VU Research Portal

Observation-driven models for realized variances and overnight returns applied to Value-at-Risk and Expected Shortfall forecasting

Opschoor, Anne; Lucas, André

published in

International Journal of Forecasting
2021

DOI (link to publisher)

[10.1016/j.ijforecast.2020.07.009](https://doi.org/10.1016/j.ijforecast.2020.07.009)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Opschoor, A., & Lucas, A. (2021). Observation-driven models for realized variances and overnight returns applied to Value-at-Risk and Expected Shortfall forecasting. *International Journal of Forecasting*, 37(2), 622-633. <https://doi.org/10.1016/j.ijforecast.2020.07.009>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Observation-driven models for realized variances and overnight returns applied to Value-at-Risk and Expected Shortfall forecasting[☆]

Anne Opschoor^{*}, André Lucas

Vrije Universiteit Amsterdam, The Netherlands
Tinbergen Institute, The Netherlands

ARTICLE INFO

Keywords:

Overnight volatility
Realized variance
F distribution
Score-driven dynamics
Value-at-Risk
Expected Shortfall

ABSTRACT

We present a new model to decompose total daily return volatility into high-frequency-based open-to-close volatility and a time-varying scaling factor. We use score-driven dynamics based on fat-tailed distributions to obtain robust volatility dynamics. Applying our new model to a 2001–2018 sample of individual stocks and stock indices, we find substantial in-sample variation of the daytime-to-total volatility ratio over time. We apply the model to out-of-sample forecasting, evaluated in terms of Value-at-Risk and Expected Shortfall. Models with a non-constant volatility ratio typically perform best, particularly in terms of Value-at-Risk. Our new model performs especially well during turbulent times. All results are generally stronger for individual stocks than for index returns.

© 2020 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

1. Introduction

Volatility is a key ingredient for volatility traders, (see for example Sinclair, 2013, w.r.t. volatility strategies), risk managers, and asset managers. Most of the econometric models in use are in some way based on the popular GARCH model of Bollerslev (1986) or the stochastic volatility model of Taylor (2008), including extensions, such as the threshold GARCH model of Glosten et al. (1993), among many others.

Due to the availability of high-frequency data, volatility measurements, modeling, and forecasting have improved significantly (Andersen et al., 2003). The literature

on (daily) volatility modeling moved from modeling the distribution of daily close-to-close returns (with volatility treated as latent) to modeling the ex-post realized intra-day variance either directly (Corsi, 2009; Engle & Gallo, 2006), or jointly with returns in a hybrid model (Shephard & Sheppard, 2010).

Although high-frequency data improve measurements and forecasts of volatility, such data are typically only available during trading hours, particularly for individual stocks. The literature has thus far overcome this challenge by simply discarding overnight volatility, whether by only considering realized measures, by using open-to-close returns, or by a combination of both (Corsi, 2009; Engle & Gallo, 2006; Noureldin et al., 2012; Shephard & Sheppard, 2010, among others). However, investors also receive relevant information during off-trading hours, including earnings and merger announcements. Moreover, due to the worldwide connectedness of financial markets, overnight information in one market may be important for the overall volatility of stock returns in another market in another time-zone (Taylor, 2007; Todorova & Souček, 2014).

[☆] We appreciate the comments of participants at the 13th Meeting of the NESG (Amsterdam, May 2019) and seminar participants at the Accounting/Finance and Economics Seminar Series at Cork University Business School (September 2019) and the Centre for Financial Markets (CFM) Research Seminar Series at the Michael Smurfit Graduate Business School (September, 2019).

^{*} Correspondence to: Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands.

E-mail address: a.opschoor@vu.nl (A. Opschoor).

An alternative way of dealing with overnight volatility is to scale up (or down) the realized variance measured during trading hours by a *fixed* constant (Ahoniemi et al., 2015; Hansen & Lunde, 2005; Koopman et al., 2005). Ahoniemi and Lanne (2013) compare several of these scaled volatilities and conclude that the optimal scaling method differs depending on whether one considers individual stocks or stock indices.

This paper develops a new score-driven model for modeling daily close-to-close volatility. Our model takes into account the availability of high-frequency data during the day (open-to-close). We allow the ratio of open-to-close volatility to close-to-close volatility to vary over time; compare Hansen and Lunde (2005). Considering a time-varying ratio of volatilities is also in line with the work of Ahoniemi et al. (2015) and Linton and Wu (2020). They show that daily and overnight volatility can be distinct processes with their own dynamics and, thus, with a time-varying ratio between the two. For example, during periods of crisis, relatively more information might be released outside trading hours than during calm periods, or vice versa. This implies that assuming a constant ratio between daytime and total volatility might lead to biased predictions of daily volatility levels and associated risk measures.

We capture the time-variation in the daytime-to-total volatility ratio as an integrated part of our model by using the generalized autoregressive score (GAS) framework of Creal et al. (2013). Score-driven dynamics have been applied successfully in a number of other settings, including volatility and location modeling (Harvey, 2013; Harvey & Luati, 2014) and systemic risk modeling (Lucas et al., 2014; Oh & Patton, 2018). The availability of a closed-form expression for the likelihood function and the optimality of score-driven steps (see Blasques et al., 2015) make the GAS framework attractive for modeling time-varying parameters. Moreover, parameter estimation is straightforward using standard maximum likelihood. Our score-driven (GAS) model has two key features. First, we account for fat-tails of the realized variance by assuming an F distribution for the daytime (open-to-close) volatility. The (matrix) F distribution was recently introduced to financial econometrics by Opschoor et al. (2018). Score-driven dynamics for the F distribution result in an intuitive propagation mechanism for daytime volatility: incidentally large values of the realized variance (such as the 2010 flash crash) are downweighted in the volatility equation and thus only have a limited effect on the estimated volatility path. Second, the volatility ratio in our model also follows score-driven dynamics with highly intuitive propagation dynamics. Moreover, the fat-tails in our model again ensure that incidentally large observed volatility ratios have only limited impact on the estimated path of this parameter.

We apply the new model in a forecasting experiment, focusing on commonly used downside risk measures such as Value-at-Risk and Expected Shortfall. Both are widely used across the financial industry for measuring risk and for setting margins and buffers, both for individual assets and for asset portfolios. We use a 2001–2018 sample of 93 randomly selected S&P 500 stock returns as well as

a sample of 19 stock index returns across the world. In-sample, we find clear evidence of time-variation in the daytime-to-total volatility ratio. This holds particularly for individual stocks, but also for at least half of the stock indices. Many stocks had relatively small ratios before the global financial crisis and relatively large ratios during it. Hence, relatively more information accumulates outside regular trading hours during these volatile periods. For many stocks, our model improves significantly upon the HEAVY model of Shephard and Sheppard (2010) or a model with a static volatility ratio. We also confirm Linton and Wu (2020) by showing that modeling the overnight volatility separately leads to low values of the degrees of freedom parameter, with subsequent potential problems for the existence of sufficient moments (see also Berkman et al., 2012).

Out-of-sample, we compare one-step-ahead forecasts of the 97.5% and 95% Expected Shortfall and the 99% and 95% Value-at-Risk for our new model and its competitors. We apply the recently developed unconditional and conditional backtests of Du and Escanciano (2016) to backtest the Expected Shortfall predictions, and the familiar Christoffersen (1998) (un)conditional coverage backtesting procedure for VaR.

The results are clearest for individual stocks, in particular during periods of crisis. Models with a time-varying volatility ratio perform best for VaR. Score-driven updates of the volatility ratio are particularly helpful during turbulent times. For individual stocks and the Expected Shortfall, score-driven updates again perform well for unconditional backtests, with the HEAVY model as a close competitor for conditional backtests. For the 19 stock indices, the results are more mixed, and all models behave more closely. Still, our models with score-driven or rolling volatility ratios together with the HEAVY model perform best, both in turbulent and quiet times, and in deep and moderate tail areas. The fat-tailedness and idiosyncratic nature of individual stocks thus seems to support the additional time-variation in the volatility ratio the most.

The rest of this paper is set up as follows. In Section 2, we introduce the new score-driven (GAS) models for the dynamics of the volatility and the daytime-to-total volatility ratio. In Section 3, we briefly review the backtesting procedures. Section 4 provides our in-sample and out-of-sample empirical results. Section 5 concludes.

2. Modeling framework

Let $r_t \in \mathbb{R}$ be the total daily close-to-close return of a financial asset on day t , $t = 1, \dots, T$, and $h_t = \text{Var}(r_t | \mathcal{F}_{t-1})$ its (unobserved) conditional close-to-close variance, where $\mathcal{F}_{t-1}$ is the information set containing all information up to and including time $t - 1$. The quantities r_t and h_t can be decomposed into

$$r_t = r_{o,t} + r_{d,t}, \quad (1)$$

$$\begin{aligned} h_t &= h_{o,t} + h_{d,t} + 2 \text{Cov}(r_{o,t}, r_{d,t}) \\ &= h_{o,t} + h_{d,t} + 2\rho_{o,d,t} \sqrt{h_{o,t}} \sqrt{h_{d,t}}, \end{aligned} \quad (2)$$

where $r_{o,t}$ ($h_{o,t}$) and $r_{d,t}$ ($h_{d,t}$) denote the overnight close-to-open and daytime open-to-close return (volatility), respectively, and $\rho_{o,d,t}$ denotes the correlation between the overnight and daytime return at time t .

Due to the availability of high-frequency data during trading-hours, we can obtain accurate measurements of the daytime variance $h_{d,t}$. This can be done, for example, by computing the realized variance RV_t (see Andersen & Bollerslev, 1998), defined as the sum of, e.g., five-minute intraday returns over the course of the trading day (9:30–16:00). Typically, there are no high-frequency data outside trading hours, in particular for individual stocks, such that an equivalent, accurate measure for $h_{o,t}$ is lacking.

Our object of interest in this paper is to model the forecasting distribution of r_t (and hence to model the dynamics of h_t). We test the model's performance for forecasting typical risk measures such as Value-at-Risk (VaR) and Expected Shortfall (ES). We can construct such a model in at least two different ways. First, we can model h_t directly based on a time series model that possibly includes (a scaled version of) RV_t . Second, we can model the daytime and overnight variances (and their covariance) separately. Ahoniemi et al. (2015) compare both ways in the context of equity indices. They scale RV_t to a daily close-to-close variance level $RV_t^{\text{fix-c},1}$ using the method of Hansen and Lunde (2005). This method minimizes the variance of a weighted sum of the overnight squared returns and the realized variances,

$$RV_t^{\text{fix-c},1} = \omega_1 r_{o,t}^2 + \omega_2 RV_t, \quad (3)$$

subject to the constraint $\omega_1 \mu_1 + \omega_2 \mu_2 = \mu_0$, with $\mu_1 = \mathbb{E}[r_{o,t}^2]$, $\mu_2 = \mathbb{E}[RV_t]$, and $\mu_0 = \mathbb{E}[RV_t + r_{o,t}^2]$. In a second step, this scaled realized variance $RV_t^{\text{fix-c},1}$ is then used in a time series model. As an alternative, Koopman et al. (2005) use open-to-close and overnight returns to construct

$$RV_t^{\text{fix-c},2} = c RV_t = \frac{\sum_{t=1}^T r_{d,t}^2 + \sum_{t=1}^T r_{o,t}^2}{\sum_{t=1}^T r_{d,t}^2} RV_t, \quad (4)$$

which is related to the scaling method of Hansen and Lunde (2005),

$$RV_t^{\text{fix-c},3} = c RV_t = \frac{\sum_{t=1}^T r_t^2}{\sum_{t=1}^T RV_t} RV_t, \quad (5)$$

which scales the RV_t by the ratio of close-to-close volatility to the open-to-close volatility. Ahoniemi and Lanne (2013) compare the above scaling methods and several others. They conclude that for individual stocks it is better to drop the overnight information, while for indices, the scaling method (3) of Hansen and Lunde (2005) performs best.

This paper deviates from previous approaches in that we no longer treat the proportionality parameter c as a static parameter. This is in line with findings of Ahoniemi et al. (2015), Linton and Wu (2020), among others, who show that the daytime and overnight volatility processes have their own dynamics, and that, hence, the ratio c might not be constant over time. Instead of a static c , we propose a dynamic parameter c_t that we endow with the score-driven (GAS) dynamics of Creal et al. (2013); see also Harvey (2013). The parameter c_t is now an integrated part of the model and can no longer be obtained as a separate stage in a multiple-step procedure. Our score-driven

model accounts for the fact that realized variance measures can be quite fat-tailed empirically; see for instance Opschoor et al. (2018). This is a real concern and might result in aberrant behavior of standard, static estimates of c due to a lack of existing moments of returns. As we assume fat-tailed distributions for both the returns and the realized variances, our method is robust against such data features.

Our new score-driven model reads

$$RV_t = h_{d,t} u_{d,t}, \quad u_{d,t} \sim F(1, v_1, v_2), \quad (6)$$

$$r_t = \mu + \sqrt{h_t} \varepsilon_t, \quad \varepsilon_t \sim t(0, 1, v_3), \quad h_t = c_t h_{d,t}, \quad (7)$$

$$h_{d,t+1} = \omega_1 + \alpha_1 s_{h,t} + \beta_1 h_{d,t}, \quad (8)$$

$$c_{t+1} = \omega_2 + \alpha_2 s_{c,t} + \beta_2 c_t, \quad (9)$$

where $s_{h,t}$ and $s_{c,t}$ denote the scaled scores of the predictive conditional distributions of RV_t and r_t , respectively, i.e.,

$$\begin{aligned} s_{h,t} &= S_{h,t} \nabla_{h,t} = S_{h,t} \frac{\partial \ln p_{RV}(RV_t | h_{d,t}, \mathcal{F}_{t-1}; v_1, v_2)}{\partial h_{d,t}} \\ &= \frac{v_1}{v_1 + 1} (w_{h,t} RV_t - h_{d,t}), \end{aligned} \quad (10)$$

$$\begin{aligned} s_{c,t} &= S_{c,t} \nabla_{c,t} = S_{c,t} \frac{\partial \ln p_r(r_t | h_{d,t}, c_t, \mathcal{F}_{t-1}; \mu, v_3)}{\partial c_t} \\ &= w_{c,t} \frac{\varepsilon_t^2}{h_{d,t}} - c_t, \end{aligned} \quad (11)$$

$$\begin{aligned} w_{h,t} &= \frac{v_1 + v_2}{v_2 - 2 + v_1 \cdot RV_t / h_{d,t}} \\ w_{c,t} &= \frac{v_3 + 1}{v_3 - 2 + \varepsilon_t^2 / h_t}. \end{aligned}$$

The fat-tailed $F(v_1, v_2)$ and Student's $t(v_3)$ distributions in (6) and (7) accommodate the possibly fat-tailed nature of realized variances and returns (see also Opschoor et al., 2018). To account for the curvature of the scores of the predictive density functions, we follow Creal et al. (2013) and Opschoor et al. (2018) and scale $\nabla_{h,t}$ and $\nabla_{c,t}$ by $2h_{d,t}^2/(v_1 + 1)$ and $\mathcal{I}^{-1}$, respectively, where $\mathcal{I}$ denotes the conditional Fisher-information matrix. We label our model the GAS tv- c_t model, indicating we have a model with time-varying c .

Both scores in (10)–(11) have an intuitive interpretation. First, $s_{h,t}$ is the difference between a weighted RV_t and $h_{d,t}$, where the weight accounts for the fat-tailedness of RV_t . Large values of RV_t imply a low weight in (10), such that the impact of an ‘outlying’ value of RV_t on $h_{d,t+1}$ is mitigated. Similarly, $s_{c,t}$ holds the difference between c_t and a weighted ratio of the squared total (demeaned) return and the high-frequency based volatility estimate. Again, large values of squared (demeaned) daily returns ε_t receive a smaller weight via $w_{c,t}$ in (11).

We estimate our GAS tv- c_t model by maximum likelihood. Although the model parameters can be estimated in one step, we opt for a two-step approach for numerical efficiency. In a first step, we estimate the parameters associated with the daytime (open-to-close) volatility. Let $\theta_D = \{v_1, v_2, \omega_1, \alpha_1, \beta_1\}$ and $\theta_{tot} = \{\mu, v_3, \omega_2, \alpha_2, \beta_2\}$ denote the static parameter vectors of the daily and the

total volatility processes, respectively. We first maximize the log-likelihood associated with the F distribution:

$$\max_{\theta_D} \mathcal{L}^F = \max_{\theta_D} \sum_{t=1}^T \log p_{RV}(RV_t | h_{d,t}, \mathcal{F}_{t-1}, \theta_D). \quad (12)$$

Given $\hat{\theta}_D$, we calculate the filtered $h_{d,t}$, which is one of the ingredients for the Student's t log-likelihood function for the total return r_t :

$$\max_{\theta_{tot}} \mathcal{L}^{Stud} = \sum_{t=1}^T \log p_r(r_t | \hat{h}_{d,t}, c_t, \theta_{tot}). \quad (13)$$

Comparing estimates and log-likelihoods for the one-step and two-step estimation procedure for individual stocks, we found that the efficiency loss due to two-step estimation procedure turns out to be very small.

3. Forecasting Value-at-Risk and Expected Shortfall

We assess the economic value of our new model with a dynamic daytime-to-total volatility ratio c_t in terms of VaR and ES predictions. Both are widespread risk measures in the financial industry, both for individual assets and asset portfolios.

The one-step-ahead q -level VaR is defined as the quantity $\text{VaR}(1 - q)$, such that

$$\Pr[r_t < \text{VaR}_t(1 - q) | \mathcal{F}_{t-1}] = q; \quad (14)$$

In line with the Basel accords, we consider $q = 0.01$, which corresponds to a 99% VaR. In addition, we set q to 0.05. Since all our models assume a conditional Student's t density for the total return r_t with time-varying variance h_t , the one-step conditional $\text{VaR}(q)$ equals

$$\text{VaR}_t(1 - q) = \mu + t^{-1}(q, \nu_3) \sqrt{h_t}, \quad (15)$$

with $t^{-1}(q, \nu_3)$, the q -quantile of a Student's t distribution with variance 1, and ν_3 degrees of freedom.

As the VaR does not consider any loss information beyond the q -quantile, a second risk measure is introduced: the ES. The one-step-ahead ES is defined as

$$\begin{aligned} \text{ES}_t(1 - q) &= \mathbb{E}[r_t | \mathcal{F}_{t-1}, r_t < \text{VaR}_t(1 - q)] \\ &= \frac{1}{q} \int_0^q \text{VaR}_t(q) dq. \end{aligned} \quad (16)$$

The ES thus integrates the VaR over a continuum of confidence levels. Under our Student's t distributional assumption for r_t , we have

$$\text{ES}_t(1 - q) = \mu + \frac{1}{q} p_r(x_q | \nu_3) \left(\frac{(\nu_3 - 2) + x_q^2}{\nu_3 - 1} \right) \sqrt{h_t}, \quad (17)$$

where x_q is the q -quantile of a standardized Student's t distribution with unit variance and ν_3 degrees of freedom. We consider the 97.5% (as prescribed by Basel) and the 95% Expected Shortfall.

Given the one-step-ahead VaR and ES over the out-of-sample forecasting period, we use several statistical tests to assess the adequacy of our forecasts. We define the hit-process as

$$h_t(q) = 1_{[r_t < \text{VaR}_t(1 - q)]}, \quad (18)$$

where $1_{[A]}$ denotes the indicator function of event A . We use the unconditional coverage (UC) and the independence likelihood ratio test of Christoffersen (1998), both with an asymptotic $\chi^2(1)$ distribution. The first test corresponds to the null $H_0 : \mathbb{E}[h_t(q)] = \Pr[h_t(q) = 1] = q$, while the latter tests the null hypothesis $H_0 : \Pr[h_{t+1}(q) = 1 | h_t(q)] = q$. Both tests can also be combined in a conditional coverage (CC) test that is asymptotically $\chi^2(2)$ distributed.

To evaluate the ES, we use the recently developed (un)conditional backtests of Du and Escanciano (2016). Consider the following cumulative violation process:

$$H_t(q) = \frac{1}{q} \int_0^q h_t(u) du = \frac{1}{q} (q - z_t) 1_{z_t \leq q}, \quad (19)$$

where $u_t = F(r_t; \mathcal{F}_{t-1})$ denotes the probability integral transform (PIT), or cdf transform of the return r_t based on past information. Du and Escanciano (2016) show that testing the unconditional correct specification for the ES boils down to testing whether the mean of $H_t(q)$ equals $q/2$. This can be tested based on the test-statistic:

$$t_{ES} = \frac{\overline{H(q)} - q/2}{\sqrt{v_{ES}(q)/P}}, \quad (20)$$

where $\overline{H(q)}$ is the out-of-sample mean of $H_t(q)$, $v_{ES}(q) = \text{Var}(H_t(q)) = q(1/3 - q/4)$, and P is the number of out-of-sample observations. This test-statistic follows an asymptotic standard normal distribution.

The conditional backtest aims to detect any autocorrelation in the cumulative violation process $H_t(q)$. We define $\gamma_j = \text{Cov}(H_t(q), H_{t-j}(q))$ and $\rho_j = \gamma_j / \gamma_0$ for $j \geq 0$. Then, the sample counterparts read

$$\begin{aligned} \hat{\rho}_j &= \hat{\gamma}_j / \hat{\gamma}_0, \\ \hat{\gamma}_j &= \frac{1}{T - j} \sum_{t=1+j}^T (H_t(q) - q/2)(H_{t-j}(q) - q/2). \end{aligned} \quad (21)$$

Du and Escanciano (2016) show that the asymptotic distribution of $\hat{\rho}_j$ is Gaussian. The conditional backtest now results in a well-known Box–Pierce test statistic on the joint significance of the first m autocorrelations: $BP_{ES}(m) = T \sum_{j=1}^m \hat{\rho}_j^2$, which is asymptotically χ^2 distributed with m degrees of freedom.

4. Empirical application

4.1. Data

We use two data sets. The first data set contains a balanced sample of daily open-to-close returns and daily realized variances for 93 S&P 500 constituents. The 93 stocks are randomly chosen from different industries, such as financials, materials, etc., that are available over the entire sample period from January 2, 2001 until December 31, 2018. Table 1 provides an overview of the individual stocks. We have $T = 4472$ trading days, after leaving out days where the market was open for half a day. The Financial industry covers most companies (i.e., 18), followed by Consumer Services and Energy.

Table 1
S&P 500 constituents.

Ind Nr.	Industry	# Comp.	Tickers
1	Capital Goods	10	AA,BA,CAT,HON,F,NOC,UTX, A,IR,GD
2	Finance	18	AXP,JPM,AIG,BAC,C,KEY,MTB, COF,USB,BBT,STI,WFC,GS,MS, MMC,HIG,PNC,MCO
3	Energy	11	GE,XOM,MUR,SLB,CVX,HAL, OXY,APC,SU,CNX,PXD
4	Consumer Services	14	HD,MCD,WMT,TGT,BXP,DIS,JCP, NLY,ANF,EQR,WY,RCL,WSM,TV
5	Consumer Non-Durables	8	KO,MO,SY,CL,AVP,GIS, CPB,EL
6	Health Care	11	PFE,ABT,BAX,JNJ,LLY,THC,MMM, MRK,BMY,MDT,C
7	Public Utilities	7	AEP,AEE,DUK,SO,WMB,VZ,EXC
8	Technology	4	IBM,DOV,HPQ,TSM
9	Basic Industries	6	PG,FLR,AES,IP,ATI,LPX
10	Transportation	4	LUV,UPS,NSC,FDX

This table lists ticker symbols of our sample of 93 random companies listed on the S&P 500 index during the period from January 2, 2001 until December 31, 2018. The tickers are grouped by industry.

We retrieve consolidated trades (transaction prices) from the Trade and Quote (TAQ) database from 9:30 until 16:00. Before 2015, the time-stamp precision equals one second, and after 2015, it increases to one millisecond. After cleaning the high-frequency data following the guidelines of Barndorff-Nielsen et al. (2009) and Brownlees and Gallo (2006), we construct realized variances based on five-minute returns.

The second data set consists of 19 stock indices across the world. Table 2 summarizes the composition. Daytime and overnight returns as well as the corresponding five-minute Realized Variances are obtained from the Oxford-Man Institute's 'Realized Library'.¹ The last column of the table shows that the number of daily observations differs for each index due to country-specific holidays.

4.2. Competing models

In-sample, we benchmark our new GAS tv- c_t model against three alternative models.

As a first alternative, we consider a model with static c , where we estimate c using the same methods as for the GAS tv- c_t model, but with $\alpha_2 = \beta_2 = 0$. We call this the GAS fix- c model.

The second model is the HEAVY model of Shephard and Sheppard (2010). This model considers the total daily return r_t , and uses the lagged realized variance as an explanatory variable for the variance h_t . This model is defined as

$$\begin{aligned} r_t &= \mu_t + \sqrt{h_t} \varepsilon_t, & \varepsilon_t &\sim t(0, 1, \nu_3), \\ h_{t+1} &= \omega_H + \alpha_H RV_t + \beta_H h_t. \end{aligned} \quad (22)$$

The HEAVY model does not take into account the possible difference in the overnight and daily volatility processes. Note that the HEAVY model is also invariant to any form of static scaling of the realized variance, as the scaling factor is perfectly mopped up by the α coefficient. Put

differently, the HEAVY model implicitly scales up RV_t by a constant while maximizing the log-likelihood.

The third benchmark model looks at all three components of (2) separately; compare Ahoniemi et al. (2015). Similar to our own score-driven GAS model, the first component $h_{d,t}$ is modeled by the GAS F model (8). The second term, the overnight volatility $h_{o,t}$, is modeled by a 'GAS t ON' model, where ON stands for OverNight. This model is defined as

$$\begin{aligned} r_{o,t} &= \mu_o + \sqrt{h_{o,t}} \varepsilon_{o,t} & \varepsilon_{o,t} &\sim t(0, 1, \nu_3), \\ h_{o,t+1} &= \omega_o + \alpha_o s_{o,t} + \beta_o h_{o,t} + \gamma_o \varepsilon_{D,t}^2, \\ s_{o,t} &= S_{o,t} \nabla_{o,t} = w_{o,t} \varepsilon_{o,t}^2 - h_{o,t}, \\ w_{o,t} &= \frac{\nu_3 + 1}{\nu_3 - 2 + (\varepsilon_{o,t}^2 / h_{o,t})}, \end{aligned} \quad (23)$$

where $\varepsilon_{o,t}$ denote the demeaned overnight return at time t . We follow Ahoniemi et al. (2015) and incorporate the squared daily return $\varepsilon_{D,t}^2$ into the equation to allow for a spillover effect of daily volatility on overnight volatility. Finally, we estimate $\rho_{D,o,t}$ in (2) by its sample correlation counterpart. We deviate here from Ahoniemi et al. (2015), who estimate the DCC model of Engle (2002) for the time-varying correlation between the daytime and overnight return. For individual stocks, our preliminary analyses did not show any empirical support for a DCC model during this step. We call this composite benchmark model of 'GAS t ON' and 'GAS F ' the GAS sep model, to indicate that daytime and overnight volatility are modeled separately.

4.3. In-sample results

We first estimate the GAS tv- c_t , GAS fix- c , GAS sep (i.e., GAS F plus GAS t ON), and HEAVY models by maximum likelihood using all observations.

Table 3 and Fig. 1 present the results. The table shows three interesting findings. First, the average degrees-of-freedom parameter ν_3 of the Student's t distribution across all 93 individual stocks is around 6, indicating that the conditional returns are fat-tailed. This becomes extreme

¹ See <http://realized.oxford-man.ox.ac.uk/>.

Table 2
List of stock indices.

Ind Nr.	Ticker	Name Index	# Obs.
1	AEX	AEX index	4574
2	AORD	All Ordinaries	4534
3	BFX	Bell 20 Index	4575
4	BSESIN	S&P BSE Sensex	4433
5	DJI	Dow Jones Industrial Average	4472
6	FCHI	CAC 40	4580
7	FTSE	FTSE 100	4513
8	GDAXI	DAX	4550
9	HSI	HANG SENG Index	4366
10	IBEX	IBEX 35 Index	4548
11	IXIC	Nasdaq 100	4494
12	KS11	Korea Composite Stock Price Index (KOSPI)	4428
13	MXJ	IPC Mexico	4488
14	N225	Nikkei 225	4376
15	RUT	Russell 2000	4481
16	SPX	S&P 500 Index	4476
17	SSEC	Shanghai Composite Index	4345
18	SSMI	Swiss Stock Market Index	4498
19	STOXX50E	EURO STOXX 50	4570

This table lists 19 stock indices during the period from January 2, 2001 until December 31, 2018. We list the ticker, the name, and the number of daily observations of the index.

for the overnight returns, as indicated by an average value of around 3 for the GAS t ON model. The index returns are less fat-tailed, with respect to both the daily close-to-close returns and the overnight returns. Still, the average value of ν_3 is 5 for the overnight index returns. This result confirms Linton and Wu (2020), for example, and it shows that overnight returns are considerably fat-tailed and that fourth-order moments may not exist, particularly not for individual stocks.

Second, the average persistence parameter β of the GAS tv - c_t model is rather high, both for individual stocks and for stock index returns (0.916 and 0.961, respectively). Hence almost all stocks exhibit a slowly moving daytime-to-total volatility ratio c_t .

Third, the average ν_2 parameter of the GAS F model equals 13.53 (14.37 for stock indices), indicating that realized variances are fat-tailed as well. This result has also been found for realized covariance matrices by Opschoor et al. (2018). As discussed in Section 2, our score-driven model automatically accounts for sudden outliers in $R_{V,t}$. These are downweighted through the score, such that the future path of volatilities is not disrupted; see (10).

Fig. 1 shows the difference in maximized log-likelihood between the GAS tv - c_t and the HEAVY t model (upper panels), and between the GAS tv - c_t and GAS fix - c model (lower panels). A positive value means that the GAS tv - c_t model has a higher log-likelihood. The likelihood of the GAS tv - c_t model exceeds that of the HEAVY model for a large number of individual stocks. We see the same pattern for the indices, though the pattern is somewhat less pronounced. Differences can be substantial: for example for the IPC Mexico and Russell 2000 indices, the log-likelihood difference exceeds 80 points (y -axis capped).

The lower panels show that modeling c_t as a time-varying rather than as a static parameter also improves the likelihood. In particular, for 77 out of the 93 stocks and 12 out of the 19 indices, twice the log-likelihood difference exceeds the typical critical value (5.99) of a

$\chi^2(2)$, suggesting the rejection of the null $H_0 : \alpha_2 = \beta_2 = 0$ of a static c at a 5% level.

Fig. 2 shows the fitted path of c_t for the GAS tv - c_t model for eight stocks and two stock indices along with the estimated c of the GAS fix - c model. All panels show that the variance ratio c_t varies substantially over time. We see clear peaks for most stocks and indices during the heat of the global financial crisis (2008–2009), particularly for the financial stocks. Interestingly, for some stocks, the value of c falls below 1 during the years 2005–2006, hence correcting the fitted daily volatility downwards. The time-varying ratio of the Russell 2000 Index is relatively extreme, possibly due to the early 2000s recessionary period after the burst of the dot-com bubble.

Finally, comparing the GAS sep model with the GAS tv - c_t model in terms of likelihood is not straightforward: the GAS tv - c_t model describes daily returns and daytime realized volatilities, while the GAS sep model also models the overnight returns separately. The likelihoods are thus incomparable. In the next section, we compare all models by means of out-of-sample forecast performance using downside risk measures.

We conclude that, statistically and in-sample, it is beneficial for a large number of stocks and indices to model open-to-close and close-to-close volatility by two processes using a time-varying scaling factor. In addition, we have seen that the total information outside trading hours with respect to the information during the day varies heavily over time.

4.4. Out-of-sample results

To assess the short-term forecasting performance of the different models, we consider one-step-ahead 99% and 95% VaR and 97.5% and 95% ES predictions. We use a moving-window of 1000 observations, starting from 2001–2004, such that the global financial crisis and the European sovereign debt crisis are part of the out-of-sample period.

Fig. 1. Log-likelihood differences vis-a-vis the GAS $tv-c_t$ model. This figure shows the difference in the maximized log-likelihood between the GAS $tv-c_t$ model and the HEAVY t model (upper panel), and between the GAS $tv-c_t$ model and the GAS $fix-c$ model (lower panel). The figures on the left depict the case for 93 stocks listed on the S&P 500 index, while the figures on the right show the differences for the 19 indices listed in Table 2. A positive number means a better fit of the GAS model with a time-varying ratio of total and daily volatility.

Fig. 2. Time-varying volatility ratios. This figure shows the fitted ratio of the total daily volatility and the open-to-close volatility from the GAS $tv-c_t$ model for various stocks and two indices. The upper panel shows the ratio for four financial stocks (AIG, BAC, WFC, and C) and four non-financial stocks (AVP, NLY, EQR, and CPB). The lower panel depicts the ratio for the Russell 2000 and Hang Seng Index. In addition, the horizontal lines represent the estimated $\hat{c}$ from the GAS $fix-c$ model for the same stocks and indices. The sample covers January 2, 2001 until December 31, 2018.

Table 3

Full sample average parameter estimates of all stocks.

		μ	ω	α	β	ν_3	ν_1	ν_2
Panel A: 93 stocks from the S&P 500								
GAS tv- c_t	mean	0.030	0.108	0.024	0.916	6.31		
	sd	(0.02)	(0.25)	(0.03)	(0.20)	(1.47)		
GAS fix- c	mean	0.029	1.329			6.22		
	sd	(0.02)	(0.14)			(1.45)		
GAS sep (return part) (GAS t ON)	mean	0.031	0.015	0.126	0.990	2.98		
	sd	(0.03)	(0.02)	(0.02)	(0.01)	(0.57)		
GAS sep (return part) (GAS F)	mean		0.053	0.972	0.981		24.70	13.53
	sd		(0.03)	(0.06)	(0.01)		(2.99)	(1.91)
HEAVY	mean	0.029	0.103	0.532	0.567	6.32		
	sd	(0.02)	(0.11)	(0.16)	(0.13)	(1.46)		
Panel B: 19 stock indices								
GAS tv- c_t	mean	0.033	0.054	0.014	0.961	9.76		
	sd	(0.02)	(0.07)	(0.01)	(0.06)	(2.32)		
GAS fix- c	mean	0.032	1.732			9.22		
	sd	(0.02)	(0.50)			(2.05)		
GAS sep (return part) (GAS t ON)	mean	0.038	0.004	0.172	0.994	5.03		
	sd	(0.05)	(0.00)	(0.14)	(0.01)	(3.40)		
GAS sep (vol part) (GAS F)	mean		0.017	0.895	0.985		19.281	14.37
	sd		(0.01)	(0.11)	(0.01)		(6.22)	(3.86)
HEAVY	mean	0.032	0.029	0.584	0.635	9.59		
	sd	(0.02)	(0.03)	(0.22)	(0.12)	(2.13)		

This table reports maximum likelihood parameter estimates of the GAS tv- c_t , GAS fix- c , GAS sep (GAS t ON plus GAS F), and HEAVY t models. Panel A shows the results of these models applied to a balanced panel of daily equity returns, realized variances, or overnight returns of 93 assets listed on the S&P 500 index. Panel B lists the results of the same models applied to 19 stock indices across the world. We list the mean of the estimated parameters over all 93 stocks (19 indices), with the standard deviation (SD) in parentheses. The sample covers January 2, 2001 until December 31, 2018. In the case of U.S. stocks, this boils down to 4472 observations. We refer to Table 2 for the number of observations of each index.

We re-estimate our models every 50 observations, or about every two months. Effectively, this means that even the GAS tv- c_t model with $\alpha_2 = \beta_2 = 0$ now has time-variation in c_t . We therefore re-label this model GAS roll- c_t to indicate that c is slowly updated (every 50 observations) using a rolling window. To continue looking at the effect of having a static c , we also implement a new GAS fix- c model using the scaling method of Koopman et al. (2005) (see (4)), based on the *whole* sample. Using all data gives this method an advantage insofar as future data are used in the forecasts. Despite this advantage, it turns out that such a static c model performs rather poorly for forecasting purposes, compared to the time-varying c_t specifications.

4.4.1. Value-at-Risk and expected shortfall results on individual stocks

Table 4 shows the results for individual stocks. Panels A.1 and A.2 show the unconditional (UC) and conditional coverage (CC) test of Christoffersen (1998) for VaR, while Panels B.1 and B.2 show unconditional and conditional (with $m = 10$) backtests of Du and Escanciano (2016). For each test, we report the number rejections (across all 93 stocks) of the null-hypothesis using a 10%, 5%, or 1% level.

The results show that the model with fixed c (GAS fix- c) generally performs the worst among all models considered. This holds both in terms of VaR and ES forecasts. A time-varying volatility ratio c_t thus seems to be called for.

The results also show that, overall, the model with infrequently updated c_t (GAS roll- c_t) performs the best. The HEAVY model performs better only for the ES and a

conditional coverage criterion. In terms of VaR, the GAS roll- c_t model with infrequent updating performs particularly well for the 99% VaR. Interestingly, the GAS sep model appears a close competitor at the 99% VaR. We return to this when looking at a sub-period analysis. The GAS sep model, however, fails considerably when forecasting less far out into the tails, such as at the 95% VaR. By contrast, the GAS tv- c_t and GAS roll- c_t models continue to perform quite well in this setting. Also note that the HEAVY model performs less well in terms of unconditional coverage, but is similar to the GAS roll- c_t for conditional coverage.

To gain more insight into the role of the GAS models with a time-varying volatility ratio c_t , we split the sample into a turbulent (2005–2011) and a more tranquil (2012–2018) period. Table 5 shows the UC and CC backtesting results for the VaR for both sub-samples.

The results for the turbulent period are striking: Panel A clearly shows that the GAS tv- c_t model is far superior to all other models. It is precisely during turbulent times that the time-variation pays off for risk prediction in terms of VaR. During these times, more information also comes out after trading hours and at a faster pace. The frequent updating in the GAS tv- c_t model keeps track of this.

Panel B shows that in tranquil periods, the results are different. In particular, for 99% VaR modeling, the day and night volatility processes perform well separately. This appears in line with previous studies such as Ahoniemi et al. (2015). However, during tranquil times when forecasting less far out into the tails, such as at the 95% VaR, the GAS tv- c model again outperforms both the fixed c and GAS sep models. The HEAVY and GAS roll- c_t perform comparably.

Table 4
One-step-ahead VaR and ES predictions (individual stocks).

	Unconditional test					Conditional test				
	GAS tv- c_t	GAS roll- c_t	GAS fix- c	GAS sep	HEAVY	GAS tv- c_t	GAS roll- c_t	GAS fix- c	GAS sep	HEAVY
Panel A.1: 99% VaR										
# p -val < 0.10	36	26	50	30	37	40	34	54	34	37
# p -val < 0.05	27	15	43	18	27	28	21	45	19	26
# p -val < 0.01	7	5	28	7	15	10	7	27	10	10
Panel A.2: 95% VaR										
# p -val < 0.10	26	18	58	44	32	40	39	69	55	35
# p -val < 0.05	15	16	50	39	23	25	28	59	47	27
# p -val < 0.01	4	6	42	21	15	10	15	44	33	13
Panel B.1: 97.5% ES										
# p -val < 0.10	23	23	35	19	31	39	48	40	35	29
# p -val < 0.05	16	17	29	17	25	32	38	34	31	24
# p -val < 0.01	7	10	16	9	13	25	30	29	22	15
Panel B.2: 95% ES										
# p -val < 0.10	18	22	36	25	27	41	49	49	41	36
# p -val < 0.05	7	13	29	19	19	32	40	42	38	30
# p -val < 0.01	5	7	16	15	7	24	31	34	33	21

This table reports results on one-step-ahead 99% and 95% Value-at-Risk (VaR) and 97.5% and 95% Expected Shortfall (ES) predictions, using the GAS tv- c_t , GAS roll- c_t , GAS fix- c , GAS sep, and HEAVY t models, applied to daily equity returns of 93 assets listed on the S&P 500 index. We use a moving window of 1000 observations. Panel A lists results on backtesting the VaR. Specifically, the sub-panels A.1 and A.2 show results of the unconditional and conditional coverage tests of Christoffersen (1998). Panel B shows results of the unconditional and conditional backtests of Du and Escanciano (2016). For all tests, we list the number of cases where the p -value falls below 10%, 5%, or 1%. Bold numbers represent the minimum across all models in the same row. The out-of-sample covers January 14, 2005 until December 31 2018 and contains 3472 observations.

Table 5
One-step-ahead VaR predictions of individual stocks decomposed into two samples.

	Unconditional test					Conditional test				
	GAS tv- c_t	GAS roll- c_t	GAS fix- c	GAS sep	HEAVY	GAS tv- c_t	GAS roll- c_t	GAS fix- c	GAS sep	HEAVY
Panel A: Out-of-sample period: 2005–2011										
Panel A.1: 99% VaR										
# p -val < 0.10	16	23	54	40	27	16	18	44	24	18
# p -val < 0.05	8	12	42	20	17	12	13	39	20	13
# p -val < 0.01	3	4	31	10	9	3	4	26	8	7
Panel A.2: 95% VaR										
# p -val < 0.10	19	33	64	49	36	18	29	61	43	25
# p -val < 0.05	11	23	58	40	25	10	23	54	38	18
# p -val < 0.01	4	12	49	27	17	4	9	46	23	13
Panel B: Out-of-sample period: 2012–2018										
Panel B.1: 99% VaR										
# p -val < 0.10	30	23	31	17	25	43	34	41	27	32
# p -val < 0.05	22	13	22	10	18	32	25	32	20	25
# p -val < 0.01	6	4	12	4	10	10	12	16	7	9
Panel B.2: 95% VaR										
# p -val < 0.10	15	17	33	23	21	34	42	55	46	29
# p -val < 0.05	6	11	23	15	15	25	29	40	33	25
# p -val < 0.01	5	6	14	6	9	8	11	17	13	10

This table reports results on one-step-ahead 99% and 95% Value-at-Risk (VaR) predictions, using the GAS tv- c_t , GAS roll- c_t , GAS fix- c , GAS sep, and HEAVY t models, applied to daily returns of 93 stocks listed on the S&P 500 index. We use a moving window of 1000 observations. Panel A lists results on backtesting the VaR on the out-of-sample period from 2005–2011. Panel B shows backtesting results on the out-of-sample period from 2012–2018. Both panels show results of the unconditional and conditional coverage tests of Christoffersen (1998). For both tests, we list the number of cases where the p -value falls below 10%, 5%, or 1%. Bold numbers represent the minimum across all models in the same row.

Fig. 3 shows the time-series plots of $\hat{c}_t$ for the GAS roll- c_t and GAS tv- c_t models for three example stocks: BAC, EQR, and CPB. The fourth panel contains the average across all 93 stocks. The panels for the individual stocks strongly indicate that there is a time-varying scaling factor. The GAS roll- c_t succeeds in capturing the secular trend of this movement. However, the estimates from

the GAS tv- c_t model show that there is even more time variation in c_t . This difference is especially clear during turbulent times. During these times, more information appears to come after trading hours, resulting in a relatively higher overnight variance. This confirms our results from Panels A1 and A2 in Table 5. The last panel in the figure shows that, even when averaging the rolling c_t across all

Fig. 3. Estimated paths of the volatility ratio c_t for the GAS roll- c_t and GAS tv- c_t model. This figure shows the time-series of the estimated volatility ratio $\hat{c}_t$ of the GAS roll- c_t model for the stocks BAC, EQR, and CPB, using a moving window of 1000 observations updated every 50 observations. In addition, we show the forecasted ratio $\hat{c}_t$ of the GAS tv- c_t model of the same stocks. The lower-right panel shows the average rolling c_t ratio across all 93 stocks of our sample, plus a 95% confidence bound. The out-of-sample covers January 13, 2005 until December 31, 2018.

93 stocks, the *cross-sectional average* volatility ratio still shows variation between 1.20 and 1.40.

We conclude that a model with a time-varying c_t produces the best overall VaR forecasts. By splitting the sample into crisis and tranquil periods, we found that the better overall performance is due particularly to turbulent periods. It is precisely here that the dynamic nature of the model matters most. In tranquil periods, the model still performs well less far out in the tails (95%), whereas deeper into the tails in tranquil periods, the GAS sep performs well.

In terms of the ES, Table 4 indicates that the GAS tv- c_t performs best when using unconditional coverage. For conditional coverage, the HEAVY model behaves better than the others. The HEAVY model is clearly inferior to the GAS tv- c_t , however, in terms of unconditional coverage, whereas the difference between the two models appears much smaller in terms of conditional coverage. The model with separate components (GAS sep) again performs well for the 97.5% ES, but not for the 95% ES.

Table 6 again splits the results into turbulent and tranquil periods. In turbulent times, the GAS tv- c model is again clearly superior to all its competitors in terms of unconditional coverage. In terms of conditional coverage, the HEAVY wins by a slight margin, but GAS models with time-varying c_t follow closely. Taking both tests together, the GAS model with a time-varying c_t appears to produce the best ES forecasts during crisis periods. This matches our earlier result for VaR forecasts.

During tranquil periods, the results are more mixed, and the models perform similarly. There is again one exception: for the deeper tail (97.5%) in tranquil periods, considering the ES by modeling separate daytime and overnight volatility processes appears to work best. Again, this is in line with our earlier results on VaR.

We conclude that, for ES, it is also worthwhile splitting the analysis across turbulent and tranquil periods. GAS models with frequently updated volatility ratios perform best during turbulent times, when good volatility forecasts are arguably most important. During tranquil periods, other models, such as the GAS sep and HEAVY models, also perform well or even better, though the performance of most models is similar.

4.4.2. VaR and ES results for stock indices

We now turn to the results for stock indices. Table 7 shows the results.

There are some important similarities with the previous results for individual stocks. We again see that the fixed c model performs the worst in terms of VaR forecasts. A time-varying volatility ratio thus also appears worthwhile for index returns and downside risk forecasting. For indices, the time-varying volatility ratio model with infrequent updating (GAS roll- c_t) generally performs the best, although it is sometimes slightly outperformed by the GAS tv- c_t model or the HEAVY model. Unlike with individual stocks, however, the discrepancies between all the different models are generally very small for index returns. This appears to be even more so for the ES analysis.

Table 6
One-step ahead ES predictions of stock returns decomposed into two samples.

	Unconditional test					Conditional test				
	GAS tv- c_t	GAS roll- c_t	GAS fix- c	GAS sep	HEAVY	GAS tv- c_t	GAS roll- c_t	GAS fix- c	GAS sep	HEAVY
Panel A: Out-of-sample period: 2005–2011										
Panel A.1: 97.5% ES										
# p -val < 0.10	12	22	45	35	27	32	37	35	40	29
# p -val < 0.05	8	18	37	26	20	23	33	26	33	18
# p -val < 0.01	3	8	23	14	12	20	22	19	26	13
Panel A.2: 95% ES										
# p -val < 0.10	12	28	46	41	30	36	45	43	47	36
# p -val < 0.05	9	24	34	32	24	26	36	34	41	22
# p -val < 0.01	2	11	22	20	12	14	23	25	33	15
Panel B: Out-of-sample period: 2012–2018										
Panel B.1: 97.5% ES										
# p -val < 0.10	24	23	25	10	23	31	35	35	26	19
# p -val < 0.05	18	14	19	7	16	24	27	27	17	15
# p -val < 0.01	6	6	11	2	7	19	21	22	13	8
Panel B.2: 95% ES										
# p -val < 0.10	9	10	18	11	10	29	31	32	23	22
# p -val < 0.05	7	8	14	7	8	22	29	27	18	20
# p -val < 0.01	5	3	6	4	5	17	21	21	18	11

This table reports results one-step-ahead 97.5% and 95% Expected Shortfall (ES) predictions, using the GAS tv- c_t , GAS roll- c_t , GAS fix- c , GAS sep, and HEAVY t models, applied to daily returns of 93 stocks listed on the S&P 500 index. We use a moving window of 1000 observations. Panel A lists results on backtesting the ES on the out-of-sample period from 2005–2011. Panel B shows backtesting results on the out-of-sample period from 2012–2018. Both panels show results of the unconditional and conditional backtests of Du and Escanciano (2016). For both tests, we list the number of cases where the p -value falls below 10%, 5%, or 1%. Bold numbers represent the minimum across all models in the same row.

Table 7
One-step-ahead VaR and ES predictions (stock indices).

	Unconditional test					Conditional test				
	GAS tv- c_t	GAS roll- c_t	GAS fix- c	GAS sep	HEAVY	GAS tv- c_t	GAS roll- c_t	GAS fix- c	GAS sep	HEAVY
Panel A.1: 99% VaR										
# p -val < 0.10	18	16	19	19	17	16	14	19	19	17
# p -val < 0.05	17	14	19	19	17	15	14	19	19	16
# p -val < 0.01	13	12	18	17	15	11	12	15	16	11
Panel A.2: 95% VaR										
# p -val < 0.10	15	13	19	19	14	14	15	19	19	16
# p -val < 0.05	13	13	19	19	14	14	14	19	19	15
# p -val < 0.01	10	5	19	19	9	10	8	19	18	10
Panel B.1: 97.5% ES										
# p -val < 0.10	17	16	19	18	17	10	12	10	11	8
# p -val < 0.05	17	16	18	18	17	10	10	9	8	8
# p -val < 0.01	13	14	18	18	15	8	8	8	7	7
Panel B.2: 95% ES										
# p -val < 0.10	17	16	19	18	16	8	10	11	9	8
# p -val < 0.05	17	15	19	18	16	6	10	9	8	8
# p -val < 0.01	13	11	19	17	12	5	8	6	7	6

This table reports the results on one-step-ahead 99% and 95% Value-at-Risk (VaR) and 97.5% and 95% Expected Shortfall (ES) predictions, using the GAS tv- c_t , GAS roll- c_t , GAS fix- c , GAS sep, and HEAVY t models, applied to 19 daily stock indices. We use a moving window of 1000 observations. Panel A lists results on backtesting the VaR. Specifically, sub-panels A.1 and A.2 show the results of the unconditional and conditional coverage tests of Christoffersen (1998). Panel B shows results of the unconditional and conditional backtests of Du and Escanciano (2016). For all tests, we list the number of cases where the p -value falls below 10%, 5%, or 1%. Bold numbers represent the minimum across all models in the same row.

In terms of the unconditional test, however, models with a fixed c and with separate volatility processes reject for (almost) all indices at all levels of significance.

We also tested whether these results for stock indices were robust to splitting the sample into a turbulent (2005–2011) and a more tranquil (2012–2018) period. Indeed, the GAS roll- c_t , GAS tv- c_t , and HEAVY models all show good performance for index returns during both

turbulent and tranquil times. Margins between the three models are generally small. Unlike in the case of individual stocks, however, there is now no setting where it makes sense from a forecasting perspective to model the daytime and overnight volatility processes as separate components, even in the deep tails during tranquil times. We conclude that modeling a time-varying volatility ratio is also helpful when forecasting downside risk measures

for index returns, closely followed by the HEAVY model. The performance of models with a fixed volatility ratio or separate components lags behind.

5. Conclusions

We introduced a new dynamic score driven model for the ratio of total close-to-close volatility and daytime volatility. The model differs from the existing literature by considering a time-varying ratio instead of a fixed ratio, as for instance Ahoniemi et al. (2015) and Hansen and Lunde (2005). Our model for a time-varying scaling factor takes into account the fat-tailedness of the realized variance using the (univariate analogue of the) GAS F model of Opschoor et al. (2018).

To test the model, we used daily returns and realized variances over the period 2001–2018 for a random sample of 93 stocks from the S&P500 across different industries, as well as for 19 stock indices across the world. We found that the new model improved in-sample predictions for almost all individual stocks and about half of the stock indices. The ratio of total to daytime volatility clearly varied over time.

Out-of-sample, one-step-ahead VaR and ES predictions of individual stocks also improved from using the specification with a score-driven volatility ratio (frequent updating), or from using a specification with rolling window updating (infrequent updating). The time-varying volatility ratio proved particularly useful in turbulent times, when relatively more information may be released between trading hours. The results confirm that a fixed volatility ratio over the whole out-of-sample period results in poor VaR and ES predictions. Interestingly, during quiet times and farther out into the tails, it sometimes appeared to be a good approach to model daytime and overnight volatility as separate processes, as suggested in previous studies. However, our results show that this is rather incidental and does not carry over to the index return setting nor to regions less far out into the tail.

When considering the index return setting, most of our results survived. However, for index returns, all differences were considerably smaller than for individual stocks. In this setting, the score-driven models with a time-varying volatility ratio and the HEAVY model all behaved roughly at par, in both moderate and deep tail regions, and in both turbulent and tranquil times. The models did, however, beat models with a fixed volatility ratio or with separate modeling components for daytime and overnight volatility.

References

- Ahoniemi, K., Fuertes, A. M., & Olmo, J. (2015). Overnight news and daily equity trading risk limits. *Journal of Financial Econometrics*, 14, 525–551.
- Ahoniemi, K., & Lanne, M. (2013). Overnight stock returns and realized volatility. *International Journal of Forecasting*, 29, 592–604.
- Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 885–905.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71, 529–626.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal*, 12, 1–32.
- Berkman, H., Koch, P. D., Tuttle, L., & Zhang, Y. J. (2012). Paying attention: overnight returns and the hidden cost of buying at the open. *Journal of Financial and Quantitative Analysis*, 47, 715–741.
- Blasques, C., Koopman, S. J., & Lucas, A. (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika*, 102, 325–343.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31, 307–327.
- Brownlees, C. T., & Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis*, 51, 2232–2245.
- Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39, 841–862.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Economics*, 7, 174–196.
- Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28, 777–795.
- Du, Z., & Escanciano, J. C. (2016). Backtesting expected shortfall: accounting for tail risk. *Management Science*, 63, 940–958.
- Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20, 339–350.
- Engle, R. F., & Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131, 3–27.
- Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48, 1779–1801.
- Hansen, P. R., & Lunde, A. (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics*, 3, 525–554.
- Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: With applications to financial and economic time series*. Cambridge University Press.
- Harvey, A. C., & Luati, A. (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, 109, 1112–1122.
- Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, 12, 445–475.
- Linton, O., & Wu, J. (2020). A coupled component DCS-EGARCH model for intraday and overnight volatility. *Journal of Econometrics*.
- Lucas, A., Schwaab, B., & Zhang, X. (2014). Conditional euro area sovereign default risk. *Journal of Business & Economic Statistics*, 32, 271–284.
- Noureldin, D., Shephard, N., & Sheppard, K. (2012). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 27, 907–933.
- Oh, D. H., & Patton, A. J. (2018). Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads. *Journal of Business & Economic Statistics*, 36, 181–195.
- Opschoor, A., Janus, P., Lucas, A., & van Dijk, D. (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, 36, 642–657.
- Shephard, N., & Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*, 25, 197–231.
- Sinclair, E. (2013). *Volatility trading*. John Wiley & Sons.
- Taylor, N. (2007). A note on the importance of overnight information in risk management models. *Journal of Banking & Finance*, 31, 161–180.
- Taylor, S. J. (2008). *Modelling financial time series*. world Scientific.
- Todorova, N., & Souček, M. (2014). Overnight information flow and realized volatility forecasting. *Finance Research Letters*, 11, 420–428.