Erasmus University Rotterdam
Erasmus School of Economics

Master Thesis Econometrics & Management Science
(Quantitative Finance)

Forecasting Growth-at-Risk by dynamic models

Author:
Wouter de Voogd

Student number:
452840

Supervisor:
prof. dr. C. Zhou

Second assessor:
B. van Os MSc

Date: September 16, 2021

Abstract

We predict the Growth-at-Risk (GaR) for 23 OECD countries by multiple dynamic models that do not need explanatory variables. Taking the GARCH model as a starting point, we investigate the incorporation of several adaptations. We find that information pooling in the GARCH modelling yields more stable and reliable estimations for univariate GARCH models. Autoregressive models show high explanatory power for monthly GDP data, in contrast to the often used quarterly data, filtering out extreme GaR estimations from the constant-mean GARCH models. We also find that regime-switching models do not enhance our GaR estimations due to the short length of the available time series. We also consider the probability that at least one country violates the joint GaR. Pooled GARCH models with multiple autoregressive lags for the mean show for joint GaR the best results, also mitigating the number of countries that violate the joint GaR.

Contents

1 Introduction

Recently, policymakers have increased their focus on downside risk. The IMF [2017] introduced Growth-at-Risk (GaR), which has rapidly become a popular measure for macroeconomic downside risk. Similarly to Value-at-Risk, it is a quantile of economic growth, considered for the GDP. In other words, the future GDP growth rate is lower than the $(1-p)$-GaR with probability p. Policymakers can use GaR analysis in their macro-financial surveillance toolkit, to enhance the methods for mitigating the risk to the financial system as a whole. In much research, this measure is modelled with explanatory macroeconomic or financial variables, including the IMF in the work of Prasad et al. [2019] and the OECD [Caldera Sánchez and Röhn, 2016]. Adrian et al. [2019] also advocate the use of quantile regression. In contrast, the results presented in Plagborg-Møller et al. [2020] indicate that financial variables have very limited predictive power. The IMF uses in the research of Prasad et al. [2019] a National Financial Conditions Index (NFCI) to perform a quantile regression. However, this index is only quarterly computed and has not been updated since 2016.

We investigate in this thesis dynamic models that do not need explanatory variables to model Growth-at-Risk. We model and estimate the distribution of the GDP growth and then estimate the GaR therefrom. In particular, we compare different methods to derive certain characteristics of GDP growth, to get more insight into GaR modelling in general. In literature, the quarterly GDP growth rates are the most common frequency used in literature so far, probably because many macroeconomic variables are quarterly available. In this sense, our research extends this existing research.

Univariate GARCH models are common-used in the context of risk management and specifically Value-at-Risk measures. However, they are relatively little used in the GaR literature. GARCH models capture volatility clustering, where periods of high and low volatility tend to alternate. This common feature of financial data also appears in monthly GaR data. Therefore, they are a good choice to model the GDP of a single country. Our starting point for estimating GaR is the GARCH(1,1) model of Bollerslev [1986]. In addition, we consider the GJR-GARCH model of Glosten et al. [1993], as this model takes asymmetry into account in conditional volatility dynamics, which can particularly be useful to distinguish downside and upside risk.

When considering each country individually, the univariate GARCH models generally need larger samples to obtain stable parameter estimates, when Quasi Maximum Likelihood (QML) estimation is used [Hwang and Valls Pereira, 2006]. Because GDP data is available only monthly, its sample size is relatively small. Therefore, univariate GARCH models may fail to adequately model the conditional heteroskedasticity in data. To overcome this problem, we assume that all univariate GARCH models for GDP growth time series in different countries share the same GARCH coefficients. Then we can use cross-sectional pooling of information to estimate the GARCH models together. Although this assumption is relatively strong, it may contribute to the stability of our models. We use the Composite Likelihood (CL) method of Pakel et al. [2011] to estimate these models. Pakel et al. [2011] show their model is capable of estimating conditional heteroskedasticity correctly, using previously infeasible sample sizes. We investigate whether the assumption of this approach is adequate and whether this approach can enhance our GaR estimations.

As economies become more dependent on each other in the globalized world of nowadays, we model such cross-sectional dependence. For this purpose, we construct multivariate GARCH models for the GDP growth rate, assuming that the GDP growth of countries have explanatory power on each other, using the Scalar BEKK model of Engle et al. [2019], which captures the dynamics in the covariances between the considered GDPs. For the estimation, we use Composite Likelihood to obtain consistent parameter estimators, according to Pakel et al. [2020].

Chauvet and Hamilton [2006] and Smith and Summers [2009] show that GDP growth has a strong regime-dependence structure. We investigate the power of incorporating regime switching in the modelling of the GDP growth. With assuming that the GDP growth follows different distributions over time, we apply the Markov Switching GARCH model of Haas et al. [2004] for the GDP to investigate different states of the economies for all countries individually.

Besides marginal Growth-at-Risk, we also predict the joint Growth-at-Risk as in Brownlees and Souza [2021]. The marginal GaR is defined such that for each country *individually* the probability of a violation (a growth rate below the marginal GaR) is p. In contrast, the joint GaR is defined such that the probability is p that the growth rate of *at least one* considered country falls below its joint GaR prediction. We predict the joint GaR by the Bootstrap Joint Prediction Region (BJPR) method of Wolf and Wunderli [2015].

We evaluate our GaR predictions for the coverage levels of 0.75, 0.95 and 0.99 and for time horizons up to three months. Besides the average coverage level, we also perform the following evaluation tests. First, we adopt the commonly used Dynamic Quantile Test of Engle and Manganelli [2004] for our marginal GaR estimations and adapt the test also to apply it for our joint GaR predictions. Second, we compare our estimation methods for marginal GaR by a comparative backtest in Nolde and Ziegel [2017] on the Tick Loss of Giacomini and Komunjer [2005]. This loss function calculates the weighted average difference between the real growth rate and the GaR, where higher weights are assigned if the GaR is violated. Third, we use the Dynamic Binary Test of Dumitrescu et al. [2012], which is especially suitable for small samples. We apply the DB test not only for marginal GaR, but we also evaluate the joint GaR, analogously to the Dynamic Quantile tests.

We show in this thesis that GARCH models are very useful to outperform the historical GaR. However, they suffer from instability of their GARCH parameters, which especially worsens long-term forecasts. We show that the Composite Likelihood method, whereby the GARCH parameters are assumed common, successfully solves this problem. A simple approach yields here the best results, as the use of monthly updated clusters does not provide better results. In addition, we contribute to the literature by investigating monthly GDP growth data. We show this data is advantageous in terms of the estimation of the conditional mean, in comparison to quarterly data. Therefore, more research on this topic would be useful. Furthermore, we show that a regime-switching GARCH model for the purpose of Growth-at-Risk mostly does not perform better in comparison to the univariate GARCH models and an unadapted multivariate normal approach yields overestimation of the coverage level. Finally, we contribute to the joint Growth-at-Risk research. First, we propose the Hits Given Failure (HGF) measure to evaluate to which extent countries are violating the joint GaR together. Second, we show that Composite Likelihood also yields the best results for the joint GaR, which is also expressed by the fact that mostly one or few countries violate the GaR at one time.

The remainder of this thesis is organized as follows. In Section 2, we review the main studies about this topic, where we mention the main findings of previous researches. Section 3 describes the GDP growth rate data that is used in this thesis. Section 4 contains the methodology we use to estimate the Growth-at-Risk of multiple countries. In Section 5, we present the results of our different estimation methods. Section 6 concludes.

2 Literature review

Since the principle of Growth-at-Risk was investigated by both the IMF [2017] and OECD [Caldera Sánchez and Röhn, 2016], the measure has gained popularity and is incorporated into the macro-financial surveillance toolkit of the IMF. Since then, the literature about Growth-at-Risk is rapidly growing. Adrian et al. [2019] investigate the use of quantile regressions for the GDP growth rate of the USA. Brownlees and Souza [2021] investigate the use of univariate GARCH models and backtest the results, concluding that GARCH models and quantile regression models have similar forecasting performance. In addition, they introduce joint Growth-at-Risk, which measures downside risk for a set of countries.

Most existing research about GaR, including Brownlees and Souza [2021], Adrian et al. [2019] and Plagborg-Møller et al. [2020], use quarterly data, probably because many macroeconomic variables are available on a quarterly basis. In contrast, we investigate monthly GDP growth data in our thesis, which extends this existing research. Plagborg-Møller et al. [2020] consider US data and a panel of twelve OECD countries. They conclude that financial variables have very limited predictive power for the Growth-at-Risk, beyond the information contained in real indicators. Therefore, we focus on estimating the Growth-at-Risk using volatility models rather than using explanatory variables. This allows us to use monthly data as the purpose of our research conveniently.

GARCH models occupy a prominent place in the literature on quantitative finance since their introduction by Bollerslev [1986]. An important extension of the standard GARCH model is made by Glosten et al. [1993], who introduced asymmetry in the error terms. However, Hwang and Valls Pereira [2006] show that the Quasi Maximum Likelihood estimates of univariate GARCH(1,1) models are significantly negatively biased in small samples. In addition, the parameter estimates are often instable, when Quasi Maximum Likelihood is used. The Composite Likelihood method, first investigated by Engle et al. [2008], is a pooling method designed to solve these problems. In this thesis, we also use the CL method for the estimation of univariate GARCH models, because GDP growth rates are available at a low frequency. Pakel et al. [2011] investigate CL further and use cross-sectional information, leading to the pooled GARCH estimation method that we use in this thesis. They show that CL is able to enhance univariate GARCH model estimations for data sets of hundreds of observations. They state that although the assumption that all series share common parameters is almost certainly violated, CL still produces better results for short time series. CL was also utilized by Brownlees and Souza [2021] to jointly estimate the quarterly Growth-at-Risk of multiple countries, only considering univariate GARCH models. In our research, we focus on monthly GDP growth data and extend the research by considering clusters for pooled GARCH models.

The literature also offers many multivariate GARCH models. Amongst them is the BEKK model of Engle and Kroner [1995], which was adapted to the Scalar BEKK model of Engle et al. [2019]. Pakel et al. [2020] applied Composite Likelihood to multivariate GARCH models, showing this yields consistent parameter estimators. In addition, this decreases the risk of parameter instability and avoids parameter bias for small samples. We use multivariate GARCH to forecast Growth-at-Risk, which is a novel approach to our knowledge.

Chauvet and Hamilton [2006] show evidence that GDP growth has a regime-dependence structure, which was supported more recently by the findings of Chang et al. [2017]. Smith and Summers [2009] find a general change in the volatility of GDP growth, which was not permanent. Whereas they rely on an autoregressive model as a starting point for their regime-switching model, we focus on GARCH models as the basis of our regime-switching framework. The Markov Switching framework was first applied on econometric models by Hamilton [1989]. Gray [1996] proposed a MS-GARCH model under the hypothesis that the conditional variance for all regimes depends on the expectation of previous conditional variances, which was further modified by Klaassen [2002]. As these models suffer from analytical intractability, Haas et al. [2004] uses an approach where each conditional variance depends only on its own lag, which is attractive for our data set of limited length. This model is advantageous because its structure allows for deriving expressions of the covariance structure of the process. From these expressions, we also can derive constraints that ensure covariance stationarity. Therefore, we use the approach of Haas et al. [2004] in this thesis for our regime-switching framework.

One of the most popular Value-at-Risk backtests in the literature is the Dynamic Quantile (DQ) test, which was introduced by Engle and Manganelli [2004], which is constructed by utilizing the criterion that each period the probability of exceeding the VaR must be independent of all the past information. However, Hurlin and Tokpavi [2008] state that the power of the backtesting test is generally low for short time series, i.e. it does not reject the validity of a model often enough. Therefore, we also use the Dynamic Binary backtest of Dumitrescu et al. [2012] and evaluate whether our models show different performance for both backtests.

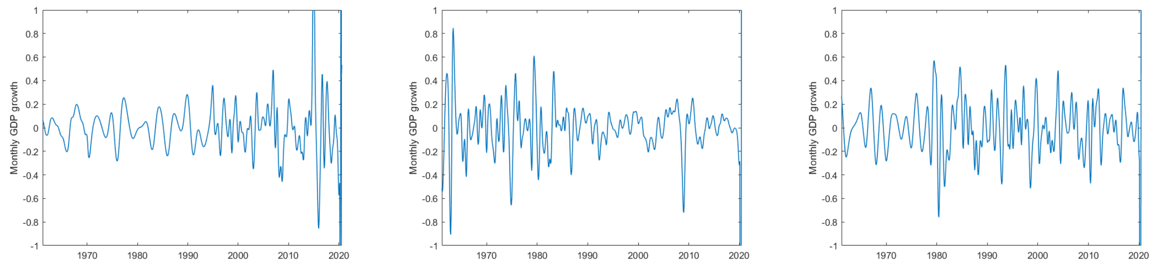
3 Data

3.1 Monthly GDP growth

We use monthly GDP growth rates for 23 OECD countries[1], with a time window from March 1961 up to August 2020. Here, GDP growth rates are defined as the monthly percentage change

[1] We consider the following countries in our analysis: Australia (AUS), Austria (AUT), Belgium (BEL), Canada (CAN), Switzerland (CHE), Germany (DEU), Denmark (DNK), Spain (ESP), Finland (FIN), France (FRA), United Kingdom (GBR), Greece (GRC), Ireland (IRL), Italy (ITA), Japan (JPN), South Korea (KOR), The Netherlands (NLD), Norway (NOR), Portugal (PRT), Sweden (SWE), Turkey (TUR), United States (USA) and South Africa (ZAF).

Figure 1: GDP growth of three countries over time.



(a) Ireland (b) The Netherlands (c) Norway

in seasonally adjusted GDP. This data is obtained from the FRED[2].

Table 18 in the Appendix, Section B, shows the cross-correlations of the whole data set. All countries have a positive correlation, which supports the fact that economies depend on each other. However, the cross-correlations vary from 0.04 to 0.97. In addition, we observe that the cross-correlations between France, Germany, Portugal and Italy are clearly the highest. This finding suggests that the data exhibits a certain form of time series clustering.

Figure 1 shows the GDP growth over time of three countries in the data set. We observe that the GDP data exhibits volatility clustering, which also varies over time. This observation suggests that a model that captures both time-varying volatility and volatility clustering would suit our data well. In addition, we observe for all countries both periods of low and high volatility. This would suggest a model that allows for regime-switching in predicting the volatility of the GDP growth. However, those regimes turn out to be not omnipresent, except for the coronavirus recession, which we observe in the GDP growth rates of all countries.

In our data set, this recession is represented by the last six months of our data set, namely March 2020 to August 2020. Figure 6 in the Appendix shows the mean and variance of these months, compared with the two years before. We observe that the mean growth rate is generally close to zero, but moves in the coronavirus recession between -5% to +2%. The variance across countries is even multiplied by more than 10 for all months. Therefore, it is likely that our methods behave differently in this estimation period.

To investigate the dynamics of the GDP growth rates, we fit a univariate GARCH(1,1) model for all GDP growth rates. Table 1 shows the results. We observe that $\alpha_i + \beta_i$ is close to one for many countries[3]. Since we estimate the GARCH models with the restriction $\alpha_i + \beta_i < 1$,

[2]FRED is the dataset of the Federal Reserve Bank of St. Louis, available at `https://fred.stlouisfed.org/`.

[3]For some countries, the parameters of α_i and β_i seem to add to one in Table 1, which is due to rounding to three decimal places.

the parameters of these time series would likely exceed one without this restriction. These parameter estimations result in high variance estimations. Therefore, the univariate GARCH models have the risk to suffer from instability and produce biased GaR estimations. The fact that the squared returns do not match with the autocorrelation functions of the squared returns also suggests the presence of estimation bias. Figure 7 in the Appendix shows that the autocorrelations of almost all countries drop relatively quick to zero. The finding that univariate GARCH models are biased is in line with the conclusions of Pakel et al. [2011]. This motivates the choice of a model where cross-sectional information is used to decrease the risk of parameter instability in GARCH models.

Table 1: Estimated GARCH parameters for all countries, based on the full data set, where the GARCH model is denoted as $\sigma^2_{i,t+1|t} = \omega_i + \alpha_i \varepsilon^2_{i,t} + \beta_i \sigma^2_{i,t}$ and $\mu_{i,t+1|t} = \mu_i$.

Par.	AUS	AUT	BEL	CAN	CHE	DEU	DNK
ω_i	0.001	0.001	0.000	0.002	0.001	0.065	0.001
α_i	0.847	0.952	0.846	0.964	0.907	0.325	0.905
β_i	0.153	0.048	0.154	0.036	0.093	0.000	0.095
ESP	FIN	FRA	GBR	GRC	IRL	ITA	JPN
0.093	0.001	0.180	0.096	0.003	0.001	0.084	0.001
0.051	0.926	0.064	0.224	1.000	0.871	0.051	0.480
0.698	0.074	0.731	0.063	0.000	0.129	0.693	0.520
KOR	NLD	NOR	PRT	SWE	TUR	USA	ZAF
0.000	0.001	0.002	0.048	0.002	0.034	0.047	0.001
0.790	0.838	0.884	0.046	0.917	0.474	0.295	0.976
0.210	0.162	0.116	0.649	0.084	0.000	0.000	0.024

Although the assumption that all series share the same parameters is violated almost certainly, Pakel et al. [2011] find that CL also can enhance the results for short time series in this case. Therefore, we first assume that all time series share the same GARCH coefficients. However, Table 1 however both shows countries for which the volatility is mainly determined by the last error term and countries for which the volatility is mainly estimated by the past volatility. Therefore, we also allow for multiple clusters of countries in which we perform information pooling to allow for clustering countries with similar dynamics.

To get more insight into the dynamics of the GARCH models for a given data set, we estimate univariate GARCH models for the individual countries, based on the data from the start of the data set, until a given point in time. Figure 5 in the Appendix, Section B shows some coefficients over time, for illustration. We observe for these countries that the GARCH model for smaller data sets have strongly varying parameters, mostly for a length smaller than 15 years (which corresponds to approximately 180 observations).

3.2 Quarterly GDP growth

In addition, we also consider quarterly GDP data, which is mostly evaluated in the literature. We use the data which was considered earlier by Brownlees and Souza [2021], namely 24 OECD countries from 1973 to 2016. Most countries from these data set are also in the monthly data that we consider[4]. We compare the basic features of the two data sets together in Table 2. The quarterly GDP growth rates have a clear higher mean than the monthly rates, as the data for quarterly growth rates is not inflation-corrected and has a lower frequency. In addition, the variance is about twelve times higher. When we estimate an AR(1) and AR(3) model for both data sets, 44 percent of the variance for the monthly data is already explained by an AR(1) model, where an AR(1) model for the quarterly data only yields a R^2 of 0.061. Therefore, we investigate how these data features are reflected in our GaR estimations, as this suggests that a model with an autoregressive mean performs clearly better for the monthly GDP data.

Table 2: Comparison of our main GDP data (of monthly frequency) and the data set where quarterly GDP is considered. We fit for both models autoregressive models on the whole data set and report the parameter averages over all countries.

			AR(1) model			AR(3) model				
Data	μ	σ^2	ϕ_0	ϕ_1	R^2	ϕ_0	ϕ_1	ϕ_2	ϕ_3	R^2
Monthly	-0.006	0.155	-0.001	0.650	0.441	-0.003	-0.124	-0.179	0.810	0.510
Quarterly	0.796	1.804	0.692	0.129	0.061	0.541	0.102	0.134	0.081	0.116

4 Methodology

4.1 General framework

4.1.1 GDP growth model

Consider the GDP growth rate $y_{i,t}$ of a single country i at time t, for $i = 1, ..., N$ countries. We define the general framework, that allows for volatility clustering, as follows:

$$y_{i,t+1} = \mu_{i,t+1|t} + z_{i,t+1}\sqrt{\sigma^2_{i,t+1|t}}, \text{ where } z_{i,t+1} \sim f_{z_i}(0,1) \text{ i.i.d.}, \tag{1}$$

[4]The difference between the two data sets is that Brownlees and Souza [2021] use the countries of Mexico, Iceland and Luxembourg, where we use Turkey and South Africa.

where $\mu_{i,t+1|t}$ denotes the conditional mean, $\sigma^2_{i,t+1|t}$ denotes the conditional variance and $f_{z_i}(0,1)$ denotes a distribution with zero mean and unit variance. The $(1-p)$-GaR is defined as the lower one-sided prediction interval that contains future realizations of GDP growth of a given country with a coverage level $1-p$. The corresponding conditional p-quantile is derived from the general framework by

$$Q_p(y_{i,t+1|t}) = \mu_{i,t+1|t} + F^{-1}_{z_i}(p) \cdot \sqrt{\sigma^2_{i,t+1|t}} \tag{2}$$

where $F^{-1}_{z_i}(p)$ denotes the inverse cumulative distribution function of f_{z_i}.

4.1.2 Estimation of conditional mean

From Equation (2), the conditional mean $\mu_{i,t+1|t}$ is the starting point of the volatility models. We use different assumptions of the conditional mean to model the GDP growth rates. As a benchmark, we consider the case where the GDP of each country has a constant mean, i.e.

$$\mu_{i,t+1|t} = \mu_i, \quad \forall i = 1, ..., N. \tag{3}$$

An alternative for the constant mean is a simple autoregressive model, allowing the mean to vary over time. As such, we model the conditional mean at each time t for based on an autoregressive model. That is, the conditional mean is estimated as follows:

$$\mu_{i,t+1|t} = \phi_{i,0} + \sum_{l=1}^{L} \phi_{i,l} y_{i,t-l+1}, \tag{4}$$

where L denotes the number of used lags. We consider an AR(1) model, which was also used as the conditional mean for GDP growth rates by Brownlees and Souza [2021], and an AR(3) model. In addition, we also investigate a multivariate estimation of the conditional mean. As the GDPs of countries have a high dependency on each other, it is useful to investigate whether this dependence can be captured using cross-sectional information in the construction of the conditional mean. For this purpose, we use a VAR(1) model to estimate the conditional means simultaneously:

$$\boldsymbol{\mu}_{t+1|t} = \boldsymbol{c} + \boldsymbol{A}\boldsymbol{y}_t, \tag{5}$$

where $\boldsymbol{\mu}_{t+1|t} = (\mu_{1,t+1|t}, ..., \mu_{N,t+1|t})'$ and the matrix $\boldsymbol{A}$ is an $N \times N$ matrix of parameters.

4.2 Volatility models

4.2.1 Univariate GARCH models

Next, we explain the univariate volatility models we use, taking the conditional mean $\mu_{i,t+1|t}$ as given. The **GARCH(1,1)** model of Bollerslev [1986] assumes the conditional variance as

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

where $\varepsilon_{i,t} = y_{i,t} - \mu_{i,t|t-1}$ and the following constraints are imposed: $\sigma^2_i(1-\alpha_i-\beta_i) > 0$, $\alpha_i > 0$, $\beta_i > 0$ and $\alpha_i + \beta_i < 1$, $\forall i = 1, ..., N$.

The **GJR-GARCH(1,1)** model of Glosten et al. [1993] takes asymmetry into account in conditional volatility dynamics, which can in particular be useful to distinguish downside and upside risk. This model estimates the conditional volatility by

$$\sigma^2_{i,t+1|t} = \sigma^2_i(1-\alpha_i-\beta_i-\frac{1}{2}\gamma_i) + (\alpha_i + \gamma_i\mathbb{1}[\varepsilon_{i,t} < 0])\varepsilon^2_{i,t} + \beta_i\sigma^2_{i,t+1|t}, \tag{7}$$

where $\varepsilon_{i,t} = y_{i,t} - \mu_{i,t|t-1}$ and $\mathbb{1}[\cdot]$ denotes the indicator function. The following constraints are satisfied: $\sigma^2_i(1-\alpha_i-\beta_i-\frac{1}{2}\gamma_i) > 0$, $\alpha_i > 0$, $\beta_i > 0$ and $\gamma_i > 0$, $\forall i = 1, ..., N$. We estimate our univariate GARCH models by Quasi Maximum Likelihood (QML) using a Gaussian distribution.

4.2.2 Pooled GARCH

Following Pakel et al. [2011], we use cross-sectional pooling of information, assuming that all univariate GARCH models share the same parameters α and β, and γ for the GJR-GARCH(1,1) model. For generality, we denote these parameters as $\boldsymbol{\theta}$. However, we still assume that the unconditional parameters σ^2_i are asset-dependent. This means that the conditional variances for the GARCH(1,1) model in Equation (6), that satisfies the usual constraints, are given as

$$\sigma^2_{i,t+1|t} = \sigma^2_i(1-\alpha-\beta) + \alpha y^2_{i,t} + \beta\sigma^2_{i,t|t-1}. \tag{8}$$

For the GJR-GARCH(1,1) model, we perform information pooling analogously. Our estimation of Pooled GARCH is performed in two steps. First, the unconditional variances σ^2_i are initialized by the sample variance. Second, the parameters $\boldsymbol{\theta}$ are estimated by optimizing the composite likelihood function. For $f(y_{i,t}|\mathcal{F}_{t-1})$ being the conditional density of $y_{i,t}$, given all information up to $t-1$, $\mathcal{F}_{t-1}$, the composite likelihood function is given by

$$CL(\phi_{(N)}; y) = \frac{1}{T}\Big[\frac{1}{N}\sum_{i=1}^{N}\log f(y_{i,t}|\mathcal{F}_{t-1};\boldsymbol{\psi}_i)\Big], \tag{9}$$

where $\boldsymbol{\psi}_i = (\boldsymbol{\theta}', \sigma_i^2)$, T denotes the length of the estimation window and N denotes the total number of time series. The parameters $\boldsymbol{\theta}$ are estimated by the maximization of the composite likelihood function. Assuming that $y_{i,t}$ conditionally follows a normal distribution, we maximize the following likelihood function:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{T-1} \frac{1}{N} \sum_{i=1}^{N} \Big(-\frac{1}{2} \log \hat{\sigma}^2_{i,t+1|t}(\boldsymbol{\theta}) - \frac{1}{2} \frac{y^2_{i,t+1}}{\hat{\sigma}^2_{i,t+1|t}(\boldsymbol{\theta})} \Big), \tag{10}$$

where all $\hat{\sigma}^2_{i,t+1|t}$ are estimated according to Equation (8).

In addition, we also make use of information pooling for clusters of countries. Some countries may differ in terms of volatility dynamics, making the estimations of the GARCH models inadequate. Therefore, we investigate whether Clustered GARCH mitigates the estimation error while reducing the instability of the GARCH models. We cluster our N time series of GDP growth rates by k-means clustering, dividing them into k clusters, where we iteratively assign each GDP time series to the cluster with the nearest mean. This is calculated by a minimization problem, where squared Euclidean distances are used to minimize the within-cluster variances. For our purpose, we cluster the $N = 23$ time series frequently based on their past GDP observations. In this way, we aim to group the countries with similar behaviour over time and use information pooling within these groups. We update the clusters every month.

4.2.3 Markov Switching GARCH

We apply a Markov Switching GARCH model for the GDP growth rates to allow for two regimes in our analysis. Generally, MS-GARCH models the variance for country i as follows:

$$\sigma^2_{i,t}(r_{i,t}) = \omega_i(r_{i,t}) + \alpha_i(r_{i,t})\varepsilon^2_{i,t-1} + \beta_i(r_{i,t})\sigma^2_{i,t-1}. \tag{11}$$

Here, $r_{i,t}$ is a country-specific variable which indicates the state r of the world at time t which follows a Markov chain with country-specific finite state space $R_{i,t} = 1, 2$ and an country-specific transition matrix $\boldsymbol{P}_i$, equal to

$$\boldsymbol{P}_i = \begin{bmatrix} P(r_{i,t}=1|r_{i,t-1}=1) & P(r_{i,t}=1|r_{i,t-1}=2) \\ P(r_{i,t}=2|r_{i,t-1}=1) & P(r_{i,t}=2|r_{i,t-1}=2) \end{bmatrix} = \begin{bmatrix} p_{i,11} & 1-p_{i,22} \\ 1-p_{i,11} & p_{i,22} \end{bmatrix}. \tag{12}$$

In our analysis, we use the analytically tractable approach of Haas et al. [2004], who propose the following approach in which each specific conditional variance depends on its own lag:

$$\sigma^2_{i,t}(r_{i,t}) = \alpha_{i,0}(r_{i,t}) + \alpha_i(r_{i,t})\varepsilon^2_{i,t-1} + \gamma_i(r_{i,t})\sigma^2_{i,t-1}(r_{i,t}), \tag{13}$$

satisfying positive variance and the stationarity constraints of Haas and Paolella [2012]. In the model of Haas et al. [2004], the mean $\mu_{i,t}$ is assumed to be equal across regimes. In our implementation, we make use of the MSGARCH toolbox of Chuffart [2017] and estimate the model with Maximum Likelihood, using the filter of Hamilton [1989].

4.2.4 Multivariate GARCH

Next, we investigate the dependence between the economies, assuming that the GDP growth of different countries have explanatory power on each other. To incorporate this dependence structure in the GDP modelling, we use multivariate GARCH, using the assumption that the returns $\boldsymbol{y}_t$ follow a normal distribution with time-varying mean and variance:

$$\boldsymbol{y}_t|\mathcal{F}_{t-1} \sim N(\boldsymbol{\mu}_t, \boldsymbol{H}_t). \tag{14}$$

For the estimation of our multivariate GARCH model, we use the Scalar BEKK model of Engle et al. [2019], which has the advantage that it is a computationally simple model. Scalar BEKK estimates the conditional covariance matrix as

$$\boldsymbol{H}_t = (1 - \alpha - \beta)\boldsymbol{\Sigma} + \alpha \boldsymbol{y}_{t-1}\boldsymbol{y}_{t-1} + \beta \boldsymbol{H}_{t-1}, \tag{15}$$

where α and β are dynamic parameters and $\boldsymbol{\Sigma}$ denotes the unconditional covariance matrix $E[\boldsymbol{y}_t\boldsymbol{y}_t']$. First, $\boldsymbol{\Sigma}$ is estimated by a simple moment estimator $\hat{\boldsymbol{\Sigma}} = \frac{1}{T}\sum_{i=1}^{T} \boldsymbol{y}_t\boldsymbol{y}_t'$. Next, α and β are estimated, based on a likelihood function. As larger dimensions can accuse computational problems in conventional Maximum Likelihood, Pakel et al. [2020] developed a procedure to estimate Scalar BEKK by Composite Likelihood, showing that maximizing the composite likelihood yields consistent estimators of α and β for Scalar BEKK. For this purpose, they approximate the objective function by the average of bivariate densities, that are constructed from asset pairs. For details regarding about the estimation procedure, we refer to this paper. For the estimation of Scalar BEKK, we rely on the MFE Toolbox of Kevin Sheppard[5].

4.3 Growth-at-Risk estimation

4.3.1 Definition of marginal and joint Growth-at-Risk

Our analysis of Growth-at-Risk focuses on h-step ahead growth rates. For the estimation, we use a bootstrapping algorithm for the GDP shocks, from which the quantiles are taken. We

[5]The code and documentation of the MFE Toolbox are available at `https://www.kevinsheppard.com/code/matlab/mfe-toolbox/`.

estimate the Growth-at-Risk for each period in time again. The first 174 observations are chosen as starting in-sample period, because we showed in Section 3 that GARCH models have an instable behaviour for a very short estimation window, but become in our dataset more stable for approximately 180 observations. First, we consider the marginal GaR, which is defined as follows. For a given country i, the h-step-ahead marginal Growth-at-Risk for country i at time t utilizes

$$P\big(y_{i,t+h|t} \leq \mathrm{GaR}^{M}_{i,t+h|t}(p)\big) = p. \tag{16}$$

In addition, we also calculate the joint Growth-at-Risk as in Brownlees and Souza [2021]. Here, the quantiles of the countries' growth rates are estimated such that the probability is p that the growth rate of at least one considered country falls below its own joint GaR prediction for time t. Mathematically, this is denoted as

$$P\Big(\sum_{i=1}^{N} \mathbb{1}\left[y_{i,t+h|t} \leq \mathrm{GaR}^{J}_{i,t+h|t}(p)\right] \geq 1\Big) = p. \tag{17}$$

4.3.2 General estimation of marginal Growth-at-Risk

For marginal GaR, we first describe the estimation of GaR for a univariate GARCH model, which is from Equation (16) mathematically constructed as

$$\mathrm{GaR}^{M}_{i,t+h|t}(p) = Q_p(y_{i,t+h} \mid \mathcal{F}_t), \tag{18}$$

for $Q_p(\cdot)$ being the quantile at a coverage level $(1-p)$. Generally, the marginal GaR is estimated as follows. First, we run this univariate model based on the data available at time t, based on one of the definitions of the conditional mean $\mu_{i,t+1|t}$ as described in Subsubsection 4.1.2. From these estimations, we obtain the conditional means $\mu_{i,t}$, conditional variances $\sigma^2_{i,t}$ for $i = 1, ..., N$ and $t = 1, ..., T$ and the parameters of the conditional mean $\boldsymbol{\theta}_\mu$, depending on the definition of the conditional mean, and the parameters $\boldsymbol{\theta}_\sigma$ of the conditional variance, depending on the definition of $\sigma_{i,t+1|t}$. For notational convenience, we define T here as the current period in time or the length of the in-sample period.

We obtain the standardized residuals by $\hat{z}_{i,t} = (y_{i,t} - \mu_{i,t})/\sigma_{i,t}$. Next, we draw S times an index from the values $1, ..., T-h$, that are used for the purpose of bootstrapping, such that we draw for each time series i a number of $S = 5000$ bootstrapped innovations $z_{i,1}, ..., z_{i,S}$. Next, we calculate the bootstrapped path as $\tilde{y}^{(s)}_{i,T+1|T} = \mu_{i,T+1|T} + \sigma_{i,T+1|T} z^{(s)}_{i,t}$, for $s = 1, ..., S$. For $j > 1$, we repeat this procedure and estimate again the conditional mean and variance iteratively per day-ahead path as $\tilde{y}^{(s)}_{i,T+j|T} = \mu_{i,T+j|T} + \sigma^{(s)}_{i,T+j|T} z^{(s)}_{i,t}$ at intermediate horizon j, until we reached horizon of interest h. Finally, we take the p-quantile from the simulated paths as the marginal GaR. The whole procedure is summarized in Algorithm 2 in the Appendix.

4.3.3 Markov Switching GARCH estimation

To estimate the Growth-at-Risk for Markov Switching GARCH models, the general procedure is adapted to a model with two states. The idea of the estimation is that we simulate S paths, including simulating the state in each step and finally take the quantile of the simulated growth rates for a given time horizon.

Similarly to the standard GARCH models, we first run the MS-GARCH model for country i based on the in-sample data. In the use of the MS-GARCH model, we focus on the estimation of the volatility. Therefore, we assume for both regimes equal mean, which we set equal to the historical mean μ_i. From the estimated MS-GARCH model, we obtain the state-specific parameters $\omega_{i,r}$, $\alpha_{i,r}$ and $\beta_{i,r}$, for the states $r = 1, 2$. We also derive the transition matrix $\boldsymbol{P}_i$ and the conditional state-specific variances $\sigma^2_{i,t,r}$, from which we calculate the standardized residuals $z_{i,t,r}$. In addition, we also retrieve for each point in time $t = 2, ..., T$ the predicted state probabilities $\left[q_{1,t}, q_{2,t}\right]' = \left[P(r_{i,t} = 1), P(r_{i,t} = 2)\right]'$, where $r_{i,t}$ denotes the state of country i at time t.

Next, we bootstrap for horizon $j = 1$ the standardized residuals $z^{(s)}_{i,t,r}$ and calculate the state-specific 1-step-ahead predicted variances $\sigma^2_{i,T+1|T,r} = \omega_{i,r} + \alpha_{i,r}\left(y_{i,T} - \mu_i\right)^2 + \beta_{i,r}\sigma^2_{i,T,r}$, for both states $r = 1, 2$. We update the state probabilities of time $T + 1$ as

$$\left[q_{1,T+1}, q_{2,T+1}\right]' = \boldsymbol{P}_i\left[q_{1,T}, q_{2,T}\right]'. \tag{19}$$

From these probabilities, we draw for S simulation paths the state at time $T + 1$. According to each drawn state $r(s)$, we assign to each simulation the state-specific variance $\sigma^{2(s)}_{i,T+1|T} = \sigma^2_{i,T+1|T,r(s)}$ for state $r(s) = 1, 2$. From this variance, we also simulate S values for the next GDP growth observation, based on the bootstrapped values of the standardized residuals as

$$\tilde{y}^{(s)}_{i,T+1|T} = \mu_i + \sigma^{(s)}_{i,T+1|T} z^{(s)}_{i,t,r}, \tag{20}$$

where μ_i denotes the unconditional mean, $\sigma^{(s)}_{i,T+1|T}$ is the forecasted variance for the drawn state of simulation s and $z^{(s)}_{i,t,r}$ is the bootstrapped standardized residual, from the drawn state r.

For $j > 1$, we repeat this procedure. Again, we update the state probabilities for time $T+j$ and draw for each simulation the state and standardized residuals. We calculate the conditional means and variances to obtain the simulated paths for time $T + j$ until we reach time $T + h$.

Next, we calculate the marginal GaR as the quantile of the S simulated paths of the returns at time $T+h$. The procedure is summarized in Algorithm 1.

Run the MS-GARCH model with data available at time T. Obtain:

- The means μ_i and variances $\sigma^2_{i,t}$, to calculate the standardized residuals $z_{i,t,r}$;

- State-specific parameters $\omega_{i,r}$, $\alpha_{i,r}$ and $\beta_{i,r}$, and the state probabilities $q_{r,t}$ for states $r = 1, 2$, time $t = 2, ..., T$;

- Transition matrix $\boldsymbol{P}_i$;

for $j = 1, ..., h$ day-ahead paths **do**

Draw S times an index from $[1, ..., T-h]$ and bootstrap the values $z^{(1)}_{i,t,r}, ..., z^{(S)}_{i,t,r}$ for $r = 1, 2$;

if $j = 1$ **then**

Calculate the state-specific 1-step-ahead predicted variances $\sigma^2_{i,T+1|T,r} = \omega_{i,r} + \alpha_{i,r}\big(y_{i,T} - \mu_i\big)^2 + \beta_{i,r}\sigma^2_{i,T,r}$, for both states $r = 1, 2$;

Estimate the state probabilities at time $T+1$ as given in Equation (19) and draw S times a state $r(s)$ for simulation s;

Assign per simulation s the state-specific variance $\sigma^{2(s)}_{i,T+1|T} = \sigma^2_{i,T+1|T,r(s)}$ for state $r(s) = 1, 2$;

Obtain $\tilde{y}^{(s)}_{i,T+1|T} = \mu_i + \sigma^{(s)}_{i,T+1|T} z^{(s)}_{i,t,r}$, for $s = 1, ..., S$ (recall μ_i is constant across states);

else

For all $s = 1, ..., S$ simulated paths:

Draw a new state $r(s)$, from the state transition probabilities given the state simulated at time $T+j-1$ for this path, using the probabilities as in $\boldsymbol{P}_i$, using the column corresponding to the state at time $T+j-1$;

Construct $\sigma^{2(s)}_{i,T+j|T} = \omega_{i,r(s)} + \alpha_{i,r(s)}\big(\tilde{y}^{(s)}_{i,T+j-1|T} - \mu_i\big)^2 + \beta_{i,r(s)}\sigma^{2(s)}_{i,T+j-1|T}$,

Obtain $\tilde{y}^{(s)}_{i,T+j|T} = \mu_i + \sigma^{(s)}_{i,T+j|T} z^{(s)}_{i,t,r}$.

end if

end for

Construct $\text{GaR}^M_{i,t+h|t}(p)$ for country i as the p-quantile from the values of all S simulated paths for time $T+h$: $Q_p(\tilde{y}^{(s)}_{i,T+h|T})$.

Algorithm 1: Pseudocode for the h-step ahead MS-GARCH GaR for country i at time T

4.3.4 Multivariate GARCH estimation

For multivariate GARCH, we rely for our GaR estimations on simulated returns from a multivariate normal distribution. We start by estimating the model, using the data of $y_{i,t}$ for $i = 1, ..., N$, up to time T. Next, we simulate 1-day-ahead returns for all countries from the model for S simulations. Based on these simulations, we estimate the 1-day-ahead conditional covariance matrix $\boldsymbol{H}_{t+1}$, using the simulated returns r_t and the 'current' value of $\boldsymbol{H}_t$. We repeat this iterative procedure until we reached the estimations of h-step ahead returns. Next, we take the p-quantile to obtain the marginal Growth-at-Risk from our multivariate GARCH model.

We compare our multivariate GARCH model by univariate GARCH, for which we assume normal distributed errors, to investigate how the incorporation of the correlations affects the results. In this case, the GaR is calculated as follows. First, we estimate for country i a GARCH model, based on the GDP growth rates up to time $t-1$ and obtain the parameters $\hat{\mu}_i$, $\hat{\omega}_i$, $\hat{\alpha}_i$ and $\hat{\beta}_i$, and the predicted volatility $\hat{\sigma}_{i,t+1}$. From this, we simulate for S simulations the next value of $y_{i,t+1}$ from $N(\hat{\mu_i}, \hat{\sigma}_{i,t+1})$. For each simulation, we now predict the volatility for time $t+2$ as $\sigma_{i,t+2} = \sqrt{\mu_i + \hat{\alpha}_i y_{i,t+1}^2 + \hat{\beta}_i \hat{\sigma}_{i,t+1}^2}$, and repeat this procedure. The GaR is obtained as the p-quantile across the S path simulations for horizon h.

4.3.5 Joint Growth-at-Risk estimation

The joint GaR is determined by the Bootstrap Joint Prediction Region (BJPR) of Wolf and Wunderli [2015], which is calculated as

$$\text{GaR}_{i,t+h|t}^{BJPR}(p) = \mu_{i,t+h|t} + d_p^{(1)} \sigma_{i,t+h|t}, \tag{21}$$

Here $d_p^{(1)}$ denotes the p-quantile of $U_t^{(1)}$, which is the smallest value of the standardized growth rates $\hat{Z}(y_{i,t}) = (y_{i,t} - \bar{y}_{i,t})/\sigma(y_{i,t})$:

$$U_t^{(1)} = \min_{i=1,...,N} \hat{Z}(y_{i,t}), \quad \forall t = 1, ..., T. \tag{22}$$

The joint Growth-at-Risk for h-step-ahead forecasts at time T is predicted as follows. First, we calculate the bootstrapped marginal GaR predictions $\text{GaR}_{i,t+h|t}^{M}(p)$ as described previously. Next, we standardize all bootstrapped paths for time T at horizon h and obtain the values $\hat{Z}(\tilde{y}_{i,T+h|T}^s)$, for $i = 1, ..., N$. Here, the mean $\bar{y}_{i,t}$ and standard deviation $\sigma(y_{i,t})$ are calculated from the S simulated paths. Next, we compute the p-quantile of the minimum values across the countries for all these standardized bootstrapped paths, as

$$d_p^{(1)} = Q_p\Big[\min_{i=1,...,N}(\hat{Z}(\tilde{y}_{i,T+h|T}^s)\Big], \text{ for } s = 1, ..., S. \tag{23}$$

Finally, we compute the Joint Growth-at-Risk as in Equation (21).

4.4 Historical Growth-at-Risk

To evaluate the performance of our methods, we compare them with a simple historical approach. The Marginal Historical Growth-at-Risk, denoted by $\text{GaR}_{i,T}^{MH}(p)$, is computed as the historical univariate quantile per country i, calculated at each time, based on the past observations.

The historical benchmark for joint GaR at time T is constructed by using the BJPR in a historical approach, following Brownlees and Souza [2021]. For all time $t = 75, ..., T$, the growth rates are first standardized by their country-specific sample mean $\bar{y}_{i,t}$ and standard deviation $\sigma(y_{i,t})$, based on the past observations until $t-1$, as $y_{i,t}^{Z,Hist} = (y_{i,t} - \bar{y}_{i,t})/\sigma(y_{i,t})$. The first 74 observations are used as a burn-in period to avoid extreme values. Next, analogously to Equation (22), the smallest value U_t^H of the standardized growth rates at time t is retrieved from the values of $y_{i,t}^{Z,Hist}$.

From these minimum values, we take the historical quantile at probability level p of all calculated values of $y_{i,t}^{Z,Hist}$ for time $1, ..., T$, as

$$d_{p,T}^{Z,Hist} = Q_p\Big[\min_{i=1,...,N}(y_{i,t}^{Z,Hist})\Big]. \tag{24}$$

Finally, we calculate the historical joint (HJ) Growth-at-Risk for all countries at time T as

$$\text{GaR}_{i,T}^{HJ}(p) = \bar{y}_{i,T} + \sigma(y_{i,T})d_{p,T}^{Z,Hist}, \text{ for } i = 1, ..., N, \tag{25}$$

where we use the sample means and standard deviations at time T as input.

4.5 Evaluation

To evaluate our results, we first report the average empirical coverage of a method for the considered probability level p. Second, we prefer GaR estimations that are relatively high while also predicting the coverage correctly. For this purpose, we define the average Quantile Length (QL), based on the work of Brownlees and Souza [2021]. The QL of respectively the marginal and joint GaR estimations is calculated as follows:

$$QL^M = \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}Q_{0.995}(\boldsymbol{y}_i) - \text{GaR}_{i,t|t-h}^{M}\Big), \quad QL^J = \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}Q_{0.995}(\boldsymbol{y}_i) - \text{GaR}_{i,t|t-h}^{J}\Big), \tag{26}$$

where $Q_{0.995}(\boldsymbol{y}_i)$ denotes the empirical quantile of the time series $\boldsymbol{y}_i$, estimated from the whole data set. The quantile of 0.995 is chosen as the most estimations of our methods fall below this value.

Next, we define the Hits Given Failure (HGF) measure for the joint GaR. This is defined as the number of countries for which the joint GaR is violated, given that the joint GaR is violated for at least one country, meaning that the GaR jointly is violated. Mathematically, this is denoted as

$$\text{HGF} = E(V_i \mid V_i \geq 1), \quad \text{for } V_i = \sum_{i=1}^{N} \mathbb{1}\left[y_{i,t+h|t} \leq \text{GaR}^{J}_{i,t+h|t}(p)\right]. \tag{27}$$

The HGF indicates whether the joint GaR is violated by the countries' time series independently, which is the case for a low HGF value.

In addition, we evaluate the performances of our methods by three types of tests. First, we perform two Dynamic Quantile Tests to test whether our methods unconditionally have a correct coverage and the GaR violations are independent over time, which we apply on both marginal and joint GaR. In addition, we compute for the marginal GaR estimations the Tick Loss for all methods and apply a comparative backtest to compare our estimation methods. Thirdly, we evaluate both marginal and joint GaR by the Dynamic Binary Response Test of Dumitrescu et al. [2012], which shows good results for small samples, to evaluate if this test shows different results than the DQ-type tests. These measures are explained in the next subsections. We evaluate our estimation methods for horizons $h = 1, 2, 3$ and coverage levels $1 - p = 0.75, 0.95, 0.99$.

4.5.1 Dynamic Quantile tests

The Dynamic Quantile (DQ) test was introduced by Engle and Manganelli [2004]. It is constructed by utilizing the criterion that each period the probability of exceeding the GaR must be independent of all the past information. We perform the DQ test for both marginal and joint GaR.

The DQ test for marginal GaR at a given probability level p is defined as follows. The hit function $H_{i,t}$ is equal to $-p$ when the observation $y_{i,t}$ is below the GaR, and $1 - p$ otherwise. Mathematically, this is denoted for the h-step ahead GaR as

$$H_{i,t+h|t} = \mathbb{1}[y_{i,t} < \text{GaR}^{M}_{i,t+h|t}(p)] - p, \tag{28}$$

which should by construction have zero mean, if the GaR has a coverage level p.

For marginal GaR, we perform three DQ-type tests. First, we test whether the hits have unconditional zero mean, which means testing whether the GaR has unconditionally the correct coverage level p. Therefore, we regress the hit functions on only an intercept so that the number of explanatory variables $R = 0$.

Secondly, we test whether the hit functions are independent over time. This property is important since it should not be possible for a good GaR estimation method to predict a violation today if there was one yesterday. Therefore, we regress the hits on their R lags as follows:

$$H_{i,t} = \beta_0 + \sum_{r=1}^{R} \beta_i H_{i,t-r} + u_t, \tag{29}$$

for which optimal GaR forecasts generate by construction zero-mean hits.

Thirdly, we also test whether the hit sequences are independent of the lagged GDP growth rates, which can be tested from the following regression:

$$H_{i,t} = \beta_0 + \sum_{r=1}^{R} \beta_r y_{i,t-r} + \nu_t, \tag{30}$$

In our analysis, we use $R = 4$ lags to perform our regressions. All independence tests use in line with Engle and Manganelli [2004] the null hypothesis that $\beta_0 = ... = \beta_R = 0$, which is tested by an augmented Wald test, which is for 1-step-ahead predictions defined as

$$\frac{\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}}{p(1-p)} \sim \chi^2(R+1), \tag{31}$$

where $\boldsymbol{\beta}$ is a vector of the beta coefficients in the regression. In addition, $\boldsymbol{X}$ denotes the $T \times (R+1)$ matrix of explanatory variables, where the first column consists of ones. Note that T denotes there the number of time periods for which the regression is performed, which varies across regressions. For $h > 1$, we use Newey-West HAC standard errors. In Section 5, we report for the marginal GaR the number of countries for which the test is passed for our chosen significance level.

For joint GaR, we report the mean uniform coverage and perform two Dynamic Quantile tests. However, we perform here the tests on the Uniform Hit function (UH), which is defined for time t as $-p$ if the GaR for all countries is not violated, and $1-p$ if the GaR is violated for at least one country. Mathematically, this is denoted as

$$UH_{t+h|t} = \mathbb{1}\Big[\sum_{i=1}^{N} \mathbb{1}[y_{i,t}^{BJPR} < \text{GaR}_{i,t+h|t}(p)] \geq 1\Big] - p. \tag{32}$$

The two DQ-type tests are performed similarly as for marginal GaR, with the difference that we for joint GaR test whether the uniform hits have unconditional zero mean and are independent over time.

In our analysis, we use an adaptation of the implementation of Brownlees and Souza [2021].

4.5.2 Comparative Tick Loss backtest

Next, we compare the performance of marginal GaR forecasts of different estimation models by the Tick Loss (TL) of Giacomini and Komunjer [2005]. This loss function has the advantage that it considers relative evaluation, taking also the difference between the estimated GaR and the observation into account. This feature is suitable for our research, as we consider monthly GDP data, which means that the number of violations may be relatively small.

The Tick Loss is an appropriate loss function for this purpose, as the function is designed to retrieve an optimal forecast for the object of interest, which is in this case equal to the p-quantile of $y_{i,t+1}$. The Tick Loss considers the negative hit function, multiplied by the actual difference between the estimated GaR and the observation $y_{i,t}$, averaging these values over time and all countries, resulting in the following function:

$$\mathrm{TL}_{i,t}(p) = \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}(y_{i,t} - \mathrm{GaR}_{i,t|t-h})(p - \mathbb{1}[y_{i,t} < \mathrm{GaR}_{i,t|t-h}])\Big). \tag{33}$$

We report the average value of the Tick Loss for each method, which we multiply by 100 for readability.

In addition, we perform a comparative test on the Tick Loss, to investigate whether the models' performances differ significantly. We use the comparative backtesting framework of Nolde and Ziegel [2017], where the difference between the scoring functions of two estimation methods is tested. In the scoring function, the estimated GaR and the observations $y_{i,t}$ at some quantile p are compared. We use as our scoring function $S(\mathrm{GaR}, y)$ the Tick Loss.

Given the estimations of two methods $\mathrm{GaR}^A_{i,t}$ and $\mathrm{GaR}^B_{i,t}$, we define

$$\begin{aligned}\lambda^* &= \limsup_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} S(\mathrm{GaR}^A_{i,t}, y_{i,t}) - S(\mathrm{GaR}^B_{i,t}, y_{i,t})\\ \lambda_* &= \liminf_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} S(\mathrm{GaR}^A_{i,t}, y_{i,t}) - S(\mathrm{GaR}^B_{i,t}, y_{i,t}).\end{aligned} \tag{34}$$

We use the Equal Forecasting Performance test of Gneiting and Ranjan [2011], with the following test statistic:

$$Q_{EFP} = \frac{\Delta_T \bar{S}}{\hat{\sigma}_T/\sqrt{T}} \sim N(0,1), \tag{35}$$

where the numerator is defined as

$$\Delta_T \bar{S} = \frac{1}{T}\sum_{t=1}^{T} S(\mathrm{GaR}^A_{i,t}, y_{i,t}) - S(\mathrm{GaR}^B_{i,t}, y_{i,t}) \tag{36}$$

and $\hat{\sigma}_T$ is a HAC estimator of the asymptotic variance [Andrews, 1991]. Considering a probability level ν, we reject the hypothesis that $\lambda^* \leq 0$, i.e. method A performs significantly worse than method B, if $1 - \Phi(Q_{EFP}) \leq \nu$, where $\Phi(.)$ denotes the CDF of the normal distribution. Similarly, we reject the hypothesis that $\lambda_* \geq 0$, i.e. that method B performs significantly worse than method A, if $\Phi(Q_{EFP}) \leq \nu$. If we cannot reject both tests, the methods do not differ significantly from each other.

We perform the tests for all countries and report in the comparison between two methods the percentage of countries for which both methods perform significantly better than the other method.

4.5.3 Dynamic Binary test

In addition, we also test for correct conditional coverage by the Dynamic Binary (DB) test in Dumitrescu et al. [2012]. As they show that the DB test exhibits good small sample properties, this test is suitable for our research. The test uses the following Dynamic Binary Response model, in which the conditional probability of violation at time t is given by

$$P(I_t(p) = 1 \mid \mathcal{F}_{t-1}] = E[I_t(p) \mid \mathcal{F}_{t-1}] = F(\pi_t), \tag{37}$$

where $F(.)$ denotes the CDF of the conditional violation probability, where we use the Dynamic Logit model, i.e. $F(\pi_t) = \exp(\pi_t)/(1 + \exp(\pi_t))$ and $I_t(p)$ is an violation indicator function. For marginal GaR, $I_t(p)$ is equal to $\mathbb{1}[y_{i,t} < \text{GaR}_{i,t+h|t}]$ for probability level p. For joint GaR, we define $I_t(p)$ as the indicator function $\mathbb{1}[y_{i,t}^{BJPR} < \text{GaR}_{i,t+h|t}(p)] \geq 1\Big]$, analogously to its definition. We assume that the index π_t satisfies the following autoregressive model:

$$\pi_t = c + \beta\pi_{t-1}, \tag{38}$$

which is optimized by the following likelihood function:

$$\ln L(\beta, c; I(p), \pi_{t-1}) = \sum_{i=1}^{T} \Big[I_t(p) \ln F(\pi_t(\beta, c, \pi_{t-1})) + (1 - I_t(p)) \ln(1 - F(\pi_t(\beta, c, \pi_{t-1}))) \Big], \tag{39}$$

where we use the constrained maximum likelihood estimation method of Kauppi and Saikkonen [2008]. For details regarding this procedure, we refer to this paper.

We test whether our estimation methods yield correct conditional coverage by testing whether $p = F(\pi_t(\beta, c, \pi_{t-1}))$, which is equivalent to $\beta = 0$ and $c = F^{-1}(p)$. Following Dumitrescu et al. [2012], we perform a Likelihood Ratio test, as this test has good small sample properties in terms of power. This test takes the following form:

$$DB_{LR} = -2\Big[\ln L(0, F^{-1}(p); I(p), \pi_{t-1}) - \ln L(\hat{\beta}, \hat{c}; I(p), \pi_{t-1})\Big] \sim \chi^2(1), \quad (40)$$

where $\hat{\beta}$ and $\hat{c}$ are the estimated parameters of the binary-choice model, without using the test conditions. For marginal GaR, we perform the DB test for each country individually and report the percentage of countries that pass the test at a 5% significance level. For joint GaR, we report the p-value of the test.

5 Results

5.1 Univariate GARCH

5.1.1 Marginal GaR

We start the analysis of Growth-at-Risk by the univariate GARCH models. First, we estimate the GaR for every country individually, using QML. Table 3 shows the results for the historical GaR, GARCH and GJR-GARCH using a constant and autoregressive mean, with one and three lags. For clarity, we present here the results for the coverage levels 75% and 99%, as the pattern of the results for 95% is mainly similar to the results for 99%. The Quantile Length is evaluated in Subsection 5.2 together with the results of Pooled GARCH.

For the 1-step-ahead predictions, we observe that all univariate GARCH models consistently decrease the Tick Loss in comparison to the historical GaR. For the coverage level 75%, the TL is decreased by 22 percent for the constant mean models and by 64 percent for the autoregressive mean models. For the coverage levels of 95% and 99%, the TL reduction of the constant mean models is 38 percent, although the autoregressive mean models are still the best in terms of the TL. In addition, all methods have a more accurate empirical coverage than the historical benchmark, also shown in a better performance for the unconditional DQ test. For the multistep-ahead predictions, the GARCH models with constant mean show however a decreasing mean coverage and an increasing TL. By contrast, the historical GaR approximately yields the same results over the time horizons. The models with a time-varying mean keep their empirical coverage relatively stable compared to the constant mean models. Although their TL values increase relatively to the historical GaR, they still perform better. The power of the AR(3)-GARCH becomes in particular clear for 3-step-ahead predictions. Here, it yields the most accurate empirical coverage, the lowest TL value and the most countries for which the DB test is passed, showing independence from the first lag. Therefore, the added lags show to stabilize the results of the GaR estimations. Overall, the AR-GARCH models yield the best performance, where the AR(3)-GARCH model is the most reliable estimation method in the long term.

Table 3: Marginal results for historical GaR (Hist.) and univariate GARCH models estimated with QML, evaluated at the empirical coverage level (Cov.); the percentage of countries for which the Dynamic Quantile test is passed, using respectively a constant (Unc.), 4 hit lags (Hits) and 4 lags of the GDP growth rate; the average Tick Loss times 100 (TL) and the percentage of countries for which the Dynamic Binary test (DB) is passed.

Coverage			75%						99%			
$h = 1$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
Hist.	79.63	34.78	0.00	0.00	6.7425	39.13	98.16	56.52	0.00	13.04	2.2490	65.22
GARCH	78.58	56.52	0.00	4.35	5.2379	65.22	99.00	86.96	52.17	78.26	1.4150	86.96
GJR	78.62	60.87	0.00	8.70	5.2344	65.22	99.01	86.96	56.52	78.26	1.4047	86.96
AR-GARCH	77.45	60.87	0.00	8.70	2.4353	73.91	98.98	78.26	73.91	78.26	1.1156	91.30
AR-GJR	77.45	69.57	0.00	8.70	2.4251	73.91	98.98	78.26	69.57	78.26	1.1147	91.30
AR(3)-G	76.87	86.96	0.00	34.78	2.5644	95.65	98.64	73.91	52.17	52.17	1.1717	78.26
$h = 2$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
Hist.	79.59	73.91	0.00	21.74	6.7613	43.48	98.08	91.30	69.57	82.61	2.2796	65.22
GARCH	74.67	86.96	0.00	56.52	6.1342	78.26	98.88	95.65	100.00	91.30	3.7611	86.96
GJR	74.81	91.30	0.00	56.52	6.1626	73.91	98.92	95.65	100.00	91.30	3.9834	91.30
AR-GARCH	76.60	78.26	0.00	0.00	4.1519	73.91	99.27	95.65	95.65	86.96	1.6774	95.65
AR-GJR	76.68	78.26	0.00	0.00	4.1530	73.91	99.30	95.65	95.65	91.30	1.6748	95.65
AR(3)-G	76.50	95.65	8.70	26.09	4.1278	95.65	99.23	95.65	82.61	69.57	1.8868	100.00
$h = 3$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
Hist.	79.52	73.91	0.00	0.00	6.7759	43.48	98.01	91.30	86.96	86.96	2.2995	60.87
GARCH	73.10	82.61	0.00	0.00	6.8647	47.83	99.02	82.61	60.87	86.96	68.0349	78.26
GJR	73.26	82.61	0.00	4.35	6.9119	52.17	99.03	82.61	60.87	86.96	74.5033	78.26
AR-GARCH	76.57	82.61	4.35	0.00	5.5214	73.91	99.47	91.30	30.43	100.00	2.0781	91.30
AR-GJR	76.68	82.61	4.35	0.00	5.5428	69.57	99.47	86.96	30.43	95.65	2.0741	86.96
AR(3)-G	75.69	86.96	4.35	0.00	5.4748	78.26	99.41	95.65	30.43	100.00	1.9812	95.65

We observe for the coverage levels of 95% and 99% that the TL of the constant mean models exponentially rises for a larger time horizon. This feature is caused by applying the models to Turkey, it leads to some extreme GaR estimations. This is further investigated in Subsubsection 5.1.2. When we exclude Turkey, the TL values of the models with constant mean are for 3-step-ahead predictions approximately equal to the TL of the Historical GaR, just as shown for the coverage level 75%.

Evaluating the independence of the hits, we observe for the 75% coverage level that all univariate GARCH models pass the Dynamic Binary test for more countries. This suggests that the violations are also less dependent over time when we consider one hit lag. On the other hand, the DQ-Hits test, which is performed with four lags, is not passed for all countries. Therefore, we conclude that the hits of the autoregressive models show more evidence of independence. However, this finding is not robust for the usage of multiple lags. For the 99% coverage level, this result is only found for the 3-step-ahead estimations.

Next, we compare the performance of the GARCH and GJR-GARCH models. The empirical coverage stays approximately the same for all horizons and coverage levels, which also applies to the models that use a time-varying mean. The tests and Tick Losses do also not show clear differences. Thus, the usage of an asymmetric term does not significantly improve the GaR estimations in the context of univariate GARCH models.

Finally, we observe from the results of the VAR-GARCH model in the Appendix, Section C that VAR-GARCH fails to correctly estimate the GaR compared to the other estimation methods, according to its too low empirical coverage level. Although this level slightly increases for a longer time horizon, it is still between 7 and 10 percent too low for all coverage levels at $h = 3$. This finding is supported by the small percentage of countries that pass both the unconditional DQ test and the DB test. However, the clearest indication that this model performs poorly is found in the Tick Loss, which is for almost all coverage levels and horizons higher than the historical benchmark. Therefore, we conclude that using the cross-sectional parameters to model the mean does not have added value for our GaR estimations.

5.1.2 Tick Loss Comparative Backtest

To investigate the Tick Loss in more detail, we perform the comparative backtest country-by-country. Table 4 reports the results of these tests. As expected, we observe that the historical Growth-at-Risk has a significantly higher Tick Loss than all methods for the majority of the countries. However, historical GaR beats the models with a constant mean for a minority of the countries. For the 3-step-ahead estimations of the 99% coverage level, historical GaR even outperforms the models with constant mean for the majority of the countries. The historical GaR even beats the models with an AR(1) mean for around 40 percent of the countries and is only outperformed for 17 percent of the countries. This finding contrasts with Table 3, where we found that the historical GaR has a higher TL on average. However, the AR(3) mean models here still outperform the historical GaR for around 40 percent of the countries. We conclude that the univariate GARCH models outperform the historical GaR for shorter time windows, and the AR(3)-GARCH again is the clear winner.

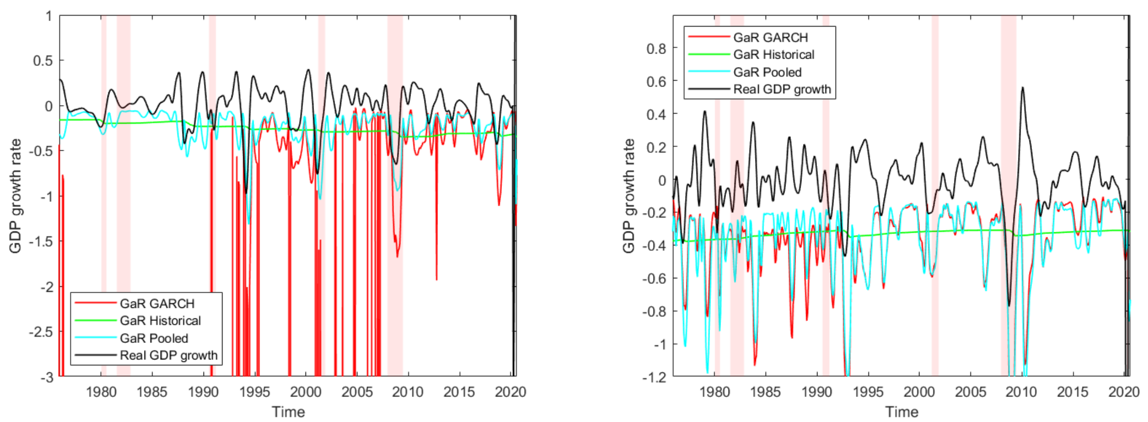
Table 4: Tick Loss Backtest table for Historical GaR (Hist.) and the univariate models, where G denotes GARCH. The displayed result is the percentage of countries for which the method of the corresponding column performs significantly better at a 5% level than the method in the corresponding row.

Coverage			75%							99%		
$h = 1$	Hist.	G	GJR	AR-G	AR-GJR	AR(3)-G	Hist.	G	GJR	AR-G	AR-GJR	AR(3)-G
Hist.		0.17	0.17	0.00	0.00	0.00		0.04	0.04	0.00	0.00	0.00
G	0.48		0.22	0.00	0.00	0.00	0.87		0.22	0.00	0.00	0.00
GJR	0.48	0.26		0.00	0.00	0.00	0.87	0.00		0.00	0.00	0.00
AR-G	0.96	0.91	0.91		0.26	0.00	1.00	1.00	1.00		0.04	0.00
AR-GJR	0.96	0.91	0.91	0.22		0.00	1.00	1.00	1.00	0.04		0.00
AR(3)-G	0.74	0.74	0.70	0.48	0.48		0.91	0.83	0.83	0.52	0.52	
$h = 2$	Hist.	G	GJR	AR-G	AR-GJR	AR(3)-G	Hist.	G	GJR	AR-G	AR-GJR	AR(3)-G
Hist.		0.22	0.22	0.00	0.00	0.00		0.35	0.35	0.04	0.04	0.00
G	0.57		0.30	0.00	0.00	0.00	0.52		0.26	0.00	0.00	0.00
GJR	0.52	0.13		0.00	0.00	0.00	0.48	0.00		0.00	0.00	0.00
AR-G	0.74	0.57	0.57		0.35	0.00	0.87	0.78	0.78		0.04	0.00
AR-GJR	0.74	0.61	0.61	0.22		0.00	0.78	0.83	0.87	0.13		0.00
AR(3)-G	0.52	0.52	0.52	0.48	0.48		0.91	0.91	0.91	0.78	0.74	
$h = 3$	Hist.	G	GJR	AR-G	AR-GJR	AR(3)-G	Hist.	G	GJR	AR-G	AR-GJR	AR(3)-G
Hist.		0.17	0.17	0.04	0.04	0.04		0.52	0.57	0.39	0.43	0.13
G	0.61		0.52	0.09	0.04	0.04	0.09		0.35	0.13	0.09	0.00
GJR	0.61	0.04		0.04	0.04	0.04	0.09	0.00		0.13	0.09	0.00
AR-G	0.13	0.17	0.17		0.26	0.00	0.17	0.61	0.57		0.22	0.00
AR-GJR	0.13	0.22	0.22	0.04		0.09	0.17	0.61	0.57	0.22		0.00
AR(3)-G	0.13	0.22	0.22	0.78	0.70		0.39	0.78	0.74	1.00	0.96	

Only for Turkey, GARCH and GJR are beaten consistently by the historical approach at all investigated horizons and coverage levels. To gain more insight into these results for Turkey, we plot the GDP growth rate of Turkey together with the 95% GaR from a historical approach, univariate GARCH with constant mean and Pooled GARCH. For comparison purpose, we show this figure also for Sweden, where the univariate GARCH models consistently beat the historical benchmark. Figure 2 shows the two figures.

We observe for both countries that the historical GaR has a smooth pattern, which results for Sweden in a relatively high Tick Loss, as it does not adapt quickly to recent changes. However, we observe for Turkey that the GaR estimated by univariate GARCH shows a very unstable pattern, resulting in strongly negative GaR estimations. Taking a closer look at the results, we observe that the growth rate of Turkey in the first and middle part of the dataset has a relatively long series of stable negative growth rates. This data feature leads to highly negative standardized growth rates for the estimations from the GARCH model with a constant mean. This strongly influences the results of the GaR estimations in the bootstrapping algorithm, which yields very low GaR estimations that are not corrected by a time-varying mean. This effect remains relatively long, as the sum of the parameters α and β is close to 1, and the weight of the past volatility for Turkey is high. Therefore, we obtain high differences between the GaR and the actual growth rate, which causes the high Tick Loss. The power of Pooled GARCH is already shown in this figure, as it yields a smoothed line of the GaR estimations over time, where the GaR still captures most of the downside risk events. On the other side, the estimated GaR of Sweden is only slightly changed and approximately yields the same result.

Figure 2: Comparison of the GDP growth rates and the 3-step-ahead 95% GaR estimations with univariate GARCH and Pooled GARCH (both with a constant mean) and the historical GaR. For readability, the y-axes of the figures are truncated at the downside.



(a) Turkey (b) Sweden

Since the last six months of our data set are months of the coronavirus recession affects the results, namely March 2020 to August 2020, we finally investigate if this period affects the results. Whereas for most performance measures the general pattern does not change, the Tick Loss is highly affected by the six months of the coronavirus recession. The reduction percentage of the Tick Loss when we leave out the last six months of our data set is shown in Table 5.

We observe that the historical GaR relatively is the least affected by the coronavirus recession at 1-step-ahead predictions. However, the coronavirus recession has the most impact on the GARCH models that use an autoregressive mean. The reason is that the AR-GARCH

Table 5: Tick Loss reduction in % for the exclusion of the last six months in the data set, for the univariate GARCH models using QML.

Horizon	$h = 1$			$h = 2$			$h = 3$		
Coverage	75%	95%	99%	75%	95%	99%	75%	95%	99%
Hist.	23.6	42.5	68.6	23.6	42.4	68.0	23.6	42.3	67.6
GARCH	27.1	49.5	76.1	25.9	23.8	35.0	24.0	1.7	0.8
GJR	26.8	48.5	75.3	25.8	22.3	32.8	23.9	1.4	0.6
AR-GARCH	53.9	76.7	91.9	44.0	65.1	84.1	36.8	52.1	69.2
AR-GJR	53.7	76.6	91.9	43.9	65.0	84.1	36.9	52.1	69.2
AR(3)-G	90.6	94.0	97.8	79.8	86.9	94.8	37.0	55.0	74.9

models follow the trend of the GDP growth rate data the most closely since their means are updated with high weight on the last observation. Generally, this reduces the Tick Loss, as the Growth-at-Risk updates quickly for recent changes in the economy. In contrast, the Tick Loss is in the coronavirus recession highly affected by these 'extreme' events that never happened before in the data. Especially the fact that the data contains very negative values, followed by highly positive values, accuses a high difference between the GaR estimations and the actual GDP growth rates. This effect is even stronger for an autoregressive mean using three lags.

The Tick Loss reduction is very small for the 3-step-ahead forecasts at the coverage levels of 95% and 99%. However, since the Tick Loss was also affected by the extreme GaR estimations for Turkey at this time horizon, this result is only caused by that feature. Overall, we conclude that the AR(3)-GARCH models even perform better when excluding the first months of the coronavirus recession, which is in the context of this research a unique event that does not follow the normal cyclical rules of the economy. Given that such an event is not likely to happen often again, this emphasizes the power of the AR-GARCH models for the GaR estimation in conventional world situations.

5.1.3 Joint GaR

Next, we evaluate the estimations of the joint Growth-at-Risk for the univariate GARCH models and show the results in Table 6. Our historical approach has for the coverage level 75% consistently a too high coverage level. This feature also applies to the univariate GARCH models, apart from the AR(3) model, which shows good coverage results for 1-step-ahead GaR estimations. In addition, this model also passes the unconditional DQ test and DB test, show-

Table 6: Joint GaR results for univariate GARCH models estimated with QML, evaluated at the mean coverage level (Cov.) and the p-value of the Dynamic Quantile test, using respectively a constant (Unc.) and 4 hit lags (Hits). In addition, the Quantile Length (QL) and the p-value of the DB test are reported.

Coverage	75%						99%					
$h = 1$	Cov.	Unc.	Hits	QL	DB	HGF	Cov.	Unc.	Hits	QL	DB	HGF
Hist.	84.23	0.00	0.00	1.01	0.00	3.15	97.96	0.02	0.00	1.49	0.10	5.73
GARCH	81.45	0.00	0.00	0.86	0.00	1.42	98.33	0.12	0.00	1.15	0.35	3.33
GJR	81.45	0.00	0.00	0.87	0.00	1.40	98.70	0.49	0.00	1.17	0.79	3.86
AR-GARCH	81.82	0.00	0.00	0.67	0.00	1.46	97.77	0.00	0.00	0.72	0.05	2.67
AR-GJR	82.00	0.00	0.00	0.67	0.00	1.47	97.96	0.02	0.01	0.73	0.10	2.82
AR(3)-G	76.07	0.57	0.07	0.64	0.62	1.73	97.40	0.00	0.00	0.67	0.01	2.57
$h = 2$	Cov.	Unc.	Hits	QL	DB	HGF	Cov.	Unc.	Hits	QL	DB	HGF
Hist.	83.83	0.01	0.00	1.01	0.00	3.15	97.96	0.29	1.00	1.49	0.10	6.09
GARCH	87.73	0.00	0.00	2.07	0.00	1.88	98.88	0.86	1.00	4.34	0.95	5.17
GJR	88.10	0.00	0.00	2.29	0.00	1.88	98.88	0.86	1.00	4.89	0.95	5.17
AR-GARCH	91.45	0.00	0.00	0.94	0.00	1.98	99.26	0.67	0.30	1.40	0.82	10.50
AR-GJR	91.26	0.00	0.00	0.95	0.00	1.96	99.26	0.67	0.30	1.45	0.82	10.50
AR(3)-G	87.92	0.00	0.00	0.95	0.00	1.91	98.88	0.87	1.00	1.49	0.95	8.17
$h = 3$	Cov.	Unc.	Hits	QL	DB	HGF	Cov.	Unc.	Hits	QL	DB	HGF
Hist.	83.61	0.01	0.00	1.01	0.00	3.14	97.95	0.29	0.00	1.49	0.10	6.09
GARCH	93.67	0.00	0.00	32.08	0.00	2.56	99.44	0.46	0.91	106.22	0.53	8.67
GJR	93.85	0.00	0.00	37.65	0.00	2.61	99.44	0.46	0.91	125.27	0.53	8.67
AR-GARCH	97.77	0.00	0.00	2.69	0.00	4.92	99.44	0.46	0.91	8.62	0.53	13.67
AR-GJR	97.39	0.00	0.00	3.31	0.00	4.36	99.44	0.46	0.91	10.18	0.53	13.33
AR(3)-G	95.72	0.00	0.00	6.86	0.00	3.30	99.44	0.46	0.91	19.87	0.53	14.33

ing that it yields correct conditional coverage. However, in the longer term, we observe that the GARCH models return too high empirical coverage. This finding indicates that the joint GaR takes too negative values for longer time horizons. Consequently, this also yields very high Quantile Length values. We conclude that only our AR(3)-GARCH model performs relatively well in the short term. For larger time horizons, all our univariate GARCH models tend to overestimate the empirical coverage.

Since the estimations for Turkey yield very low GaR estimations, we also evaluate the joint GaR while excluding Turkey to investigate if this accuses the bad results. The results are shown in the Appendix, Table 21. Indeed, the Quantile Length is decreased strongly. For the 75% coverage level, we obtain at $h = 3$ a QL that is still higher than the historical benchmark, but also between the QL values of the AR(1) models and the AR(3) models. Although the QL values are also strongly decreased for the higher coverage levels, they are still clearly higher compared to the AR-based models. However, all methods yield a higher QL value than the historical benchmark, showing that even the autoregressive models lose prediction power over time. In addition, the empirical coverage is still clearly too high for multistep-ahead predictions. In Subsection 5.2, we investigate whether we can solve this by information pooling for the estimation of the GARCH parameters.

Considering the HGF values in Table 6, the HGF of the dynamic models is for the 75% coverage level lower than the historical GaR. For the constant mean GARCH models, this applies to all considered time horizons. For $h = 1, 2$, the HGF is even below 2 for all dynamic models. Thus, for the majority of the GaR violation cases, this is caused by only one GDP growth rate that takes a value below its joint GaR. On the contrary, the historical GaR is on average violated by three countries at that time. Thus, the dynamic models have a better capability to avoid that multiple countries violate together the estimations, and therefore are more robust to system-wide crashes. For the 95% and 99% coverage levels, the HGF values are clearly higher. This was expected since those violations are likely to occur in extreme economic times, where it is likely that multiple countries face a severe economic recession.

5.2 Pooled GARCH

5.2.1 Marginal GaR

Next, we use information pooling to investigate whether the assumption of equal parameters is valid and whether this enhances our estimations. We assume that all countries share the same GARCH parameters. Table 7 shows the results. First, we emphasize that Pooled GARCH has a significantly shorter estimation time since it estimates one GARCH model for all countries together. Apart from the estimation results, this is already a clear advantage.

The results from Table 7 show that information pooling successfully increases the stability of our models. We observe that for 1-step-ahead estimations, the average coverage level is slightly lower for the 75% coverage level and higher for the 99% coverage levels. Importantly, the Tick Loss is sharply decreased in comparison to the historical GaR for 1-step-ahead forecasts. This result is similar to the univariate GARCH models estimated without information pooling, although the Tick Loss is for 75% coverage even reduced slightly more for the AR(1) models. The reasoning behind this result was already shown in Figure 2: the historical GaR does not follow the GDP over time closely. However, in contrast to the univariate GARCH models, the mean coverage level does not decrease significantly for higher horizon steps. For multistep-ahead forecasts, the GARCH models with a constant yield a better average coverage level at the coverage level 95%. By contrast, we observe for the 99% (and 95%) coverage level that models with an autoregressive mean have a better coverage level. This is shown by the average empirical coverage, but also by the unconditional DQ test and the DB test, for which the tests are passed for more countries. On the other hand, the constant mean GARCH models show more evidence for independent hit sequences for multistep forecasts. The TL values for the AR-GARCH methods are for 3-step-ahead predictions also reduced in comparison to the univariate GARCH models. Therefore, we conclude that information pooling also stabilizes our models consistently at longer horizons, especially doing well for AR(3)-GARCH.

Table 7: Marginal results for univariate GARCH models estimated with Pooled GARCH.

Coverage			75%						99%			
$h = 1$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
Hist.	79.63	34.78	0.00	0.00	6.7425	39.13	98.16	56.52	0.00	13.04	2.2490	65.22
GARCH	77.66	73.91	0.00	0.00	5.2632	78.26	98.99	82.61	52.17	65.22	1.2953	86.96
GJR	77.67	73.91	0.00	0.00	5.2156	78.26	99.00	82.61	60.87	69.57	1.2855	86.96
AR-GARCH	76.65	69.57	0.00	4.35	2.3661	78.26	99.01	82.61	78.26	86.96	1.1059	86.96
AR-GJR	76.69	69.57	0.00	4.35	2.3656	73.91	98.98	78.26	78.26	78.26	1.1100	82.61
AR(3)-G	75.99	56.52	0.00	21.74	2.6383	65.22	98.52	73.91	43.48	47.83	1.2635	78.26
$h = 2$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
Hist.	79.59	73.91	0.00	21.74	6.7613	43.48	98.08	91.30	69.57	82.61	2.2796	65.22
GARCH	75.58	95.65	0.00	56.52	5.9290	86.96	97.77	69.57	86.96	69.57	1.8268	60.87
GJR	75.55	95.65	0.00	56.52	5.9234	86.96	97.79	69.57	86.96	73.91	1.8076	65.22
AR-GARCH	63.54	0.00	0.00	0.00	4.1870	0.00	96.43	26.09	43.48	43.48	1.6750	17.39
AR-GJR	63.53	0.00	0.00	0.00	4.1906	0.00	96.43	26.09	43.48	47.83	1.6761	17.39
AR(3)-G	75.76	95.65	0.00	0.00	4.9895	82.61	98.68	86.96	95.65	52.17	2.1035	91.30
$h = 3$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
Hist.	79.52	73.91	0.00	0.00	6.7759	43.48	98.01	91.30	86.96	86.96	2.2995	60.87
GARCH	76.33	95.65	0.00	13.04	6.2457	82.61	98.79	91.30	78.26	91.30	2.0105	91.30
GJR	76.33	95.65	0.00	13.04	6.2455	82.61	98.79	91.30	78.26	91.30	2.0106	91.30
AR-GARCH	69.40	60.87	0.00	0.00	5.3885	39.13	98.08	73.91	82.61	73.91	1.9281	65.22
AR-GJR	69.41	60.87	0.00	0.00	5.3946	39.13	98.07	73.91	82.61	73.91	1.9291	65.22
AR(3)-G	75.44	91.30	0.00	0.00	5.3195	73.91	97.88	65.22	86.96	60.87	1.9224	65.22

5.2.2 Tick Loss Comparative Backtests

Table 8 shows the comparative backtest results for the Tick Loss, that are performed for every single country. The pooled univariate GARCH model is outperformed for most countries in terms of the Tick Loss, apart from the univariate GARCH models where information pooling is not used. However, for the 99% coverage level, the pooled GARCH outperforms the non-pooled equivalent for multistep-ahead GaR estimations. For the models that use AR-based means, there is stronger evidence that information pooling is useful, in the longer term. Although for $h = 1$ the comparison for AR(1)-GARCH does not show a clear winner, the pooled variant outperforms for multistep-ahead estimations the non-pooled variant for almost all countries. For AR(3)-GARCH, this effect is in particular seen for 3-step-ahead predictions.

Table 8: Tick Loss Backtest table for the Pooled GARCH models, without asymmetry terms, tested against the univariate GARCH models (indicated by an added 'u'). The displayed result is the percentage of countries for which the method of the corresponding column performs significantly better than the method in the corresponding row.

Coverage			75%						99%			
$h = 1$	G	AR-G	AR3-G	uG	uAR-G	uAR3-G	G	AR-G	AR3-G	uG	uAR-G	uAR3-G
G		0.00	0.00	0.13	0.00	0.00		0.00	0.00	0.13	0.00	0.00
AR-G	0.96		0.00	0.91	0.13	0.00	1.00		0.00	0.96	0.13	0.00
AR3-G	0.74	0.48		0.74	0.48	0.13	0.83	0.52		0.83	0.52	0.00
uG	0.39	0.00	0.00		0.00	0.00	0.35	0.00	0.00		0.00	0.00
uAR-G	0.96	0.09	0.00	0.91		0.00	1.00	0.04	0.00	1.00		0.00
uAR3-G	0.74	0.43	0.13	0.74	0.48		0.83	0.57	0.00	0.83	0.52	
$h = 2$	G	AR-G	AR3-G	uG	uAR-G	uAR3-G	G	AR-G	AR3-G	uG	uAR-G	uAR3-G
G		0.00	0.00	0.22	0.00	0.00		0.00	0.00	0.74	0.04	0.00
AR-G	0.87		0.83	0.87	1.00	0.00	0.91		0.87	0.96	0.96	0.00
AR3-G	0.78	0.00		0.65	0.04	0.00	0.96	0.00		1.00	0.61	0.00
uG	0.35	0.00	0.00		0.00	0.00	0.00	0.00	0.00		0.00	0.00
uAR-G	0.65	0.00	0.39	0.57		0.00	0.35	0.00	0.00	0.83		0.00
uAR3-G	0.52	0.17	0.48	0.52	0.48		0.74	0.17	0.43	0.91	0.74	
$h = 3$	G	AR-G	AR3-G	uG	uAR-G	uAR3-G	G	AR-G	AR3-G	uG	uAR-G	uAR3-G
G		0.00	0.00	0.22	0.04	0.04		0.00	0.00	0.61	0.43	0.17
AR-G	0.39		1.00	0.43	0.91	0.83	1.00		0.04	0.96	0.96	0.96
AR3-G	0.22	0.00		0.17	0.52	0.39	0.91	0.78		0.91	1.00	0.96
uG	0.43	0.00	0.00		0.09	0.04	0.00	0.00	0.00		0.13	0.00
uAR-G	0.04	0.00	0.09	0.17		0.00	0.17	0.00	0.00	0.57		0.00
uAR3-G	0.09	0.00	0.13	0.22	0.70		0.43	0.00	0.00	0.74	1.00	

Comparing the information pooling methods, we observe that the constant mean models are outperformed by the autoregressive mean models for the majority of the countries in almost every case. In addition, Pooled GARCH with a constant mean never outperforms the models with an AR-based mean. The AR(1) and AR(3) models show an ambiguous pattern: for the coverage level 75%, the AR(3) model outperforms AR(1) for $h = 1, 2$, but for $h = 3$, AR(3) is always outperformed. For the coverage level 99%, the AR(3) model only outperforms the AR(1) model for $h = 2$, although the average Tick Loss of both models was approximately equal. The AR(3) model is almost always outperformed by the AR(1) model, apart from one country, which strongly affects the mean Tick Loss. In general, the AR(3) model is more appropriate for the 75% coverage level, where the AR(1) model shows better results for the GaR estimations in the tail.

5.2.3 Marginal Quantile Lengths

In addition, we also evaluate the Quantile Lengths of the marginal GaR estimations of the univariate GARCH models with and without estimation pooling. Table 9 shows the results[6]. For 1-step-ahead predictions, all univariate and pooled GARCH methods yield as expected a lower Quantile Length in comparison to the historical benchmark. Here, the models with an autoregressive mean take a lower value than the GARCH models with a constant mean. The pooled GARCH models reduce the QL for the constant mean models, but we do not observe a significant difference for the autoregressive mean models. This is in line with the observations of the Tick Loss, where the use of information pooling also only reduced the TL for the constant mean models.

For multistep-ahead predictions, only the autoregressive models keep their QL value clearly under the QL of the historical GaR. In contrast, the univariate GARCH models again show the effects of the exploding negative estimations from Turkey. From the results of the 75% coverage level, we observe that the GARCH models also without this feature yield a slightly higher QL value in comparison to the historical benchmark. However, the pooled GARCH models also face an increasing QL, although the autoregressive mean models still outperform the historical GaR. These good but diminishing results are again in line with Table 7. Here was observed that the average Tick Loss is below the value of the historical GaR, although the value increases for a

[6]For the results of AR(3)-GARCH, the quantile at 99.5% for the whole data set is exceeded for an average amount at most two times per country, for $h = 1, 2$. In these cases, the length is set to zero. Therefore, the lengths for the AR(3)-GARCH models are slightly positively biased.

Table 9: Quantile Lengths for the univariate GARCH models and Pooled GARCH models.

	$h=1$			$h=2$			$h=3$		
Univ.	75%	95%	99%	75%	95%	99%	75%	95%	99%
Hist.	0.66	0.88	1.10	0.66	0.88	1.10	0.66	0.88	1.10
GARCH	0.68	0.79	0.90	0.67	1.45	2.97	0.67	11.92	67.31
GJR	0.68	0.80	0.91	0.67	1.53	3.21	0.67	13.16	73.79
AR-GARCH	0.61	0.64	0.67	0.64	0.74	0.86	0.66	0.88	1.24
AR-GJR	0.61	0.64	0.67	0.64	0.74	0.86	0.66	0.88	1.24
AR(3)-G	0.62	0.63	0.64	0.64	0.68	0.73	0.66	0.84	1.09
Pooled	75%	95%	99%	75%	95%	99%	75%	95%	99%
GARCH	0.68	0.76	0.84	0.66	0.77	0.88	0.66	0.83	1.13
GJR	0.68	0.76	0.84	0.66	0.77	0.88	0.66	0.83	1.13
AR-GARCH	0.60	0.63	0.66	0.60	0.66	0.72	0.63	0.74	0.87
AR-GJR	0.60	0.63	0.66	0.60	0.67	0.72	0.63	0.74	0.87
AR(3)-G	0.62	0.62	0.63	0.64	0.72	0.78	0.65	0.75	0.85

larger time horizon. Thus, from the QL, we conclude that the power of the pooled AR-GARCH models is high for the marginal GaR, but it is decreasing for larger horizons.

5.2.4 Joint GaR

We further perform joint GaR estimation by Pooled GARCH, of which Table 10 shows the results. For the 75% coverage level, the AR-mean models yield the best mean coverage level for 1-step-ahead predictions. The unconditional DQ test is not rejected for AR(1)-mean models, but the hit functions are still dependent, according to the DQ-Hits test. Only the AR(3)-GARCH model passes both DQ tests and shows strong evidence of correct conditional coverage from the DB test. For multistep-ahead predictions, we observe for the 75% coverage level that the empirical coverage level decreases sharply for all models, apart from the AR(3)-GARCH model. Consequently, all tests are rejected. However, the empirical levels return to a higher level for 3-step-ahead predictions. Although the tests do not show a correct conditional coverage, we still yield a better result as for the univariate GARCH models, where the empirical coverage level increased towards 90%. For the 95% and 99% coverage levels, we observe results similar to 75%. The AR(3)-GARCH model again performs the best for $h = 2$, while AR(1)-GARCH and AR(1)-GJR have the best performance for $h = 3$. Although the joint GaR is not specified correctly in the long term, we conclude that Pooled GARCH mitigates the estimation error for joint GaR that resulted from the estimations of univariate GARCH. In particular, the AR(3)-GARCH model again shows a relatively good and stable performance.

From the results of Table 10, we again observe a changing pattern between 1-step-ahead predictions and further horizons. Whereas the GARCH models with a constant mean return a too high empirical coverage for 1-step-ahead predictions, the autoregressive mean models return more accurate coverage. However, for $h = 2$, the constant mean models mitigate the decrease in empirical coverage, compared to the AR(1)-GARCH model. For $h = 3$, the constant mean GARCH and AR(1)-GARCH both return to a higher empirical coverage, where for the AR(1)-GARCH incorrect conditional coverage and the DQ-Hits test both cannot be rejected, in contrast to the constant mean GARCH. Therefore, the autoregressive mean performs mostly better than the constant mean, although the behaviour is ambiguous.

As expected, the historical GaR has a higher QL than all GARCH models for 1-step-ahead estimations. The autoregressive mean-based models have the lowest QL, which mean that they succeed to drop the GaR less to unnecessary low values, which is again a clear advantage. However, the results for the AR models are slightly biased as the GaR estimations sometimes exceed the 0.995 quantile. In the long-term, we observe that the autoregressive mean models always keep the QL below the value of the historical GaR, although it is at the cost of a lower empirical coverage. In contrast, the constant mean models exceed the QL of the historical GaR for 3-step-ahead. The AR(3)-GARCH model has the lowest QL for the coverage level 99% at horizon $h = 3$, which is not caused by a bias from an exceedance of the 99.5% quantile, again showing the power of this model for joint GaR prediction.

For the HGF values, the HGF value is consistently lower for the GARCH models in comparison to the historical GaR. Although we also observed this for the univariate GARCH models, the difference is that this result for 75% is consistent over time. In particular, the AR(3)-GARCH shows here robustness against a crash of multiple countries. For 2-step-ahead predictions, we even observe that its HGF is clearly lower in comparison to the other GARCH models, although its empirical coverage is significantly higher. In addition, its number of average violations is for the coverage level 99% relatively low, although slightly higher for 2-step-ahead estimations. Still, from these results, the AR(3)-GARCH model turns out to be the method that avoids multiple violations of the joint GaR.

To evaluate whether the performance is affected by the number of observations on which we base our predictions, we also evaluate the joint GaR for pooled GARCH using the subsample starting in October 2002. In Table 10, the empirical coverages of this subsample are

Table 10: Joint GaR results for univariate GARCH models using information pooling, evaluated at the mean coverage level (Cov.) and the p-value of the Dynamic Quantile test, using respectively a constant (Unc.), hit lags (Hits) and the first principal component of the GDP growth data (GDP). In addition, the p-value of the DB test is reported. The column '*2nd s.*' denotes the empirical coverage for the subsample from the 500th observation, October 2002.

Coverage				75%								99%			
$h = 1$	Cov.	*2nd s.*	Unc.	Hits	QL	DB	HGF	Cov.	*2nd s.*	Unc.	Hits	QL	DB	HGF
Hist.	84.23	83.18	0.00	0.00	1.01	0.00	3.15	97.96	96.73	0.02	0.00	1.49	0.10	5.73
GARCH	82.56	85.51	0.00	0.00	0.84	0.00	1.51	98.70	98.13	0.49	0.00	1.01	0.79	4.00
GJR	82.56	85.51	0.00	0.00	0.84	0.00	1.50	98.70	98.13	0.49	0.00	1.01	0.79	3.86
AR-GARCH	78.11	83.64	0.10	0.00	0.67	0.19	1.47	99.07	98.60	0.87	0.00	0.75	0.98	5.20
AR-GJR	77.92	83.18	0.12	0.00	0.67	0.23	1.45	98.89	98.13	0.79	0.00	0.75	0.96	4.50
AR(3)-G	74.40	77.10	0.75	0.34	0.64	0.72	1.88	97.96	96.26	0.02	0.00	0.68	0.10	3.36
$h = 2$	Cov.	*2nd s.*	Unc.	Hits	QL	DB	HGF	Cov.	*2nd s.*	Unc.	Hits	QL	DB	HGF
Hist.	83.83	83.10	0.01	0.00	1.01	0.00	3.15	97.96	96.71	0.29	1.00	1.49	0.10	6.09
GARCH	68.77	75.59	0.03	0.00	0.86	0.00	1.88	97.21	96.71	0.01	0.37	1.16	0.00	3.67
GJR	68.77	75.59	0.03	0.00	0.86	0.00	1.88	97.21	96.71	0.01	0.37	1.16	0.00	3.67
AR-GARCH	49.07	59.62	0.00	0.00	0.72	0.00	1.97	93.68	93.90	0.00	0.00	0.86	0.00	2.24
AR-GJR	48.88	59.15	0.00	0.00	0.72	0.00	1.97	93.49	93.90	0.00	0.00	0.87	0.00	2.20
AR(3)-G	76.95	82.16	0.40	0.18	0.80	0.45	1.63	98.33	97.18	0.39	0.00	0.97	0.35	5.78
$h = 3$	Cov.	*2nd s.*	Unc.	Hits	QL	DB	HGF	Cov.	*2nd s.*	Unc.	Hits	QL	DB	HGF
Hist.	83.61	83.49	0.01	0.00	1.01	0.00	3.14	97.95	96.70	0.29	0.00	1.49	0.10	6.09
GARCH	83.24	84.91	0.00	0.00	1.14	0.00	1.98	99.07	98.11	0.92	1.00	2.13	0.98	7.20
GJR	83.43	84.91	0.00	0.00	1.21	0.00	1.98	99.07	98.11	0.92	1.00	2.31	0.98	7.20
AR-GARCH	70.02	77.36	0.07	0.13	0.94	0.02	1.89	98.70	98.11	0.66	0.96	1.50	0.78	8.14
AR-GJR	70.02	77.36	0.07	0.10	0.90	0.02	1.89	98.70	98.11	0.66	0.96	1.44	0.78	8.14
AR(3)-G	68.90	74.53	0.03	0.02	0.90	0.00	1.78	98.14	98.11	0.20	0.61	1.26	0.19	6.40

shown next to the empirical coverage of the entire out-of-sample period. We observe that all estimation methods for $h = 1$ have a larger empirical coverage, which does not improve the results. However, for multistep-ahead predictions, the estimations for the GARCH models with conditional mean yield an empirical coverage close to 75%. In addition, the AR(1) models are still not close to 75%, but their coverage is higher. For $h = 3$, all models with autoregressive mean are relatively close to 75%. We conclude that our methods perform especially better for multistep-ahead forecasts if we have more data to base on our predictions.

5.2.5 Comparison with quarterly GDP data

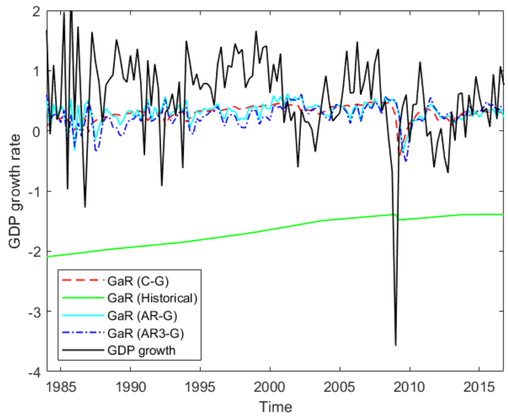
To put the results of Pooled GARCH for marginal GaR, shown in Table 7, into a broader perspective, we also perform the considered methods for quarterly data. We consider the data used in the research of Brownlees and Souza [2021], who also perform a Pooled GARCH model on GDP growth data. Table 11 shows the results. Where our models for the monthly GDP data show for 1-step-ahead an empirical coverage level that is slightly too high for 75%, we observe for the quarterly data that the empirical coverage level is underestimated. This feature of the predictions is also consistent over time, which is also in contrast with the GaR predictions of the AR(1)-GARCH model applied to our monthly data. In addition, the incorporation of a time-varying mean does not improve the GaR predictions significantly. It is even the case that the historical benchmark performs better in terms of the empirical coverage and the unconditional DQ test, while the DQ test on the Hits and the Tick Loss also do not show a clear winner. The GARCH models show better performance for the 95% coverage level, where the Tick Loss of the historical GaR is mostly outperformed. In addition, the GARCH models pass the unconditional DQ test and DB test in more cases. Although these results are slightly worse for the coverage level 99%, we conclude that the GARCH models perform in particular well for higher coverage levels at the quarterly data.

Table 11: Marginal GaR results of Pooled GARCH models for the quarterly GDP growth data.

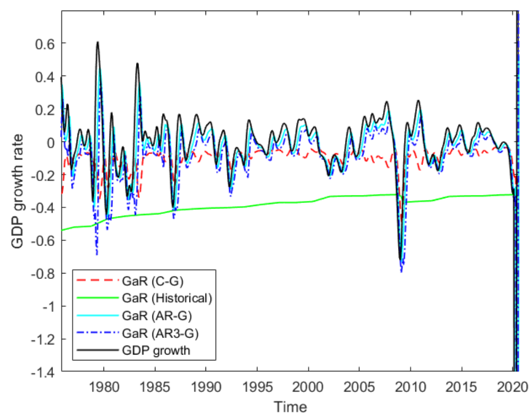
Coverage	75%						95%					
$h = 1$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
Hist.	69.73	45.83	16.67	37.50	33.6747	62.50	94.48	70.83	41.67	62.50	13.9709	70.83
GARCH	66.95	50.00	29.17	41.67	32.5605	54.17	93.84	83.33	70.83	87.50	12.1625	70.83
AR-GARCH	67.20	45.83	25.00	45.83	32.3507	50.00	93.34	79.17	62.50	87.50	12.3736	75.00
AR(3)-G	61.17	37.50	20.83	33.33	34.2049	37.50	91.07	45.83	41.67	58.33	13.2759	54.17
$h = 2$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
Hist.	69.66	70.83	16.67	45.83	33.7577	58.33	94.37	75.00	91.67	75.00	14.0799	79.17
GARCH	67.27	58.33	41.67	45.83	33.5022	58.33	93.29	79.17	91.67	91.67	13.3229	87.50
AR-GARCH	66.73	58.33	41.67	41.67	32.9670	62.50	93.19	75.00	79.17	87.50	13.1554	83.33
AR(3)-G	67.88	62.50	45.83	58.33	32.8522	62.50	93.54	79.17	87.50	87.50	12.9937	83.33
$h = 3$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
Hist.	69.84	75.00	33.33	54.17	33.7491	62.50	94.39	75.00	87.50	87.50	14.1844	75.00
GARCH	66.70	54.17	37.50	58.33	33.8305	50.00	95.80	95.83	91.67	95.83	14.4075	91.67
AR-GARCH	66.47	54.17	29.17	54.17	33.7653	50.00	96.06	95.83	95.83	100.00	14.2186	100.00
AR(3)-G	66.63	62.50	50.00	58.33	33.6596	50.00	95.83	95.83	87.50	95.83	14.0041	100.00

To gain more insight into why our GARCH models do not perform well for the 75% coverage level for the quarterly GDP growth rates, we plot the results for the Netherlands of 2-step-ahead predictions in Figure 3. In Figure 3a, we observe that the models are less able to follow the general pattern of the GDP growth rate, given that all GaR estimations stay in general between 0 and 0.5, apart from the banking crisis of 2008. Therefore, the GDP growth rates are violated relatively often, as the GDP growth rates are relatively volatile. Although this rigid pattern partly remains for the coverage level 95%, here we observe that the relative performance is slightly better. This is because the GARCH models cover the GDP growth rates well but still adapt better for the GDP growth rate pattern in comparison to the historical approach.

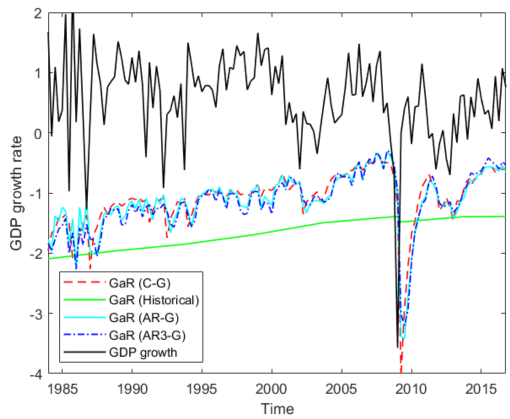
Figure 3: GDP growth at quarterly and monthly frequency for the Netherlands, compared with the 2-step-ahead GaR estimations based on the historical approach and a GARCH model using respectively a mean from a constant, an AR(1) model and an AR(3) model.



(a) Quarterly 2-step-ahead GaR (75%)



(b) Monthly 2-step-ahead GaR (75%)



(c) Quarterly 2-step-ahead GaR (95%)

The quarterly and monthly GDP also clearly behave differently in terms of the explainability of the mean. Figure 3 shows that the growth rates are closely followed by the GARCH models, in particular by the AR-GARCH models. This diminishes in particular the Tick Loss, while also the variance is relatively low, preventing the GaR estimations from being too low. This result follows from the data properties that we investigated in Section 3, where we obtained that the monthly GDP data is clearly better explained by an autoregressive model. These properties also explain that the constant mean GARCH model yields predictions that are relatively close to the estimations of the autoregressive mean: the mean of the quarterly growth rates is higher, but the coefficients of AR-GARCH are lower. Therefore, we conclude that the monthly GDP growth rates are better explained for the marginal GaR, mainly because the mean is better fitted by an autoregressive model, which yields tighter boundaries for the GaR estimations.

5.2.6 Clustered GARCH

Next, we also use clusters of countries to perform Pooled GARCH, using k-means clustering, where we use $k = 8$ clusters to split the countries. This number is chosen because it allows for both some bigger clusters as some countries that are estimated independently. The results are shown in Table 12. We observe that the incorporation of clusters slightly increases the empirical coverage, but this result does not significantly improve our estimations. If anything, the results suggest that the Clustered GARCH estimations are slightly better for longer estimation horizons in comparison to Pooled GARCH, based on the average Tick Loss values. This suggests that the exact parameter estimations per country are less important. We note that other choices of the number of clusters k yield approximately the same results. In addition, clustering the countries for which the sum of the parameters $\alpha + \beta$ are close to one, while estimating the other countries individually, does also not yield better results.

To investigate the performances at the level of the countries specifically, we also perform the Tick Loss backtests between the Pooled GARCH and Clustered GARCH. From Table 13, we observe that for most countries, we do not observe a significant difference in Tick Losses for predictions at horizon $h = 1, 2$. Here, Korea is the only country for which Clustered GARCH consistently outperforms Pooled GARCH. This country is first clustered on its own until 1994, and next clustered with other countries. Therefore, it seems that the single estimation in the earlier period paid off in terms of a more accurate estimation in comparison to Pooled GARCH. In addition, the Clustered GARCH slightly performs better for 3-step-ahead estimations for a substantial part of the countries. However, this still provides no convincing evidence that Clustered GARCH is more consistent than Pooled GARCH.

Table 12: Marginal results for univariate GARCH models estimated with Clustered GARCH (Cl. G), compared with univariate GARCH (u. G) and Pooled GARCH (P. G).

Coverage	75%						95%					
$h = 1$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
u. G	77.45	60.87	0.00	8.70	2.4353	73.91	96.17	47.83	8.70	52.17	1.4310	60.87
P. G	76.65	69.57	0.00	4.35	2.3661	78.26	95.77	34.78	8.70	56.52	1.4150	47.83
Cl. G	76.91	69.57	0.00	4.35	2.3899	69.57	95.80	39.13	13.04	34.78	1.4198	43.48
$h = 2$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
u. G	76.60	78.26	0.00	0.00	4.1519	73.91	96.60	47.83	73.91	73.91	2.3710	47.83
P. G	63.54	0.00	0.00	0.00	4.1870	0.00	88.48	8.70	13.04	8.70	2.3171	8.70
Cl. G	63.30	4.35	0.00	0.00	4.2091	0.00	88.60	4.35	8.70	4.35	2.3144	4.35
$h = 3$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
u. G	76.57	82.61	4.35	0.00	5.5214	73.91	97.43	34.78	65.22	52.17	3.2367	34.78
P. G	69.40	60.87	0.00	0.00	5.3885	39.13	92.05	60.87	73.91	30.43	2.9127	47.83
Cl. G	68.72	60.87	0.00	0.00	5.3955	34.78	91.77	56.52	65.22	34.78	2.8995	39.13

Table 13: Tick Loss backtests between Pooled GARCH and Clustered GARCH. The number that is displayed is the percentage of countries for which this method outperforms the other method.

Coverage	Better for 75%		Better for 95%		Better for 99%	
Model	P. G	Cl. G	P. G	Cl. G	P. G	Cl. G
$h = 1$	0.09	0.04	0.09	0.04	0.09	0.04
$h = 2$	0.09	0.09	0.09	0.04	0.09	0.09
$h = 3$	0.09	0.22	0.13	0.30	0.30	0.30

5.3 Markov Switching GARCH

5.3.1 Marginal GaR

Next, we predict the GaR with Markov Switching GARCH. The results for marginal GaR are shown in Table 14. At first sight, the most promising results for MS-GARCH are shown for the 95% and 99% coverage levels. Although the average coverage level for 1-step-ahead estimations is still less close to the real coverage, the Tick Loss is slightly decreased for MS-GARCH. For the 99% coverage level, the unconditional DQ test and Dynamic Binary test both are passed for more countries. Thus, MS-GARCH yields at this coverage for more countries a correct conditional coverage. For higher time horizons, MS-GARCH shows a clearly better estimated empirical coverage level. This is also expressed in the tests for conditional coverage, becoming more clear even for 3-step-ahead GaR forecasts. Comparing the results to Table 7, we however

Table 14: Marginal GaR estimation results for univariate MS-GARCH models (denoted as MS-G). For comparison, the univariate GARCH model estimated with QML (uG) is added.

	Horizon	$h = 1$		$h = 2$		$h = 3$	
Coverage		uG	MS-G	uG	MS-G	uG	MS-G
75%	Cov.	78.58	79.26	74.67	77.49	73.10	78.60
	Unc.	56.52	52.17	86.96	82.61	82.61	82.61
	Hits	0.00	0.00	0.00	0.00	0.00	0.00
	GDP	4.35	4.35	56.52	52.17	0.00	0.00
	QL	0.68	0.69	0.67	0.66	0.67	0.67
	TL	5.2379	5.4737	6.1342	6.1873	6.8647	6.5185
	DB	65.22	56.52	78.26	78.26	47.83	69.57
95%	Cov.	96.08	96.56	95.12	94.64	95.00	95.59
	Unc.	56.52	39.13	78.26	91.30	78.26	91.30
	Hits	0.00	0.00	47.83	34.78	60.87	69.57
	GDP	34.78	43.48	82.61	73.91	69.57	43.48
	QL	0.79	0.80	1.45	0.79	11.92	0.85
	TL	2.3683	2.3354	6.1116	3.0728	58.6895	3.4371
	DB	65.22	47.83	78.26	73.91	69.57	78.26
99%	Cov.	99.00	99.19	98.88	98.45	99.02	98.68
	Unc.	86.96	95.65	95.65	82.61	82.61	95.65
	Hits	52.17	39.13	100.00	82.61	60.87	82.61
	GDP	78.26	69.57	91.30	73.91	86.96	82.61
	QL	0.90	0.92	2.97	0.95	67.31	1.18
	TL	1.4150	1.3793	3.7611	1.9111	68.0349	2.1273
	DB	86.96	100.00	86.96	78.26	78.26	86.96

observe that the univariate MS-GARCH models never outperform the pooled GARCH models in terms of the TL. We conclude that the incorporation of regimes has some added value for the 95% and 99% coverage levels, but it is still outperformed by the pooled GARCH models.

The reason that MS-GARCH for the higher coverage levels beats univariate GARCH, is that MS-GARCH returns consistent estimations. On the contrary, the average coverage of the univariate GARCH models is decreasing with a larger horizon. In addition, we find for MS-GARCH that the average coverage level is not decreasing with the time horizon. This is a significant advantage in comparison to the autoregressive mean models. However, a big disadvantage is that this model has a very long computation time in comparison to the conventional GARCH model. Therefore, the MS-GARCH model is also not very useful in terms of efficiency.

For the coverage level 75%, we observe that the incorporation of two regimes does not enhance our estimations, consistently over all estimation horizons. Although the average coverage level fluctuates less over the horizons in comparison to the GARCH model estimated with QML, the coverage level is not significantly better for MS-GARCH. In addition, the Tick Loss is even lower for the MS-GARCH model and the tests for conditional coverage are also passed for fewer countries. Therefore, the incorporation of a regime-switching framework does not have added value for the GaR estimation at the 75% coverage level.

In addition, we also evaluate the TL country-by-country. Table 15 shows the results. Here is also shown that MS-GARCH only becomes competitive to the univariate GARCH models for the higher coverage levels and multistep-ahead forecasts, mainly due to the unstable GaR predictions of the univariate GARCH models. Therefore, we conclude that the MS-GARCH model does not add enough value for the marginal GaR prediction. Although the GaR estimations have a more stable pattern than the univariate GARCH predictions for some countries, we obtained better results by the information pooling we considered in Subsection 5.2.

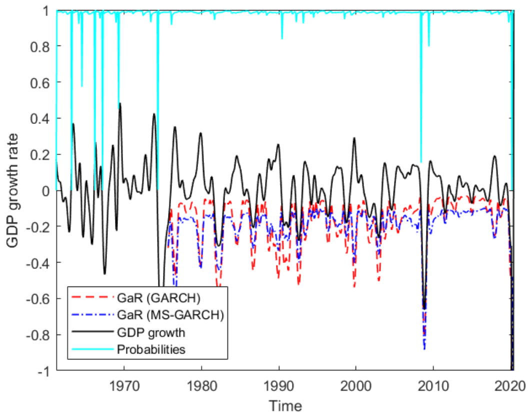
Table 15: Tick Loss backtests between univariate GARCH and Markov Switching GARCH. The number that is displayed is the percentage of countries for which this method outperforms the other method.

Coverage	Better for 75%		Better for 95%		Better for 99%	
Model	uG	MS-G	uG	MS-G	uG	MS-G
$h = 1$	0.61	0.17	0.74	0.13	0.65	0.09
$h = 2$	0.48	0.22	0.35	0.39	0.09	0.48
$h = 3$	0.70	0.22	0.26	0.39	0.04	0.04

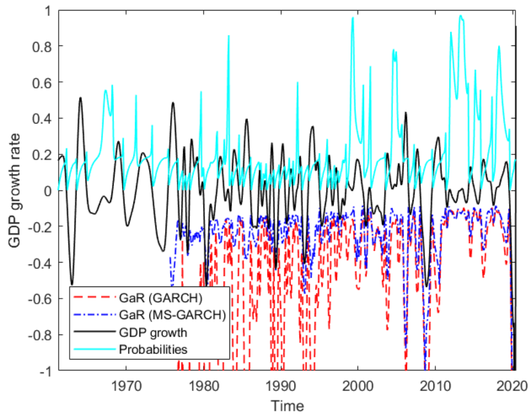
From the conditional probabilities for MS-GARCH based on the entire sample size, we observe that most countries have probabilities that are mainly close to 0 or 1. Therefore, the MS-GARCH model does not seem to estimate the GARCH parameters more accurately for these countries, although the fact that the MS-GARCH model uses the constraints of Haas and Paolella [2012] could stabilize the estimations. We take a closer look at the countries' time series that show a regime-switching pattern to investigate if the regime-switching framework does indeed explain the results better.

Figure 4 shows the countries with the most informative probabilities. In general, we observe that the countries differ in the persistence of the state probabilities. Figure 4d shows for both regimes that they persist for longer periods, where the regimes correspond to high-volatility and low-volatility states. In contrast, the other figures show a pattern where the second regime only persists for a few observations. The results of the countries both distinguish recession and expansion regimes, and high-volatility and low-volatility regimes. Figure 4a mainly filters out some strong negative rates. Figure 4b shows that the second regime occurs in the second part of the data. Although it is not at first glance seen from the figure, this second state represents a regime where the weight of the past volatility is higher in the modelling of the new volatility estimation. This causes slightly larger and more stable volatility estimations. In Figure 4c, we observe that the second state is mainly assigned to recession observations. Thus, we conclude from these figures that MS-GARCH can show some insightful GDP growth rates' features. For most countries, the Markov Switching GARCH framework does not show more extensive volatility dynamics, mainly because there seems to be too little variation in the data. Probably, this is caused by the fact that the data set is still relatively short.

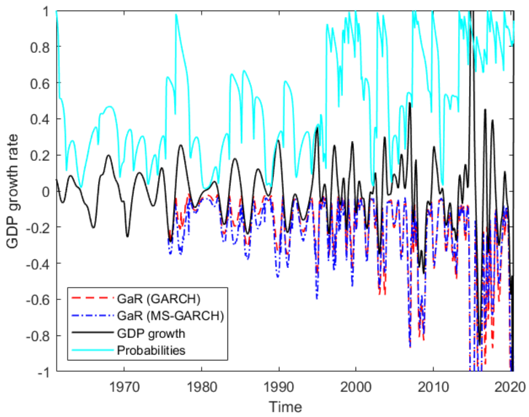
Figure 4: Predicted probabilities of MS-GARCH models, fitted on four countries' GDP growth rates, compared with the 2-step-ahead 95% GaR estimations with univariate GARCH and MS-GARCH.



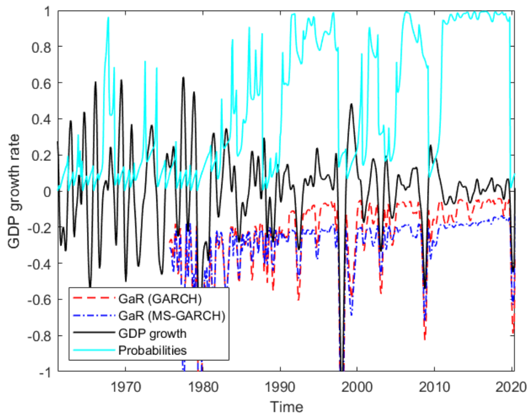
(a) Switzerland



(b) Denmark



(c) Ireland



(d) South Korea

5.3.2 Joint GaR

We evaluate the results of MS-GARCH for joint GaR in Table 16. In general, the pattern in the results is similar to the marginal GaR. The empirical coverage is structurally overestimated by MS-GARCH for 75%, which again emphasizes that MS-GARCH is more suitable for high coverage levels. Consequently, the results for the tests at this coverage level are in general not passed. However, in contrast to the marginal GaR, MS-GARCH also has a too high empirical coverage level for 95%. These estimations are not decreasing over the horizons, like the standard GARCH models, but rather increasing. However, this still does not yield promising results in terms of the performed tests. The HGF is also lower for the univariate GARCH models in comparison, which is also an advantage of univariate GARCH models.

Table 16: Joint GaR estimation results for univariate MS-GARCH models (denoted as MS-G). For comparison, the univariate GARCH model estimated with QML (uG) is added.

	Horizon	$h = 1$		$h = 2$		$h = 3$	
Coverage		Q-G	MS-G	Q-G	MS-G	Q-G	MS-G
75%	Cov.	81.45	87.94	87.73	80.30	93.67	85.29
	Unc.	0.00	0.00	0.00	0.05	0.00	0.00
	Hits	0.00	0.00	0.00	0.00	0.00	0.00
	QL	0.86	9.78	2.07	16.39	32.08	1.3e+07
	DB	0.00	0.00	0.00	0.01	0.00	0.00
	HGF	1.42	1.63	1.88	2.14	2.56	2.33
95%	Cov.	96.10	97.77	98.14	97.96	99.26	98.51
	Unc.	0.24	0.00	0.00	0.01	0.00	0.00
	Hits	0.06	0.00	0.00	0.00	1.00	0.00
	QL	0.99	14.00	3.25	27.76	66.70	2.8e+07
	DB	0.46	0.00	0.00	0.00	0.00	0.00
	HGF	2.10	2.58	4.50	5.18	8.00	5.63
99%	Cov.	98.33	99.07	98.88	99.07	99.44	99.26
	Unc.	0.12	0.87	0.86	0.92	0.46	0.71
	Hits	0.00	0.00	1.00	1.00	0.91	1.00
	QL	1.15	23.09	4.34	49.00	106.22	5.6e+07
	DB	0.35	0.98	0.95	0.98	0.53	0.82
	HGF	3.33	4.60	5.17	8.40	8.67	8.50

In terms of the empirical coverage level, MS-GARCH shows better results for the coverage level 99%. In addition, the tests for correct conditional coverage are passed for a majority of the countries. This result is even stronger visible for the DB test, although it has theoretically higher power than the DQ test. The HGF is also slightly lower, although the differences are relatively small. However, the QL explodes for all coverage levels, probably since there are always estimations close to instability. Thus, the MS-GARCH model allows for even more extreme estimations, making it an unreliable method for joint GaR estimation.

5.4 Multivariate GARCH

Next, we discuss the results of Multivariate GARCH. As we use the assumption of a multivariate normal distribution, we compare the results of our multivariate GARCH models with univariate GARCH models that assume a normal distribution of the errors. Table 17 shows the results. From the results for the coverage level 75%, we observe that the autoregressive models yield a clearly higher empirical coverage, which leads to a consistent overestimation of the empirical coverage. However, considering the constant mean models, we observe that the multivariate GARCH model mitigates the increasing empirical coverage. In addition, the AR-Multivariate GARCH model is the only model that does not show a strongly increasing empirical coverage. However, we still conclude that all models do not return reliable GaR estimations.

Table 17: Results of the marginal GaR predictions estimated by Multivariate GARCH, based on a constant mean (MG) and an autoregressive mean (AR-MG). The models are compared with univariate GARCH models for which we assume normal distributed errors, using a constant mean (G-N) and an autoregressive mean (AR-G-N).

Coverage	75%						99%					
$h = 1$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
G-N	63.88	8.70	0.00	0.00	5.4380	13.04	99.40	95.65	65.22	95.65	1.3451	100.00
AR-G-N	90.12	0.00	0.00	0.00	3.6105	0.00	99.35	100.00	91.30	100.00	1.2547	100.00
MG	70.36	34.78	0.00	4.35	6.1094	34.78	97.12	26.09	0.00	30.43	1.8052	30.43
AR-MG	94.07	0.00	0.00	0.00	3.7766	0.00	99.65	100.00	82.61	100.00	1.2862	100.00
$h = 2$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
G-N	93.07	0.00	0.00	0.00	7.1467	0.00	99.66	91.30	56.52	100.00	2.0434	91.30
AR-G-N	98.09	0.00	0.00	0.00	7.4516	0.00	99.73	78.26	60.87	100.00	1.9866	78.26
MG	80.89	65.22	0.00	43.48	6.6293	21.74	99.33	100.00	73.91	91.30	2.0644	100.00
AR-MG	91.98	0.00	0.00	0.00	5.7690	0.00	99.68	82.61	65.22	100.00	1.8562	82.61
$h = 3$	Cov.	Unc.	Hits	GDP	TL	DB	Cov.	Unc.	Hits	GDP	TL	DB
G-N	96.09	0.00	0.00	0.00	9.3889	0.00	99.68	73.91	0.00	100.00	2.8787	73.91
AR-G-N	99.43	0.00	0.00	0.00	14.8551	0.00	99.82	56.52	0.00	100.00	3.4431	56.52
MG	89.38	13.04	0.00	0.00	8.0854	4.35	99.68	82.61	13.04	100.00	2.6328	82.61
AR-MG	94.82	0.00	0.00	0.00	7.6797	0.00	99.75	65.22	0.00	100.00	2.5755	65.22

The results for the coverage level 99% are at first sight more promising, as most specification tests are passed. In most cases, this leads to a correct conditional coverage for both the last lag (shown from the DB test) as the last four lags (shown from the DQ-Hits test). The TL values of the autoregressive models are even below the univariate GARCH models from Subsection 5.1. However, this result is obtained from the autoregressive mean and not from the multivariate GARCH framework. Also, the estimation error from the fact that all these methods overestimate the empirical coverage is naturally mitigated for the coverage level 99%.

Concluding, the GaR predictions from multivariate GARCH are not well specified, which has two reasons. From the estimations of the univariate GARCH models with the assumption of the normal distribution, we deduce that this assumption is not valid. Second, recalling the VAR-GARCH results, it is also likely that our model suffers from the fact that it has too many parameters. In addition, the estimation time of multivariate GARCH is also significantly larger than for univariate GARCH models. Given these results for the marginal GaR, we decide to leave the joint GaR of multivariate GARCH out of our analysis.

6 Conclusion

In this thesis, we investigated several dynamic models to model the Growth-at-Risk of OECD countries. We both evaluated marginal and joint GaR and applied several models to get insight into the behaviour of the GDP growth rates. We included the allowance of a time-varying mean, an asymmetry term, creating clusters of countries, two volatility regimes and a multivariate framework. Starting with the univariate GARCH models, we concluded that they all outperform a simple historical approach for 1-step-ahead forecasts. In addition, GARCH models are also able to estimate a correct empirical coverage for multistep-ahead GaR forecasts. Constant mean GARCH models also produce reliable 1-step-ahead GaR estimations that follow the pattern of the growth rates quite well and are not too low. In addition, the volatility models increase the independence of the GaR violations.

However, we find in line with Pakel et al. [2011] that univariate GARCH models have for the considered data a high instability risk, which makes estimating GARCH models a challenging, dangerous task. This data feature leads to unnecessary low GaR estimations for multistep-ahead predictions, especially for models that use a constant mean. This issue appears due to the low data frequency and will therefore also exist in the future. However, this issue is solved by a simple information pooling approach, yielding consistent stabilization of the estimations. This approach also saves a lot of estimation time. We investigated the use of information pooling further by the clustering of countries. For our considered methods, we found no evidence that the use of multiple clusters outperforms the approach of assigning the same GARCH parameters to all countries. Still, an interesting research direction would be to search for other methods that use information pooling for the estimation of volatility models.

Considering an autoregressive mean, we conclude that this yields significantly better results for our GaR estimations. This finding applies both for the univariate and pooled GARCH methods. Although one lag already contains much information, including the second and third lag mostly enhances the marginal GaR estimations, especially for longer horizons. In addition, the GaR estimations are also not unnecessary low, in comparison to the univariate GARCH models. This finding is shown especially for longer time horizons. Thus, although the volatility is a logical choice to focus on in the GaR estimation, the estimation of the mean growth rate may therefore be just as important. Further research could use some explanatory variables to estimate the conditional mean more precisely, but a simple autoregressive model could still appear to be a competitive benchmark.

In our investigation of joint GaR, it turned out that the main challenge is still to obtain consistent estimations for multistep-ahead predictions. Where univariate GARCH models tend to overestimate the empirical coverage, pooled GARCH suffers from underestimation. Generally, the best estimation results are obtained by pooled GARCH models that use multiple autoregressive lags for the mean estimation. In particular, this effect increases when we can rely on much historical data, which is a feature to keep an eye on in the future. In addition, we show by the HGF measure that the GaR estimations are also more independently violated by pooled GARCH models. This extends the joint GaR research of Brownlees and Souza [2021].

We also investigated the incorporation of regime-switching models. Although the Markov Switching GARCH model showed for a few countries some useful insights, the allowance of regimes overall did not add much value to the GaR estimations, which contradicts the conclusion of Chauvet and Hamilton [2006]. The improvements for marginal GaR at high coverage levels were still smaller than the improvements from pooled GARCH. This result is caused by the fact that the data contains too little variation to clearly show a regime switch, probably because the data set still consists of relatively few observations. For further research, it might be useful to focus on testing whether the data contains a structural break, which can be allowed for both the mean and volatility.

Our methods to incorporate cross-sectional information in the GaR prediction did not work well. To investigate the joint distribution of GDP growth rates more in detail, we could combine the GARCH models with copulas and analyze pairs of countries. In addition, one could focus on a restricted version of the covariance matrix and analyze whether that specifies the cross-sectional interactions better to avoid over-specification of parameters.

Although we used monthly data, which is data at a higher frequency than in most research, the data set is still relatively small. Given that this will stay a challenge in the Growth-at-Risk research, this could be solved by using high-frequency explanatory variables, like the GARCH-MIDAS approach of Engle et al. [2013]. However, we observed in the comparison between the monthly and quarterly growth rates that the monthly growth rates are better estimated by an autoregressive model, which makes them more attractive to use. Therefore, using monthly GDP growth rates more often in literature seems to have some appealing advantages.

References

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

D. W. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, 59(3):817–858, 1991.

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

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

A. Caldera Sánchez and O. Röhn. How do policies influence GDP tail risks? Technical Report 1339, Organisation for Economic Co-operation and Development, Paris, 2016.

Y. Chang, Y. Choi, and J. Y. Park. A new approach to model regime switching. *Journal of Econometrics*, 196(1):127–143, 2017.

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

T. Chuffart. An implementation of Markov Regime Switching GARCH models in MATLAB. *Available at SSRN 2892688*, 2017.

E. I. Dumitrescu, C. Hurlin, and V. Pham. Backtesting Value-at-Risk: from Dynamic Quantile to Dynamic Binary tests. *Finance*, 33(1):79–112, 2012.

R. F. Engle and K. F. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11(1):122–150, 1995.

R. F. Engle and S. Manganelli. CAViaR: Conditional autoregressive Value-at-Risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381, 2004.

R. F. Engle, N. Shephard, and K. Sheppard. Fitting vast dimensional time-varying covariance models. *Working Papers 2008OMI10*, 2008.

R. F. Engle, E. Ghysels, and B. Sohn. Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*, 95(3):776–797, 2013.

R. F. Engle, O. Ledoit, and M. Wolf. Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375, 2019.

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

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

T. Gneiting and R. Ranjan. Comparing density forecasts using threshold-and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*, 29(3):411–422, 2011.

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

M. Haas and M. S. Paolella. *Mixture and Regime-Switching GARCH Models*, chapter 3, pages 71–102. John Wiley & Sons, Ltd, 2012. ISBN 9781118272039.

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

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

C. Hurlin and S. Tokpavi. Une évaluation des procédures de backtesting. *Finance*, 29(1):53–80, 2008.

S. Hwang and P. L. Valls Pereira. Small sample properties of garch estimates and persistence. *European Journal of Finance*, 12(6-7):473–494, 2006.

IMF. Financial Conditions and Growth at Risk. *Global Financial Stability Report*, 2017.

H. Kauppi and P. Saikkonen. Predicting us recessions with dynamic binary response models. *Review of Economics and Statistics*, 90(4):777–791, 2008.

F. Klaassen. Improving GARCH volatility forecasts with regime-switching GARCH. In *Advances in Markov-switching models*, pages 223–254. Springer, 2002.

N. Nolde and J. F. Ziegel. Elicitability and backtesting: Perspectives for banking regulation. *The Annals of Applied Statistics*, 11(4):1833–1874, 2017.

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

C. Pakel, N. Shephard, K. Sheppard, and R. F. Engle. Fitting vast dimensional time-varying covariance models. *Journal of Business & Economic Statistics*, pages 1–17, 2020.

M. Plagborg-Møller, L. Reichlin, G. Ricco, and T. Hasenzagl. When is Growth at Risk? *Brookings Papers on Economic Activity*, Forthcoming, 2020.

A. Prasad, S. Elekdag, P. Jeasakul, R. Lafarguette, A. Alter, A. X. Feng, and C. Wang. *Growth at Risk: Concept and application in IMF country surveillance.* International Monetary Fund, 2019.

P. Smith and P. M. Summers. Regime switches in GDP growth and volatility: Some international evidence and implications for modeling business cycles. *BE Journal of Macroeconomics*, 9(1), 2009.

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

A Pseudocode for GaR estimation

Run the GARCH model based on the data available at time T using a certain definition of $\mu_{i,t+1|t}$ and obtain:

- Conditional means $\mu_{i,t}$, conditional variances $\sigma^2_{i,t}$ and standardized residuals $\hat{z}_{i,t} = (y_{i,t} - \mu_{i,t})/\sigma_{i,t}$, for $i = 1, ..., N$ and $t = 1, ..., T$;
- The 1-step-ahead forecasted variances $\sigma^2_{i,T+1|T}$;
- The parameters $\boldsymbol{\theta}_\mu$ of the conditional mean, depending on the definition of $\mu_{i,t+1|t}$;
- The parameters $\boldsymbol{\theta}_\sigma$ of the conditional variance, depending on the definition of $\sigma_{i,t+1|t}$;

for $j = 1, ..., h$ day-ahead paths **do**

  Draw S times an index from the values $1, ..., T$ and bootstrap from the standardized residuals the values $z^{(1)}_{i,T}, ..., z^{(S)}_{i,T}$;

  if $j = 1$ **then**

    Construct $\mu_{i,T+1|T}$, with $y_{i,T}$ as input;

    $\tilde{y}^{(s)}_{i,T+1|T} = \mu_{i,T+1|T} + \sigma_{i,T+1|T} z^{(s)}_{i,t}$, for $s = 1, ..., S$;

  else

    Estimate $\mu_{i,T+j|T}$ with the mean equation, using $\tilde{y}_{i,T+j-1|T}$ and $\boldsymbol{\theta}_\mu$ as input;

    Estimate $\sigma^{2(s)}_{i,T+j|T}$ with the variance equation, using $\tilde{y}_{i,T+j-1|T}$ and $\boldsymbol{\theta}_\sigma$ as input;

    $\tilde{y}^{(s)}_{i,T+j|T} = \mu_{i,T+j|T} + \sigma^{(s)}_{i,T+j|T} z^{(s)}_{i,t}$, for $s = 1, ..., S$;

  end if

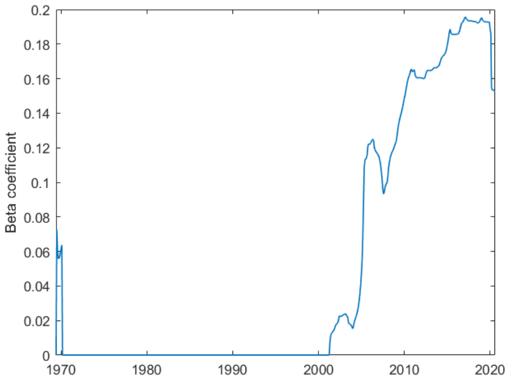
end for

Construct $\text{GaR}^M_{i,t+h|t}(p)$ for country i as $Q_p(\tilde{y}^{(s)}_{i,T+j|T})$.

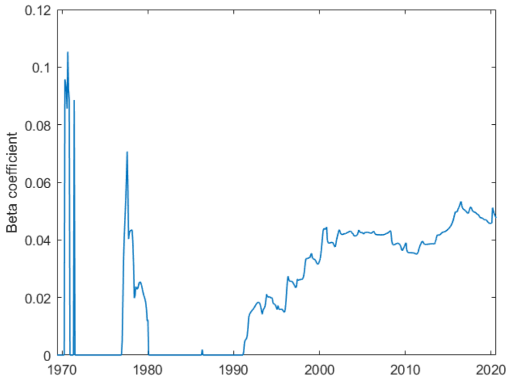
Algorithm 2: General pseudocode for the h-step ahead GaR for country i at time T

B Additional data tables and figures

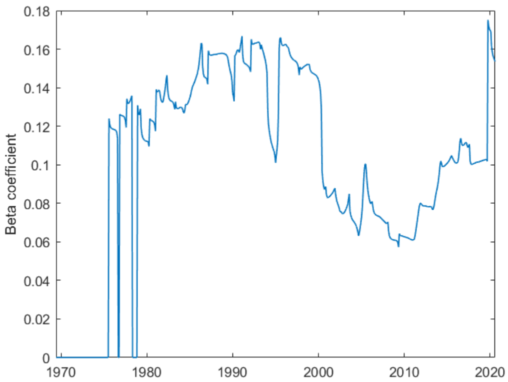
Figure 5: Beta coefficients in the GARCH models for five countries of the data sets, based on the data up to the corresponding month in the figure.



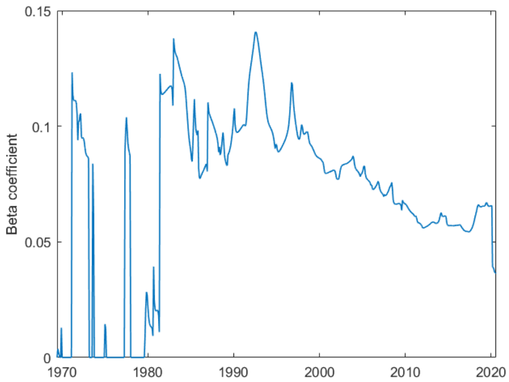
(a) Spain



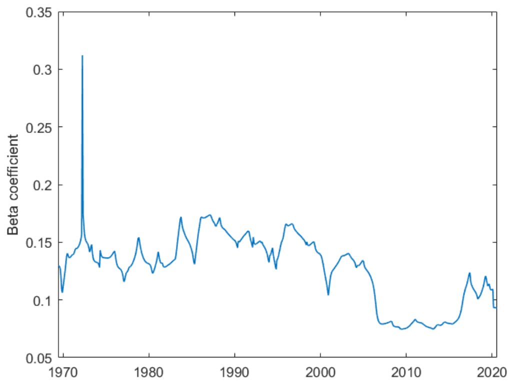
(b) Finland



(c) Great Britain



(d) Ireland



(e) Italy

Figure 6: Mean and variance of the growth rate at each time across all considered countries, where the recession period corresponds with the shaded area. The variance is plotted at a logarithmic scale.

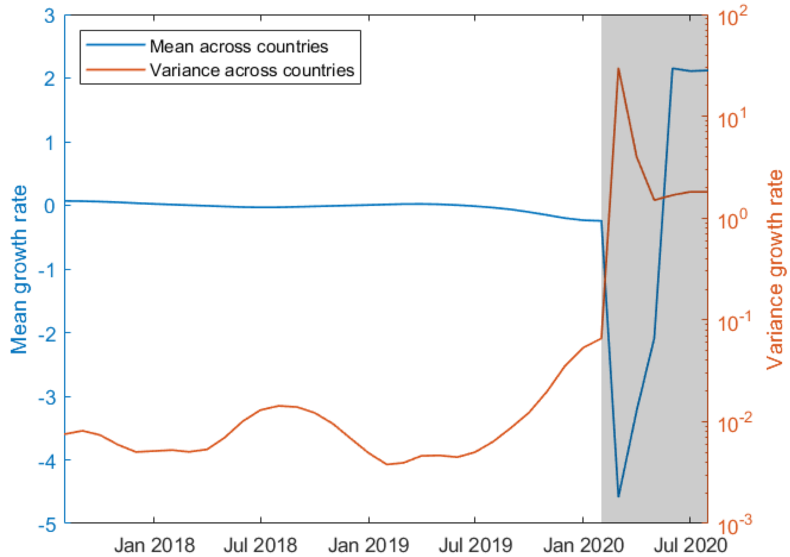
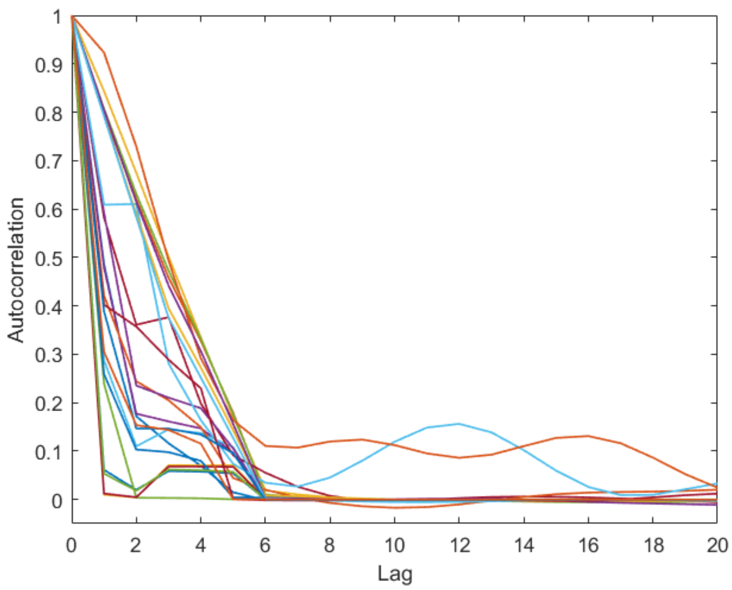


Figure 7: Autocorrelation functions of all 23 countries for the squared GDP growth rates, based on the whole dataset.



	AUS	AUT	BEL	CAN	CHE	DEU	DNK	ESP	FIN	FRA	GBR	GRC	IRL	ITA	JPN	KOR	NLD	NOR	PRT	SWE	TUR	USA	ZAF
AUS	1.00																						
AUT	0.60	1.00																					
BEL	0.66	0.91	1.00																				
CAN	0.70	0.82	0.82	1.00																			
CHE	0.61	0.83	0.83	0.75	1.00																		
DEU	0.55	0.83	0.80	0.85	0.74	1.00																	
DNK	0.42	0.64	0.67	0.53	0.54	0.55	1.00																
ESP	0.52	0.72	0.71	0.84	0.63	0.85	0.36	1.00															
FIN	0.41	0.53	0.52	0.43	0.57	0.45	0.44	0.26	1.00														
FRA	0.44	0.66	0.65	0.77	0.59	0.82	0.31	0.96	0.20	1.00													
GBR	0.69	0.87	0.85	0.89	0.72	0.86	0.53	0.84	0.43	0.76	1.00												
GRC	0.58	0.51	0.53	0.48	0.43	0.40	0.44	0.24	0.41	0.08	0.58	1.00											
IRL	0.31	0.34	0.40	0.31	0.36	0.30	0.26	0.27	0.31	0.22	0.36	0.25	1.00										
ITA	0.46	0.67	0.67	0.78	0.62	0.82	0.32	0.95	0.25	0.97	0.76	0.11	0.22	1.00									
JPN	0.58	0.66	0.66	0.59	0.57	0.55	0.60	0.33	0.54	0.22	0.65	0.70	0.43	0.25	1.00								
KOR	0.08	0.19	0.18	0.17	0.14	0.19	0.24	0.11	0.07	0.08	0.16	0.16	0.12	0.09	0.20	1.00							
NLD	0.59	0.83	0.83	0.75	0.78	0.81	0.61	0.63	0.45	0.59	0.77	0.43	0.32	0.60	0.59	0.11	1.00						
NOR	0.44	0.61	0.68	0.61	0.60	0.58	0.49	0.51	0.31	0.47	0.59	0.39	0.29	0.48	0.40	0.04	0.60	1.00					
PRT	0.51	0.72	0.70	0.81	0.65	0.83	0.39	0.94	0.25	0.92	0.82	0.24	0.25	0.93	0.34	0.12	0.63	0.49	1.00				
SWE	0.56	0.78	0.77	0.68	0.69	0.72	0.55	0.59	0.59	0.53	0.73	0.42	0.43	0.55	0.60	0.21	0.71	0.53	0.57	1.00			
TUR	0.44	0.62	0.61	0.52	0.53	0.45	0.39	0.32	0.40	0.23	0.56	0.48	0.40	0.26	0.66	0.18	0.52	0.44	0.28	0.53	1.00		
USA	0.63	0.77	0.77	0.93	0.70	0.84	0.56	0.83	0.35	0.77	0.88	0.46	0.34	0.77	0.57	0.20	0.72	0.58	0.80	0.63	0.50	1.00	
ZAF	0.70	0.74	0.78	0.66	0.66	0.55	0.50	0.39	0.51	0.26	0.70	0.72	0.36	0.29	0.76	0.14	0.69	0.53	0.36	0.65	0.67	0.57	1.00

Table 18: Cross-correlations between all countries, calculated over the whole data set

C Results VAR-GARCH

Table 19: Results for the marginal GaR estimations using VAR-GARCH.

	$h = 1$			$h = 2$			$h = 3$		
Coverage	75%	95%	99%	75%	95%	99%	75%	95%	99%
Cov.	67.24	78.88	85.04	64.50	80.10	87.61	68.76	85.53	92.96
Unc.	17.39	0.00	0.00	21.74	13.04	8.70	52.17	21.74	21.74
Hits	0.00	0.00	0.00	0.00	0.00	26.09	0.00	13.04	30.43
GDP	0.00	0.00	0.00	17.39	13.04	17.39	0.00	39.13	39.13
TL	5.2935	3.6621	2.8835	7.8678	4.3536	2.8479	9.9201	5.3705	3.1287
DB	21.74	0.00	0.00	13.04	4.35	4.35	17.39	8.70	17.39

Table 20: Results for the joint GaR estimations using VAR-GARCH. The results of the Quantile Length are biased, because the 0.995 quantile is frequently exceeded by the joint GaR estimation.

	$h = 1$			$h = 2$			$h = 3$		
Coverage	75%	95%	99%	75%	95%	99%	75%	95%	99%
Cov.	5.57	12.43	18.37	8.55	19.89	38.66	21.60	53.82	77.47
Unc.	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Hits	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
QL	0.68	0.71	0.74	0.82	0.93	1.06	1.07	1.37	1.76
DB	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HGF	3.93	3.17	2.72	3.44	2.43	2.00	2.38	1.88	1.66

D Additional results for joint GaR

Table 21: Joint GaR estimations from univariate GARCH models for quarterly GDP growth rates, excluding the GDP time series for Turkey. (99% is shown at the next page.)

Coverage	75%						95%					
$h = 1$	Cov.	Unc.	Hits	QL	DB	HGF	Cov.	Unc.	Hits	QL	DB	HGF
Hist.	87.01	0.00	0.00	1.01	0.00	3.53	95.73	0.44	0.00	1.27	0.70	4.35
GARCH	80.71	0.00	0.00	0.82	0.01	1.44	94.62	0.69	0.19	0.91	0.87	1.76
GJR	80.89	0.00	0.00	0.83	0.00	1.43	94.62	0.69	0.12	0.92	0.87	1.72
AR-GARCH	82.56	0.00	0.00	0.67	0.00	1.41	95.18	0.85	0.43	0.70	0.94	1.81
AR-GJR	83.30	0.00	0.00	0.67	0.00	1.47	95.18	0.85	0.43	0.70	0.94	1.77
AR(3)-G	76.99	0.28	0.05	0.65	0.40	1.69	94.81	0.84	0.49	0.67	0.93	1.89
$h = 2$	Cov.	Unc.	Hits	QL	DB	HGF	Cov.	Unc.	Hits	QL	DB	HGF
Hist.	86.62	0.00	0.00	1.01	0.00	3.50	95.72	0.69	0.00	1.27	0.71	4.61
GARCH	86.99	0.00	0.00	1.16	0.00	1.77	97.96	0.00	0.00	1.66	0.00	4.18
GJR	86.99	0.00	0.00	1.27	0.00	1.76	97.96	0.00	0.00	1.89	0.00	4.00
AR-GARCH	92.01	0.00	0.00	0.94	0.00	2.00	98.70	0.00	1.00	1.17	0.00	6.14
AR-GJR	92.19	0.00	0.00	0.96	0.00	2.02	98.51	0.00	0.00	1.20	0.00	5.50
AR(3)-G	88.66	0.00	0.00	0.95	0.00	1.92	98.70	0.00	0.00	1.23	0.00	7.00
$h = 3$	Cov.	Unc.	Hits	QL	DB	HGF	Cov.	Unc.	Hits	QL	DB	HGF
Hist.	86.41	0.00	0.00	1.01	0.00	3.48	95.72	0.69	0.00	1.27	0.71	4.87
GARCH	93.48	0.00	0.00	3.15	0.00	2.54	99.26	0.00	1.00	7.25	0.00	7.50
GJR	94.04	0.00	0.00	5.33	0.00	2.69	99.26	0.00	1.00	13.34	0.00	7.50
AR-GARCH	98.14	0.00	0.00	2.67	0.00	5.50	99.44	0.00	0.00	5.19	0.00	13.67
AR-GJR	97.95	0.00	0.00	3.29	0.00	5.09	99.44	0.00	0.00	6.63	0.00	13.67
AR(3)-G	95.90	0.00	0.00	6.66	0.00	3.36	99.26	0.00	0.00	12.00	0.00	11.75

Coverage	99%					
$h = 1$	Cov.	Unc.	Hits	QL	DB	HGF
Hist.	98.14	0.05	0.00	1.44	0.20	6.30
GARCH	97.59	0.00	0.00	1.00	0.02	2.54
GJR	97.77	0.00	0.00	1.01	0.05	2.58
AR-GARCH	98.14	0.05	0.01	0.73	0.20	2.90
AR-GJR	98.14	0.05	0.01	0.73	0.20	2.90
AR(3)-G	97.77	0.00	0.00	0.68	0.05	2.75
$h = 2$	Cov.	Unc.	Hits	QL	DB	HGF
Hist.	97.96	0.35	0.00	1.44	0.10	6.09
GARCH	98.88	0.86	1.00	2.28	0.95	4.83
GJR	98.88	0.86	1.00	2.63	0.95	4.83
AR-GARCH	99.44	0.46	1.00	1.41	0.53	13.00
AR-GJR	99.44	0.46	1.00	1.45	0.53	13.00
AR(3)-G	98.88	0.87	1.00	1.50	0.95	7.83
$h = 3$	Cov.	Unc.	Hits	QL	DB	HGF
Hist.	97.95	0.34	0.00	1.44	0.10	6.09
GARCH	99.44	0.46	0.91	15.83	0.53	8.00
GJR	99.44	0.46	0.91	29.66	0.53	8.00
AR-GARCH	99.44	0.46	0.91	8.44	0.53	13.00
AR-GJR	99.44	0.46	0.91	10.01	0.53	12.67
AR(3)-G	99.44	0.46	0.91	18.23	0.53	13.67