lines in the figures represent the aggregated scoring functions of the GARCH, GJR and MS-GARCH models, respectively. The models are estimated on estimation windows from the 80%-sample. The out-of-sample period ranges from 2001Q1 to 2020Q4. The sums of the aggregated scoring functions of the GARCH, GJR and MS-GARCH models, over the out-of-sample period, equal 310.550, 305.254 and 267.656, respectively.

D Short-horizon GaR forecasts

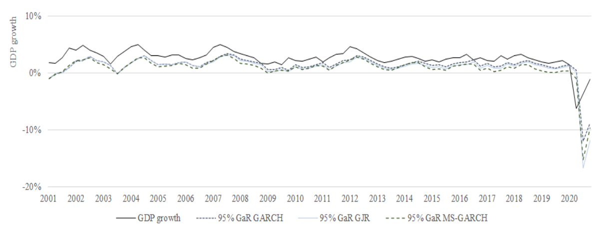
Fourth, this section, Appendix Section H.D of Appendix H, provides the short-horizon GaR forecasts. The forecasts are generated from models estimated on moving estimation windows from the 80%-sample. First, to illustrate the extremity of GaR forecasts generated from the EGARCH model, Figure 17 shows the course of the realized GDP growth rates of DEU and USA alongside the one-quarter ahead GaR forecasts of all models.



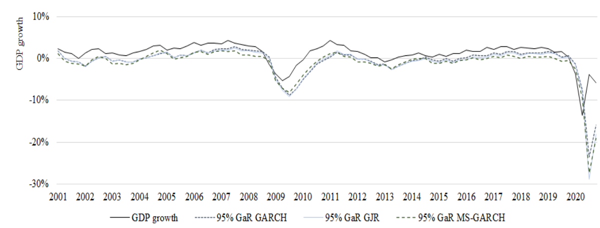
(a) DEU **(b)** USA

Figure 17. This figure plots the realized GDP growth rates alongside the one-quarter ahead 95% GaR forecasts of the GARCH(1,1), EGARCH(1,1), GJR-GARCH(1,1) and MS-GARCH(1,1) models for all countries in the 80%-sample. The solid black, dotted dark blue, solid gary, solid light blue, and dashed black lines reflect the realized GDP growth rates and the forecasts of the GARCH, EGARCH, GJR and MS-GARCH models, respectively. The models are estimated on estimation windows from the 80%-sample that comprises 20 OECD countries. The out-of-sample evaluation window ranges from 2001Q1 to 2020Q4.

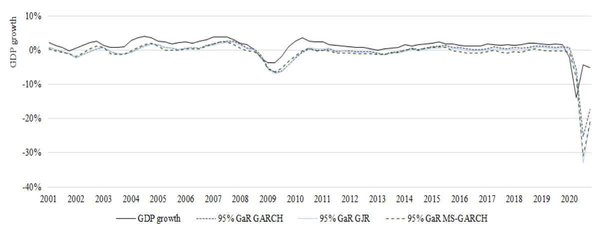
Moreover, Figure 18 below plots the course of the realized GDP growth rates alongside the one-quarter ahead GaR forecasts of the GARCH, GJR and MS-GARCH models.



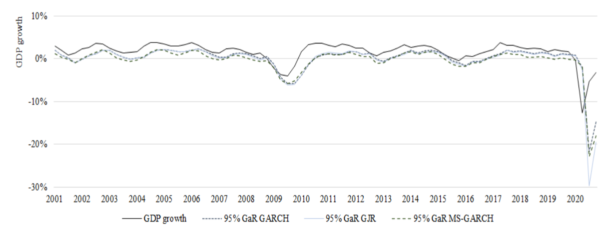
(a) AUS



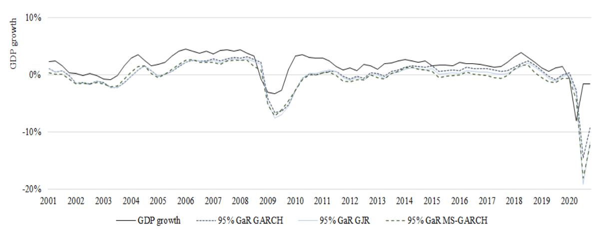
(b) AUT



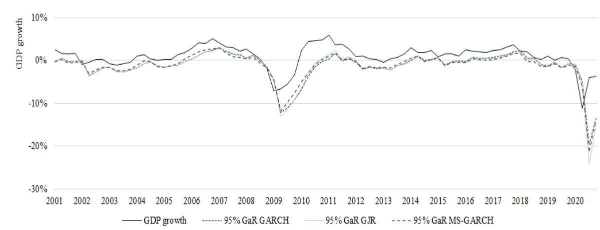
(c) BEL



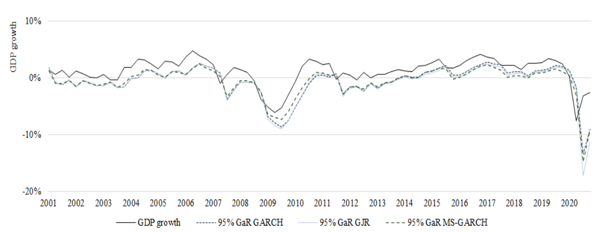
(d) CAN



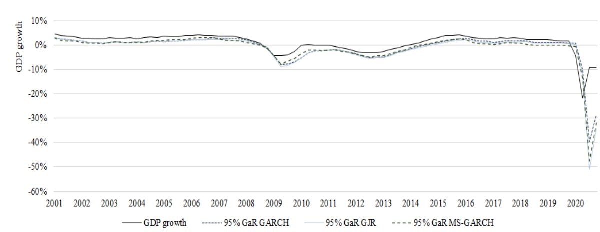
(e) CHE



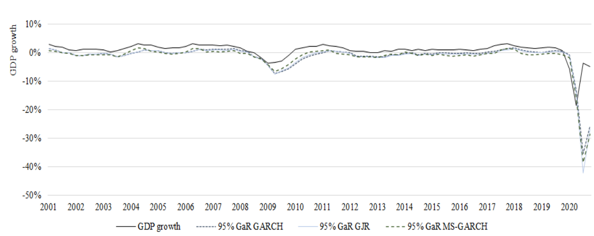
(f) DEU



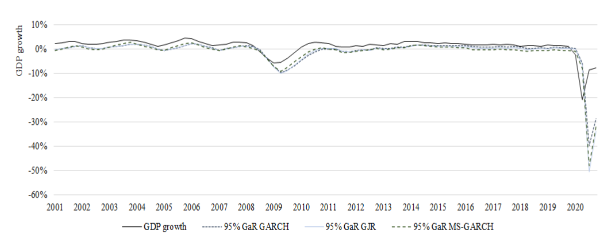
(g) DNK



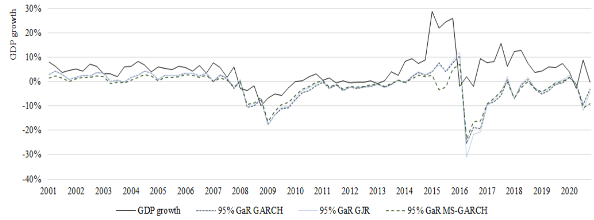
(h) ESP



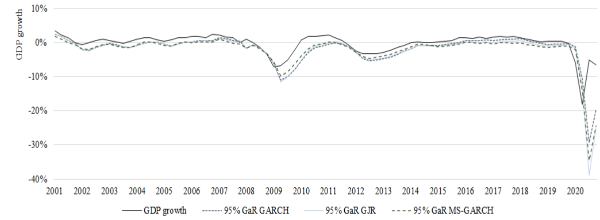
(i) FRA



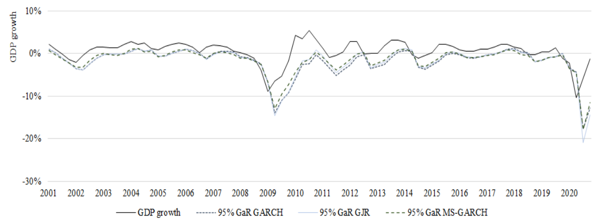
(j) GBR



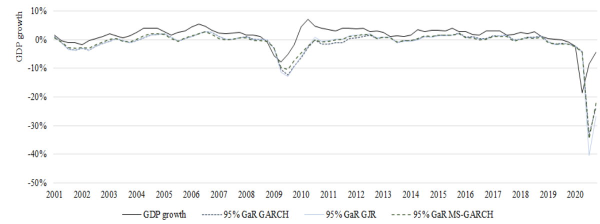
(k) IRL



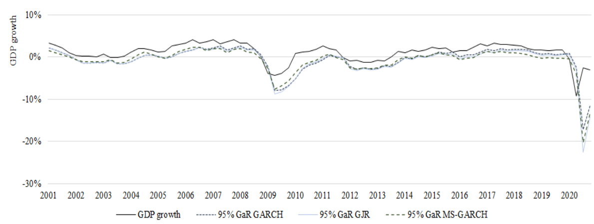
(l) ITA



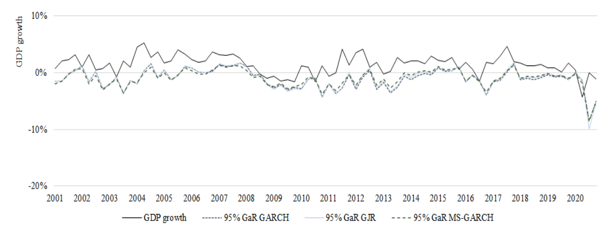
(m) JPN



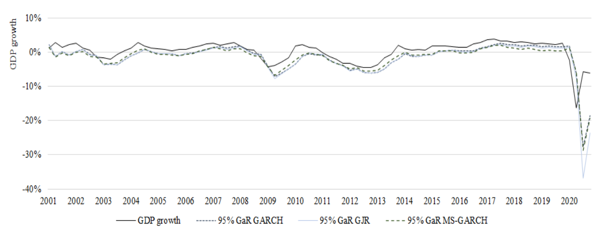
(n) MEX



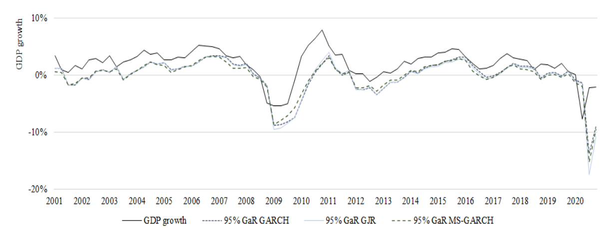
(o) NLD



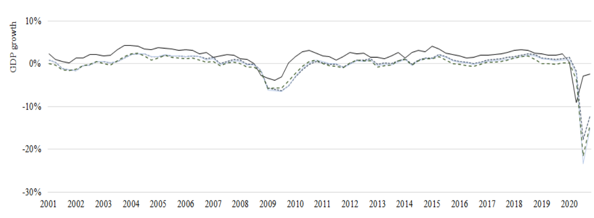
(p) NOR



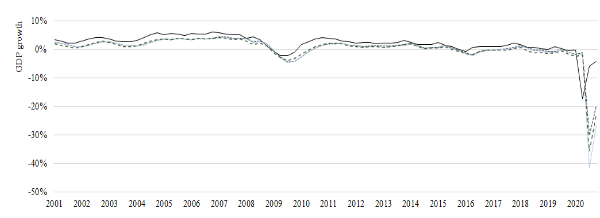
(q) PRT



(r) SWE



(s) USA

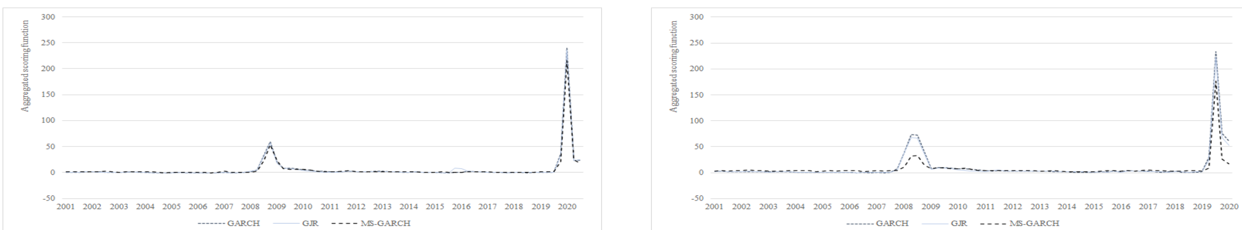


(t) ZAF

Figure 18. This figure plots the realized GDP growth rates alongside the one-quarter ahead 95% GaR forecasts of the GARCH(1,1), GJR-GARCH(1,1) and MS-GARCH(1,1) models. Subfigures (18a)-(18t) provide the graphs for all 20 OECD countries in the 80%-sample. In the figures, the solid black lines plot the realized GDP growth rates of the countries. Moreover, the dotted dark blue, solid light blue and dashed black lines reflect the one-quarter ahead 95% GaR forecasts of the GARCH, GJR and MS-GARCH models, respectively. The models are estimated on estimation windows from the 80%-sample that comprises 20 OECD countries. The out-of-sample periods range from 2001Q1 to 2020Q4.

E Scoring functions based on long-horizon GaR forecasts

Fifth, this section, Appendix Section H.E of Appendix H, provides the plots of the scoring functions, aggregated over countries, over the evaluation window. The scoring functions are based on the two-quarter ahead (Figure 19a) and one-year ahead (Figure 19b) GaR forecasts from the GARCH, GJR and MS-GARCH models, estimated on moving estimation windows from the 80%-sample. Based on the underperformance of the EGARCH model, as discussed in Section V.B.2, Figure 16 omits the scoring function of the EGARCH model. The out-of-sample evaluation windows of the two-quarter ahead and one-year ahead GaR forecasts range from 2001Q2 to 2020Q4 and from 2001Q4 to 2020Q4, respectively.



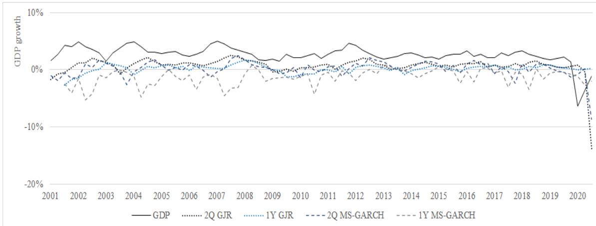
(a) Based on two-quarter ahead GaR forecasts

(b) Based on one-year ahead GaR forecasts

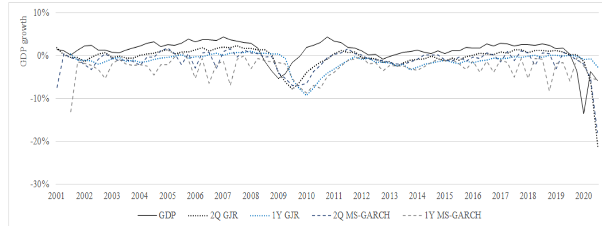
Figure 19. This figure plots the comparative scoring function of two-quarter ahead (Figure 19a) and one-year ahead (Figure 19b) GaR forecasts, aggregated over the countries in the sample. The aggregated comparative scoring function is given by $\sum_{i=1}^{n} S_{i,t} = \sum_{i=1}^{n} \big(\mathbb{I}[Y_{i,t} < GaR_{i,t|t-1}(c)] + c - 1\big)\big(GaR_{i,t|t-1}(c)\big)\big) - \mathbb{I}[Y_{i,t|t-1} < GaR_{i,t|t-1}(c)]\big(Y_{i,t}\big)$, where the expression of $S_{i,t}$ is obtained from Equation (20). The dotted dark blue, solid light blue and dashed black lines in the figures represent the aggregated scoring functions of the GARCH, GJR and MS-GARCH models, respectively. The models are estimated on estimation windows from the 80%-sample. The out-of-sample periods of the two-quarter ahead and one-year ahead GaR forecasts range from 2001Q2 to 2020Q4 and from 2001Q4 to 2020Q4, respectively.

F Long-horizon GaR forecasts

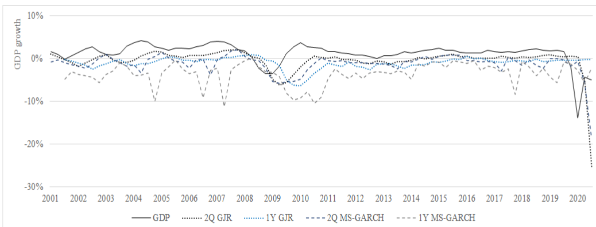
Finally, this section, Appendix Section H.F of Appendix H, provides the plots of long-horizon GaR forecasts of the models estimated on moving estimation windows from the 80%-sample. Below, Figure 20 shows the course of the realized GDP growth rates alongside the two-quarter and one-year ahead GaR forecasts of the MS-GARCH model and the GJR model. In the figures, the solid black lines plot the realized GDP growth rates. Moreover, the dotted black and dotted blue lines reflect the two-quarter ahead and one-year ahead 95% GaR forecasts of the GJR model, respectively, and the dashed black and dashed gray lines reflect the two-quarter ahead and one-year 95% GaR forecasts of the MS-GARCH model. The out-of-sample evaluation windows of the two-quarter ahead and one-year ahead GaR forecasts range from 2001Q2 to 2020Q4 and from 2001Q4 to 2020Q4, respectively.



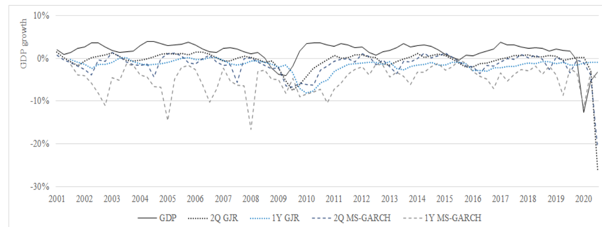
(a) AUS



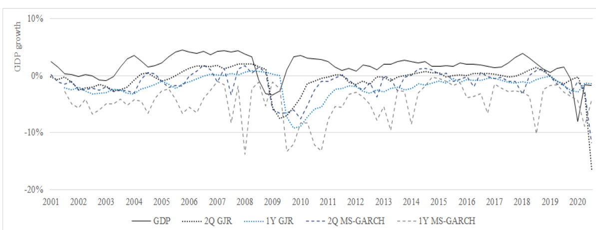
(b) AUT



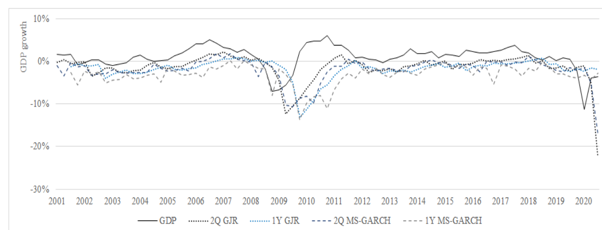
(c) BEL



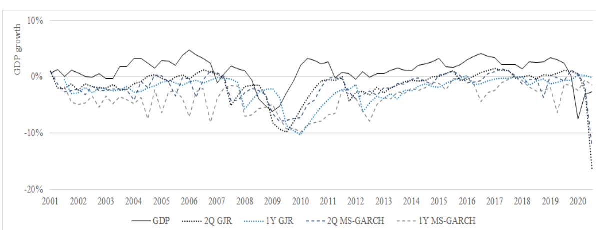
(d) CAN



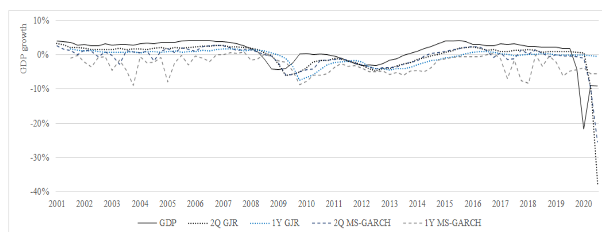
(e) CHE



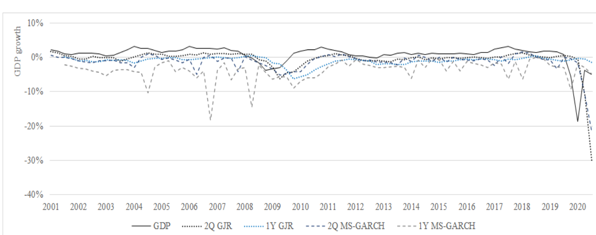
(f) DEU



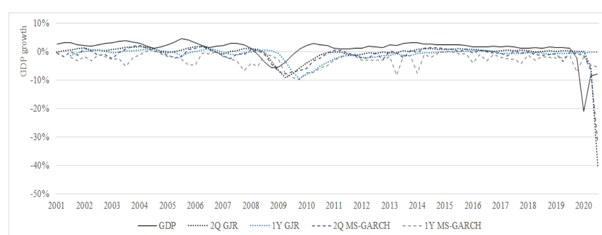
(g) DNK



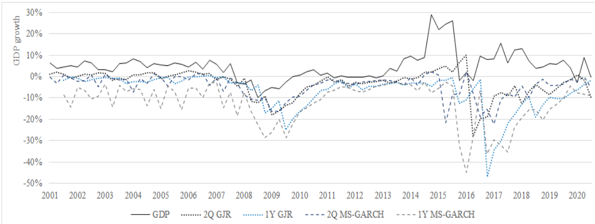
(h) ESP



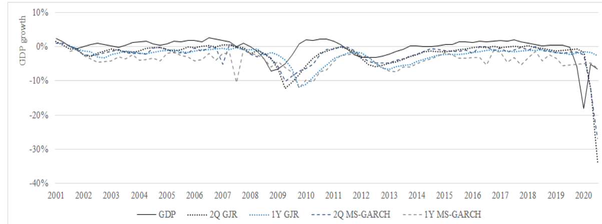
(i) FRA



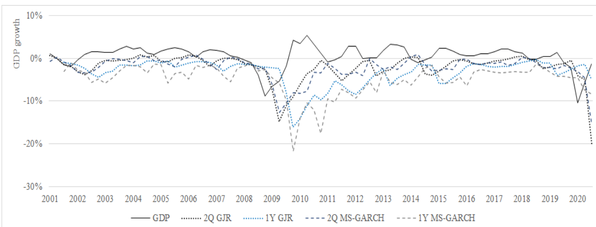
(j) GBR



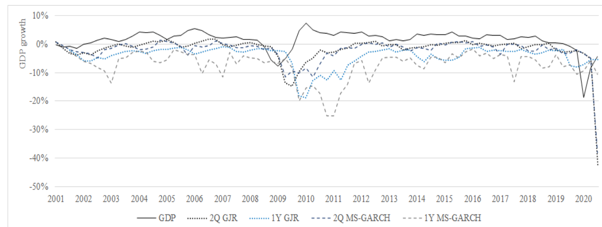
(k) IRL



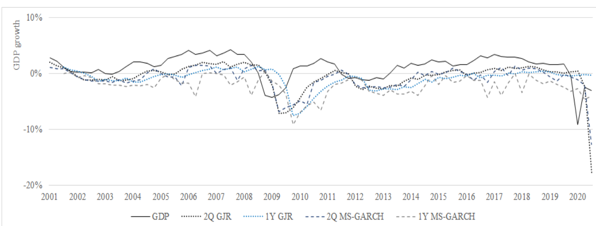
(l) ITA



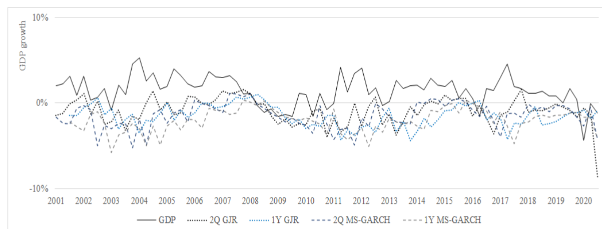
(m) JPN



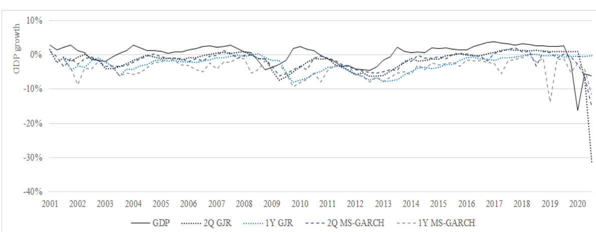
(n) MEX



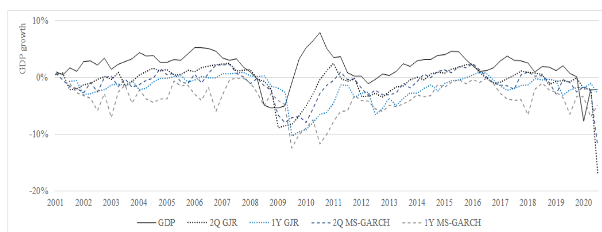
(o) NLD



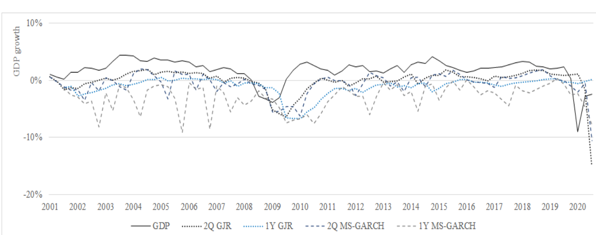
(p) NOR



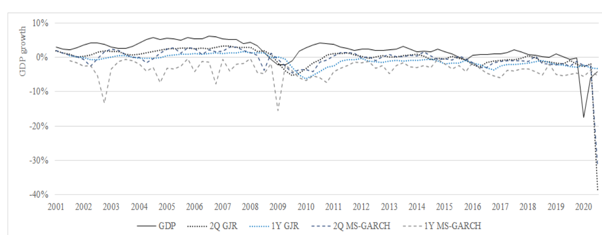
(q) PRT



(r) SWE



(s) USA



(t) ZAF

Figure 20. This figure plots the realized GDP growth rates alongside the two-quarter ahead and one-year ahead 95% GaR forecasts of the GJR-GARCH(1,1) and MS-GARCH(1,1) models. Subfigures (20a)-(20t) provide the graphs for all 20 OECD countries in the 80%-sample. In the figures, the solid black lines plot the realized GDP growth rates of the countries. Moreover, the dotted black and dotted blue lines reflect the two-quarter ahead and one-year ahead 95% GaR forecasts of the GJR model, respectively, and the dashed black and dashed gray lines reflect the two-quarter ahead and one-year 95% GaR forecasts of the MS-GARCH model. The models are estimated on estimation windows from the 80%-sample that comprises 20 OECD countries. The out-of-sample periods of the two-quarter ahead and one-year ahead GaR forecasts range from 2001Q2 to 2020Q4 and 2001Q4 to 2020Q4, respectively.

Appendix I GDP growth during expansionary and recessionary periods

This appendix provides statistics on GDP growth during expansionary and recessionary periods between 2001Q1 and 2020Q4. To classify quarters as expansionary or recessionary periods, we rely on country-specific recession indicators obtained from the OECD database. The OECD relies on country-specific component variables to identify identifies months of macroeconomic turning points that indicate the beginning of a recession. Components that the OECD takes into account relate to national share price developments, the national balance of trade, policies of monetary institutions, demographic patterns and other macroeconomic statistics. Based on a simplified version of the Bry and Boschan algorithm (Bry & Boschan (1971)), the OECD detects macroeconomic turning points by combining de-trended smoothed and normalized component series. The monthly classification is based on the 'peak method', which implies that the classified recession includes the peak in GDP growth and excludes the trough in GDP growth. To translate the monthly observations into quarterly data, we consider the third month of each quarter as identifier of the macroeconomic state of the quarter.

Table XXIII below shows the dominance of expansionary quarters in the evaluation sample from 2001Q1 to 2020Q4. Over the past 20 years, recessions have been most scarce in ZAF, where 25% of the quarters is classified as a recession, while the economy of IRL experienced most recessionary quarters (55% of the quarters). Moreover, Table XXIII shows that, on average, recessionary periods exhibit lower and more volatile GDP growth. For almost all countries, the average GDP growth during expansionary periods exceeds the average GDP growth during recessionary periods. The exception to this rule is GBR, where the average economic growth is higher during quarters that the OECD classifies as recessionary.

Table XXIII: GDP growth during recessionary and expansionary periods

	Expansionary periods		Recessionary periods	
Country	No. of quarters	Avg. GDP growth	No. of quarters	Avg. GDP growth
AUS	51	2.92	29	2.02
AUT	45	2.16	35	−0.15
BEL	52	1.78	28	0.11
CAN	45	2.38	35	0.66
CHE	39	2.55	41	0.81
DEU	34	2.35	46	−0.08
DNK	42	2.23	38	−0.11
ESP	41	1.93	39	0.12
FIN	41	2.77	39	−0.39
FRA	43	1.67	37	−0.18
GBR	50	0.72	30	1.81
GRC	49	0.97	31	−2.76
IRL	36	7.01	44	2.86
ISL	47	3.54	33	1.29
ITA	47	0.59	33	−1.42
JPN	39	1.31	41	−0.29
KOR	47	4.24	33	2.74
LUX	47	3.86	33	0.9
MEX	53	2.53	27	−0.74
NLD	41	2.04	39	0.21
NOR	44	2.28	36	0.42
PRT	52	1.32	28	−1.5
SWE	45	3.21	35	0.28
USA	42	2.36	38	1.03
ZAF	60	2.45	20	1.92
Average	45	2.45	35	0.38
Std. Dev.		2.92		3.62

This table reports statistics of GDP growth during expansionary and recessionary periods between 2001Q1 and 2020Q4, based on OECD recession indicators. The 'peak method' OECD recession indicators are obtained from the OECD database. Quarterly GDP growth rates in the table are in percentages.

Appendix J BJPR-approach for joint GaR forecasting

This appendix provides the pseudo-code to construct one-quarter ahead joint GaR forecasts, as described in Section VI, according to the BJPR-approach of Wolf & Wunderli (2015). The algorithm is an extension of the single-regime Joint GaR algorithm described by Brownlees & Souza (2021). Specifically, the algorithm details the construction of the one-quarter ahead GaR forecast of country i, forecasted at forecasting quarter t, with coverage level r and conditional probability level c.

Result: the one-quarter ahead joint GaR forecast of country i, $GaR^{J}_{i,t+1|t}(r,c)$

INPUTS ;

Require: $\mathbf{Y}$: $(n \times \tau_1)$ matrix of GDP growth rates $Y_{i,t}$;

Require: $\mathbf{H}$: $(n \times \tau_1)$ matrix of conditional means $\mu_{i,t}$;

Require: $\mathbf{M}$: $(n \times \tau_1)$ matrix of conditional variances $h_{i,t}$;

Require: Ψ: set of model parameters ;

Require: r: coverage level ;

Require: c: quantile of interest ;

Require: M: number of bootstrap replications ;

PROCEDURE ;

if *Single-regime model* **then**

  1. Construct the $(n \times \tau_1)$ matrix of residuals $\hat{Z}$ with entries (i,t) given by $\hat{Z}_{i,t} = (Y_{i,t} - \mu_{i,t})/(h_{i,t})^{1/2}$;

else if *Regime-switching model* **then**

  1. Construct the $(n \times \tau_1)$ matrices of residuals $\hat{Z}^{(k)}$ with entries (i,t) given by $\hat{Z}^{(k)}_{i,t} = (Y_{i,t} - \mu^{(k)}_{i,t})/(h^{(k)}_{i,t})^{1/2}$;

2. Construct the $(M \times 1)$ vector B with entry (m), b_m, given by a uniform draw from the integer interval $[1,\ldots,\tau_1]$;

3. **for** i *in* $1,\ldots,n$ **do**

  for m *in* $1,\ldots,M$ **do**

    if *Single-regime model* **then**

      Construct the bootstrap innovation $\tilde{Z}^{\{m\}}_{i} = \hat{Z}_{i,b_m}$;

      $\mu_{i,t+1} = \phi_0 + \phi_1 Y_{i,t}$;

      if *GARCH* **then**

        $h_{i,t+1} = \omega_i + \alpha_i \epsilon^2_{i,t} + \beta_i h_{i,t}$;

      else if *EGARCH* **then**

        $h_{i,t+1} = \exp\{\omega + \alpha \frac{|\epsilon_{i,t}|}{|h_{i,t}|} + \gamma \frac{\epsilon_{i,t}}{h_{i,t}} + \beta \ln h_{i,t}\}$;

      else if *GJR* **then**

        $h_{i,t+1} = \omega + \alpha\epsilon^2_{i,t}(1 - \mathbb{I}[\epsilon_{i,t} > 0]) + \gamma\epsilon^2_{i,t}\mathbb{I}[\epsilon_{i,t} > 0] + \beta h_{i,t}$;

      $\tilde{Y}^{\{m\}}_{i,t+1} = \mu_{i,t+1} + (h_{i,t+1})^{1/2}\tilde{Z}^{\{m\}}_{i}$;

    else if *Regime-switching model* **then**

      for $k = 1, 2$ **do**

        $\mu^{(k)}_{i,t+1} = \phi^{(k)}_0 + \phi^{(k)}_1 Y_{i,t}$;

        $h^{(k)}_{i,t+1} = \omega^{(k)} + \alpha^{(k)}\epsilon^2_{i,t} + \beta^{(k)} h_{i,t}$,

          where $h_{i,t}$ is computed as the Klaassen (2002) expectation term in Equation (11)

      end

      Simulate $\tilde{U}^{\{m\}}_{i,t+1} \sim \text{Unif}(0,1)$;

      if $U^{\{m\}}_{i,t+1} \leq \pi^{(1)\{m\}}_{i,t+1}$ **then**

        $\tilde{Z}^{(1)\{m\}}_{i} = \hat{Z}^{(1)}_{i,b_m}$;

        $\tilde{Y}^{\{m\}}_{i,t+1} = \mu^{(1)}_{i,t+1} + (h^{(1)}_{i,t+1})^{1/2}\tilde{Z}^{(1)\{m\}}_{i}$

      else if $U^{\{m\}}_{i,t+1} > \pi^{(1)\{m\}}_{i,t+1}$ **then**

        $\tilde{Z}^{(2)\{m\}}_{i} = \hat{Z}^{(2)}_{i,b_m}$;

        $\tilde{Y}^{\{m\}}_{i,t+1} = \mu^{(2)}_{i,t+1} + (h^{(2)}_{i,t+1})^{1/2}\tilde{Z}^{(2)\{m\}}_{i}$

  end

end

4. Construct the $(M \times n)$ matrix $\tilde{\mathbf{O}}$ with entry (m,i) given by $\rtimes_{m,i} = (\tilde{Y}^{\{m\}}_{i,t+1} - \hat{\mu}_{i,t+1})/(\hat{h}_{i,t+1})^{1/2}$,
where $\hat{\mu}_{i,t+1}$ and $\hat{h}_{i,t+1}$ denote the sample mean and sample variance of $\tilde{Y}_{i,t+1}$ across the M simulations ;

5. Define U^{r*n}_{m} as the $(r*n)$-smallest element of $\{\rtimes_{m,i}\}^{n}_{i=1}$;

6. Compute d^{r*n}_{c} as the empirical c-quantile of $\{U^{r*n}_{m}\}^{M}_{m=1}$;

7. Construct the joint prediction region $GaR^{J}_{i,t+1|t}(r,c) = \hat{\mu}_{i,+1} + d^{r*n}_{c}(\hat{h}_{i,t+1})^{1/2}$;

Algorithm 2: Joint GaR algorithm

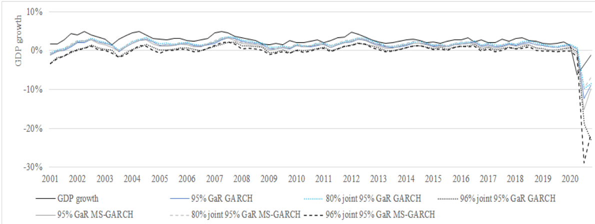
Appendix K Joint GaR forecasts

This appendix provides the plots of the joint GaR forecasts of the GARCH model and the MS-GARCH model. Moreover, this appendix K elaborates on the impact of the coverage level r on the height and accuracy of the joint GaR forecast, and identifies a suspected reason for the higher joint GaR prediction accuracy of the MS-GARCH model. On the next page, Figure 21 plots the realized GDP growth rates alongside the one-year ahead 80% joint and 96% joint GaR forecasts of the MS-GARCH model and the GARCH model. The figures omit the joint GaR forecasts of the GJR model based on the lower accuracy in terms of unconditional coverage. The out-of-sample period of the one-quarter ahead joint GaR forecasts ranges from 2001Q1 to 2020Q4.

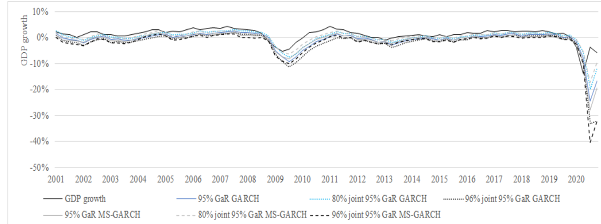
First, the figures visualize the decay of the joint GaR forecasts with increases of the coverage level r. High coverage levels force the models to build in security, which results into lower joint GaR forecasts. Conversely, low coverage levels allow for a number of overshootings in the sample. Consequently, the models are allowed to ignore the largest falls in GDP growth of individual countries and produce higher joint GaR forecasts. During unstable macroeconomic periods, the behaviour of the joint GaR forecasts with varying coverage levels diverges. The decreases of different magnitudes of the joint GaR forecasts in 2020Q3 as a response to the Covid-outbreak in 2020Q2 illustrate this divergence.

Second, the figures show that, for lower values of r, joint GaR forecasts tend to represent the average GDP growth dynamics across the sample. The dynamics around the Great Financial Crisis of the one-quarter ahead 80% joint 95% GaR illustrate the moderateness of lower r joint GaR forecasts. On average, the sample of 25 OECD countries experiences the peak of the Global Financial Crisis in 2009Q2. In DEU, the crisis hit early, with a shrink of the economy by 7.0% in 2009Q1. The larger kink of the one-quarter ahead joint GaR forecasts of DEU in 2009Q3 reflects the adaption of lower r joint GaR forecasts to average GDP growth dynamics in the sample. In opposite to the moderate behavior of lower r joint GaR forecasts, higher r joint GaR forecasts appear more extreme reflections of the country-specific GDP growth dynamics. The different extent to which lower r and higher r joint GaR forecasts account for GDP growth dynamics of other countries in the sample explains the higher forecasting accuracies of higher r joint GaR forecasts.

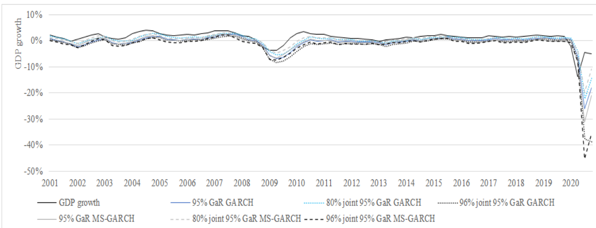
Third, after the peak of the Covid crisis in 2020Q2, the 96% joint 95% GaR forecasts of the single-regime models remain in 2020Q4 at the extraordinary levels of 2020Q3, or decrease even further. Apparently, the extreme decreases of GDP growth rates in 2020Q2, combined with the high security required to contain the GDP growth rates of 24 OECD countries at a probability of 95%, cause single-regime models to produce similarly low or lower 96% joint GaR forecasts in 2020Q4. By contrast, switching between regimes enables the MS-GARCH model to recognize the transiency of the short peak of the Covid crisis. This capability of the MS-GARCH model might explain the higher joint GaR forecasting accuracy of the MS-GARCH model. As this study focuses on marginal GaR forecasting, we consider the confirmation of the suspected driver of the possible outperformance of the MS-GARCH model in joint GaR forecasting as beyond the scope of this study.



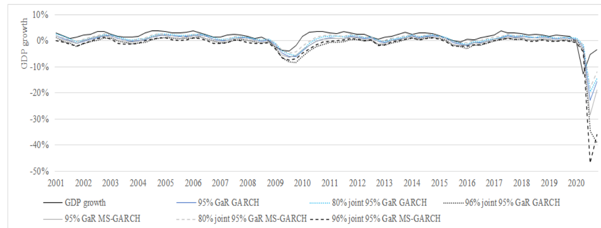
(a) AUS



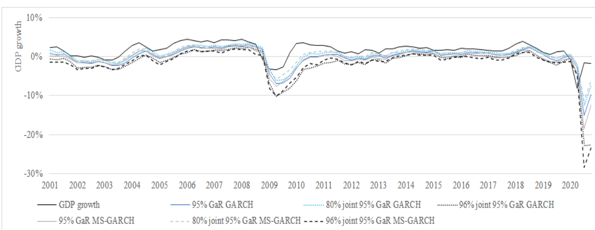
(b) AUT



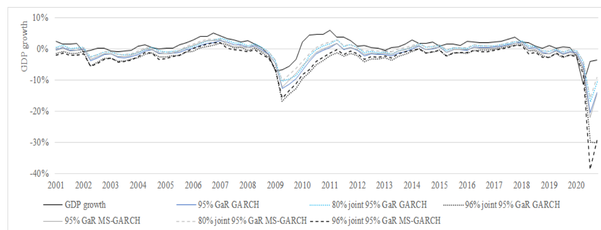
(c) BEL



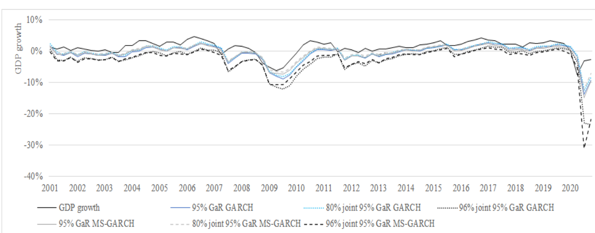
(d) CAN



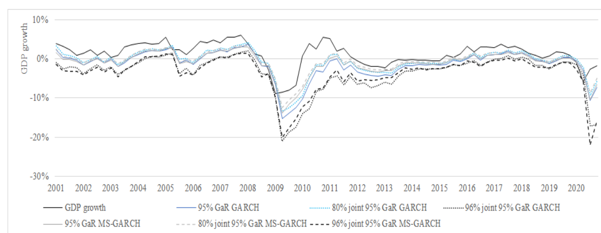
(e) CHE



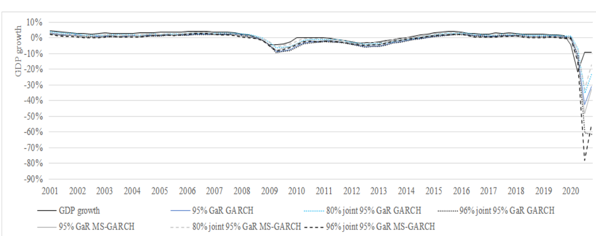
(f) DEU



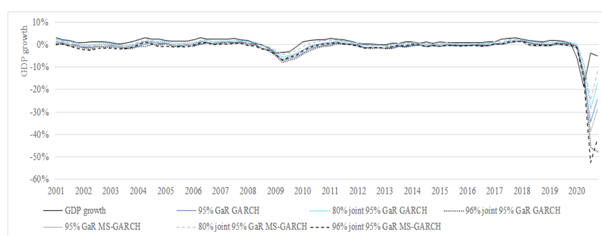
(g) DNK



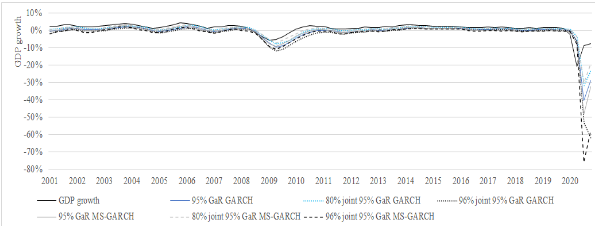
(h) FIN



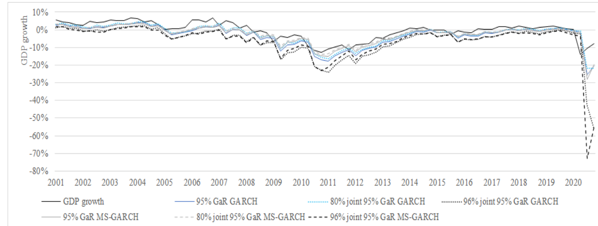
(i) ESP



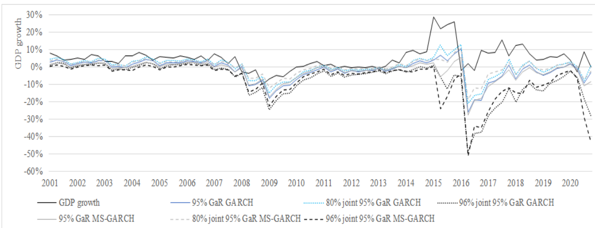
(j) FRA



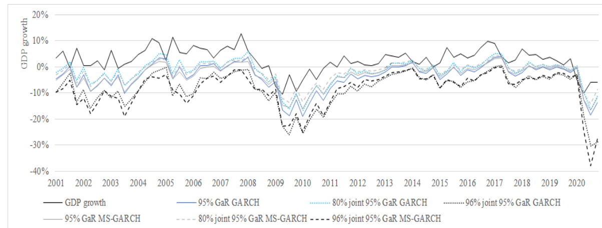
(k) GBR



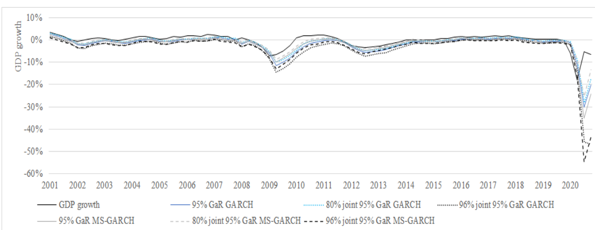
(l) GRC



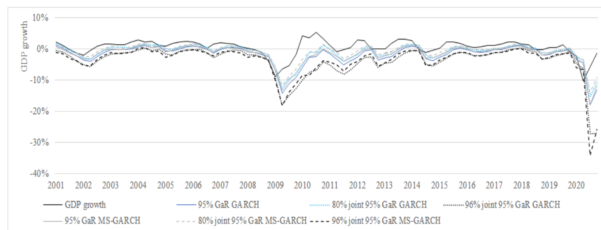
(m) IRL



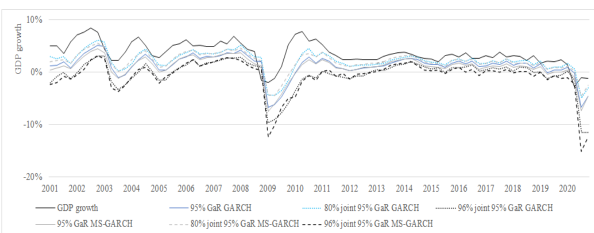
(n) ISL



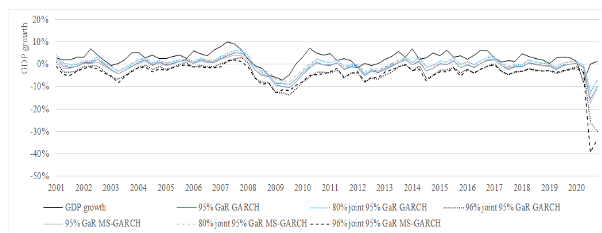
(o) ITA



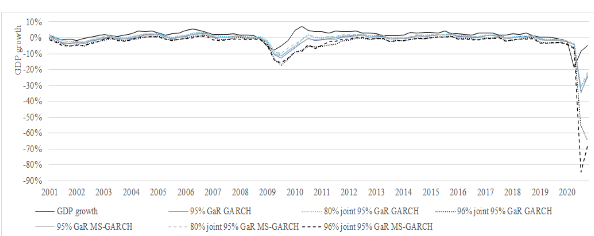
(p) JPN



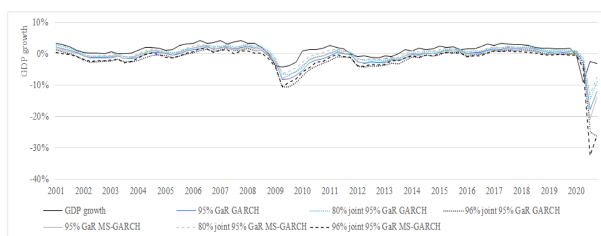
(q) KOR



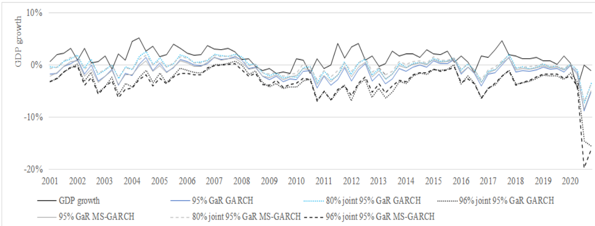
(r) LUX



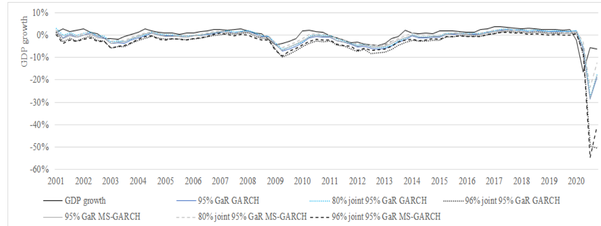
(s) MEX



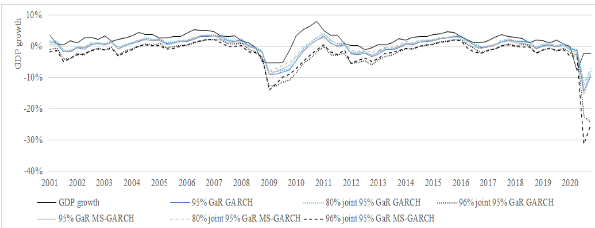
(t) NLD



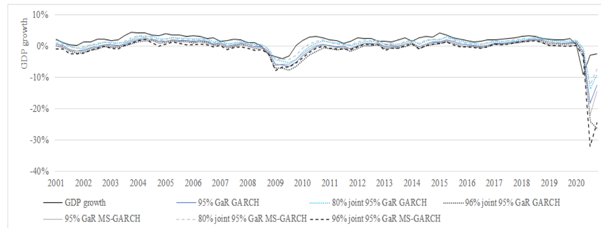
(u) NOR



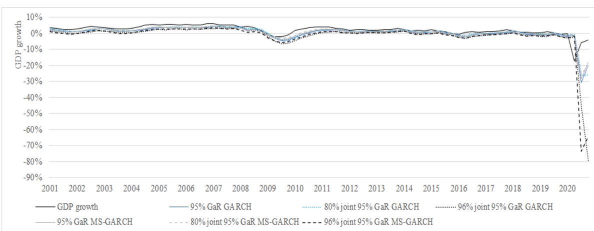
(v) PRT



(w) SWE



(x) USA



(y) ZAF

Figure 21. This figure plots the GDP growth rates alongside the one-quarter ahead 80% joint and 96% joint 95% GaR forecasts of the GARCH(1,1) and MS-GARCH(1,1) models. Subfigures (21a)-(21y) provide the graphs for all 25 OECD countries in the 100%-sample. In the figures, the solid black lines plot the realized GDP growth rates. Moreover, the solid blue and solid grey lines, respectively, reflect the one-quarter ahead marginal 95% GaR forecasts of the GARCH and the MS-GARCH model. The dotted black and dotted blue lines, respectively, reflect the 80% joint 95% GaR and 96% joint 95% GaR forecasts of the GARCH model, and the dashed black and dashed blue lines reflect the 80% joint 95% GaR and 96% joint 95% GaR forecasts of the MS-GARCH model, respectively. The out-of-sample period ranges from 2001Q1 to 2020Q4.