TI 2019-052/IV
Tinbergen Institute Discussion Paper

Observation-driven Models for Realized Variances and Overnight Returns

Anne Opschoor[1]
André Lucas[1]

[1] Vrije Universiteit Amsterdam

Observation-driven Models for Realized Variances and Overnight Returns*

Anne Opschoor André Lucas

Vrije Universiteit Amsterdam and Tinbergen Institute

This version: July 23, 2019

Abstract

We present a new model to decompose total daily return volatility into a filtered (high-frequency based) open-to-close volatility and a time-varying scaling factor. We use score-driven dynamics based on fat-tailed distributions to limit the impact of incidental large observations. Applying our new model to 100 stocks of the S&P 500 during the period 2001-2014 and evaluating (in-sample and out-of-sample) in terms of Value-at-Risk and Expected Shortfall, we find our model outperforms alternatives like the HEAVY model that uses close-to-close returns and realized variances, and models treating close-to-open en open-to-close returns as separate processes. Results also indicate that the ratio between total and open-to-close volatility changes substantially through time, especially for financial stocks.

Keywords: overnight volatility, realized variance; F distribution; score-driven dynamics

Classification codes: C32, C58.

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

* *Corresponding author*: Anne Opschoor, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV, Amsterdam, The Netherlands. *E-mail*: `a.opschoor@vu.nl`. *Phone*: +31(0)20-5982663.

Volatility (SV) model of Taylor (2008), including extensions such as the threshold GARCH model of Glosten *et al.* (1993), among others.

Due to the availability of High-Frequency (HF) data, volatility measurement, modeling and forecasting has 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 model directly the ex-post realized intraday variance (Engle and Gallo, 2006; Corsi, 2009) or a hybrid version that uses both returns and realized variances (Shephard and Sheppard, 2010).

Although HF data improves the measurement and forecasts of volatility, typically this data is only available for individual stocks during trading hours. The literature has so far typically overcome this issue by simply discarding overnight volatility and hence considering realised measures only and/or using open-to-close returns (Engle and Gallo, 2006; Corsi, 2009; Shephard and Sheppard, 2010; Noureldin *et al.*, 2012, 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 and Souček, 2014).

An alternative way of dealing with overnight volatility is to scale up the realized variance measured during trading hours by a *fixed* constant (Hansen and Lunde, 2005; Koopman *et al.*, 2005; Ahoniemi *et al.*, 2015). 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-co-close volatility. We take into account the availability of high-frequency data during the day. We deviate from earlier literature by scaling the *filtered* volatility instead of the realized variance. Moreover, we allow the scaling factor to vary over time, using a time-varying ratio between the total and open-to-close volatility as proposed by Hansen and Lunde (2005). Scaling the filtered volatility instead of the realized variance can directly result in different total daily volatility measurements, since realized variances are typically fat-tailed (see for example Opschoor *et al.*, 2018). Considering a time-varying ratio of volatilities is also in line with the work of Linton and Wu (2017). They show that daily and overnight volatility are distinct processes with their own dynamics. As a result, the ratio between total volatility and daytime (open-to-close) volatility typically varies through time. For example, during crises periods, 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 risk measures.

The time-variation in volatility levels and the ratio of daytime to total volatility in our model is captured by 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 and Luati, 2014) and systemic risk modeling (Lucas *et al.*, 2014; Oh and 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 in financial econometrics by Opschoor *et al.* (2018). Score-driven dynamics for the F distribution result in an intuitive propagation mechanism for the daytime volatility: large values of the realized variance are downweighted, hence such values do not disrupt the filtered volatility path. Second, we model the ratio between the total volatility and the filtered daytime volatility using the same score-driven dynamics and a fat-tailed distribution. The resulting propagation dynamics for the ratio are again highly intuitive: large ratios of total daily squared returns over filtered close-to-close volatilities are downweighted.

In our empirical application, we use the new model to describe daily returns and realized variances of 100 equities from the S&P 500 index over the period January 2001 to December 2014. We show in-sample that the ratio of total and daytime volatility indeed varies over time. This is particularly the case for financial stocks. For financial stocks we find relatively (small) large ratios (before) during the Global Financial crisis and the sovereign debt crisis. Hence relatively more information accumulates outside regular trading hours during these volatile periods, implying a larger total-to-daytime volatility ratio. For many stocks, our model improves significantly upon the HEAVY model of Shephard and Sheppard (2010) or a score-driven model with a fixed ratio. We also confirm Linton and Wu (2017) by showing that modeling the overnight volatility separately leads to very low values of the degrees of freedom parameter, with subsequent potential problems for the existence of sufficient moments (see also Berkman *et al.*, 2012). It indicates that overnight returns are considerably heavy-tailed and exhibit rather different statistical properties than open-to-close returns.

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 backtest of Du and Escanciano (2016) to backtest the Expected Shortfall predictions, while the VaR has been backtested by the (un)conditional coverage test of Christoffersen (1998). The results indicate that overall our new proposed model produces the best VaR and ES predictions. Especially the approach of modeling daily and overnight volatility as separate processes performs rather poor. This contrasts with the finding of Ahoniemi *et al.* (2015), although the difference could be explained by the fact that we use individual stock returns while Ahoniemi *et al.* (2015) focus on stock *indices*.

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 of daily stock returns. In Section 3, we briefly review the computation of the Value-at-Risk and Expected Shortfall, as well as recently developed backtesting procedures. Section 4 provides an overview of the data, our in-sample, and out-of-sample results for the empirical application to 100 stocks from the S&P500. Section 5 concludes.

2 The modeling framework

Let $r_t \in \mathbb{R}$ be the total daily close-to-close (ctc) return of a financial asset on day t, $t = 1, \ldots, T$, and $h_t = \mathrm{Var}(r_t|\mathcal{F}_{t-1})$ 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

$$
\begin{aligned}
r_t &= r_{o,t} + r_{d,t}, && (1)\\
h_t &= h_{o,t} + h_{d,t} + 2\,\mathrm{Cov}(r_{o,t}, r_{d,t}) = h_{o,t} + h_{d,t} + 2\rho_{o,d,t}\sqrt{h_{o,t}}\sqrt{h_{d,t}}, && (2)
\end{aligned}
$$

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 during trading-hours of high-frequency data, we can obtain accurate measurements of the daytime variance $h_{d,t}$ by for example computing the realized variance RV_t (see Andersen and Bollerslev, 1998), defined as the sum of, e.g., 5-minute intra-day returns over the course of the trading day (9:30 - 16:00). Typically, there are no high frequency data outside trading hours 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) for risk-management purposes, such as predicting the Value-at-Risk (VaR) or Expected Shortfall (ES). This can be achieved in two ways. First, one can model h_t directly based on a time series model that possibly includes (a scaled version of) RV_t. Second, one can model the daytime and overnight variances (and their covariance) separately. Ahoniemi *et al.* (2015) compare both ways in the context of equity *indexes*. They scale RV_t to a daily close-to-close variance level $RV_t^{sc,1}$ 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,

$$RV_t^{sc,1} = \omega_1 r_{o,t}^2 + \omega_2 RV_t, \tag{3}$$

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 $\mu_0 = \mathbb{E}[RV_t + r_{o,t}^2]$. This scaled realized variance $RV_t^{sc,1}$ is then used into a time series model as a second step.

For the new model put forward in our current paper, it is useful to recall a second scaling method put forward by Hansen and Lunde (2005):

$$RV_t^{sc,2} = c\,RV_t = \frac{\sum_{t=1}^T r_t^2}{\sum_{t=1}^T RV_t} RV_t, \tag{4}$$

which basically 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 two scaling methods and several other methods. They conclude that for individual stocks it is better to drop the overnight information, while for the index, the scaling method of Hansen and Lunde (2005) based on optimal weights performs superior.

This paper deviates from previous approaches in two ways. First, we focus on scaling a *filtered* volatility instead of RV_t, where the filter is based on score-driven dynamics that account for the heavy-tailed nature of the realized variances. Fat-tailedness can result in the scaling factor c in (4) becomeing ill-behaved, for instance, if fourth order moments of overnight returns fail to exist. The latter is a real concern, as overnight returns for individual stocks are considerably fat-tailed. Our method is robust against such features by assuming fat-tailed distributions for both the returns and the realized variances. Second, we note that Linton and Wu (2017), among others, show that the daytime and overnight volatility processes have their own dynamics and that, hence, the ratio c might *not* be constant over time. We accommodate this second point by making the ratio c time-varying using the score-driven (GAS) dynamics of Creal *et al.* (2013); see also Harvey (2013).

Our new model reads

$$RV_t = h_{d,t} u_{d,t}, \qquad u_{d,t} \sim F(1, \nu_1, \nu_2), \tag{5}$$

$$r_t = \mu + \sqrt{h_t}\varepsilon_t, \qquad \varepsilon_t \sim t(0, 1, \nu), \qquad h_t = c_t\, h_{d,t}, \tag{6}$$

$$h_{d,t+1} = \omega_1 + \alpha_1\, s_{h,t} + \beta_1\, h_{d,t}, \tag{7}$$

$$c_{t+1} = \omega_2 + \alpha_2\, s_{c,t} + \beta_2\, c_t, \tag{8}$$

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.,

$$\begin{aligned} s_{h,t} &= S_{h,t} \nabla_{h,t} = S_{h,t} \frac{\partial \ln p_{RV}(RV_t | h_{d,t}, \mathcal{F}_{t-1}; \nu_1, \nu_2)}{\partial h_{d,t}} = \frac{\nu_1}{\nu_1 + 1} \left(\frac{\frac{\nu_1+\nu_2}{\nu_2-2} RV_t}{1 + \frac{\nu_1}{\nu_2-2} \frac{RV_t}{h_{d,t}}} - h_{d,t} \right), && (9) \\ s_{c,t} &= S_{c,t} \nabla_{c,t} = \frac{\partial \ln p_r(r_t | h_{d,t}, c_t, \mathcal{F}_{t-1}; \mu, \nu)}{\partial c_t} = w_{c,t} \frac{\varepsilon_t^2}{h_{d,t}} - c_t, && (10) \\ w_{c,t} &= \frac{\nu + 1}{\nu - 2 + \frac{\varepsilon_t^2}{h_t}}. \end{aligned}$$

The fat-tailed $F(\nu_1, \nu_2)$ and Student's $t(\nu_1)$ distributions in (5) and (6) 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 Opschoor *et al.* (2018) and Creal *et al.* (2013) and scale $\nabla_{h,t}$ and $\nabla_{c,t}$ by $2h_{d,t}^2/(\nu_1 + 1)$ and $\mathcal{I}^{-1}$, respectively, where $\mathcal{I}$ denotes the Information matrix. We label our model as the 'GAS tv-c' model, indicating we have a model with time-varying c. As a special case for $\alpha_2 = \beta_2 = 0$, we obtain a GAS model with a fixed c, denoted by 'GAS fix-c'.

Both scores in (9)–(10) 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 implies a low weight in (9) 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 (10).

We estimate our GAS tv-c model by the method of Maximum Likelihood. Although the model parameters can be estimated in one step, we opt for a 2-step approach. In a first step, we estimate the parameters associated with the daily (open-to-close) volatility. Let $\boldsymbol{\theta}_D = \{\nu_1, \nu_2, \omega_1, \alpha_1, \beta_1\}$ and $\boldsymbol{\theta}_{tot} = \{\mu, \nu, \omega_2, \alpha_2, \beta_2\}$ denote the vector containing all static parameters of the daily process and the total volatility process, respectively. We first maximize the log-likelihood associated with the F distribution:

$$\max_{\boldsymbol{\theta}_D} \mathcal{L}^F = \max_{\boldsymbol{\theta}_D} \sum_{t=1}^{T} \log p_{RV}(RV_t | h_{d,t}, \mathcal{F}_{t-1}, \boldsymbol{\theta}_D). \tag{11}$$

Given $\hat{\boldsymbol{\theta}}_D$ 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.

$$\max_{\boldsymbol{\theta}_{tot}} \mathcal{L}^{Stud} = \sum_{t=1}^{T} \log p_r(r_t | \hat{h}_{d,t}, c_t, \boldsymbol{\theta}_{tot}). \tag{12}$$

The efficiency loss due to two-step estimation procedure is very small.

3 Forecasting Value-at-Risk and Expected Shortfall

We assess the economic value of accounting for a (time-varying) ratio of total over daytime filtered volatilities by considering Value-at-Risk and Expected Shortfall predictions. The former has been very popular in the financial industry, while the latter is gaining momentum due to the reregulations in Basel IV following the 2008 financial crisis.

The 1-step ahead q-level VaR is defined as the quantity VaR(1-q) such that

$$Pr[r_t < VaR_t(1-q) | \mathcal{F}_{t-1}] = q; \tag{13}$$

In line with the Basel accords, we consider $q = 0.01$ which corresponds to a 99% VaR. In addition, we set q equal 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 (which differs per model), the 1-step conditional $VaR(q)$ equals

$$VaR_t(1-q) = \mu + t^{-1}(q,\nu)\sqrt{h_t}, \tag{14}$$

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

The VaR is the q-quantile of the conditional return distribution but does not consider any information about the magnitude of the losses beyond this quantile. The latter are summarized by the Expected Shortfall (ES). The 1-step ahead ES is defined as

$$ES_t(1-q) = \mathbb{E}[r_t|\mathcal{F}_{t-1}, r_t < VaR_t(1-q)]. \tag{15}$$

Put differently, the ES integrates the VaR over a continuum of levels and can also be defined as

$$ES_t(1-q) = \frac{1}{q}\int_0^q VaR_t(q)\,\mathrm{d}q. \tag{16}$$

Under our Student's t distributional assumption for r_t, we have

$$ES_t(1-q) = \mu + \frac{1}{q}p_r(x_q|\nu)\left(\frac{(\nu-2)+x_q^2}{\nu-1}\right)\sqrt{h_t}, \tag{17}$$

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

Given the 1-step ahead VaR and ES over the out-of-sample forecasting period, we use several statistical tests to assess the adequacy of our forecasts. Define the hit-process

$$h_t(q) = 1_{[r_t < VaR_t(1-q)]}, \tag{18}$$

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 having an asymptotic $\chi^2(1)$ distribution. The first test corresponds to null $H_0 : \mathbb{E}[h_t(q)] = Pr[h_t(q) = 1] = q$, while the latter tests the null hypothesis $H_0 : Pr[h_{t+1}(q) = 1|h_t(q)] = \Pr[h_{t+1}(q)]$. Both tests can also be combined in a conditional coverage (CC) test that is asymptotically $\chi^2(2)$ distributed.

For Expected Shortfall, we use the recently developed test of Du and Escanciano (2016). Given (16), consider the cumulative violation process

$$H_t(q) = \frac{1}{q}\int_0^q h_t(u)du \tag{19}$$

which equals

$$H_t(q) = \frac{1}{q}(q - z_t)1_{z_t \leq q}, \tag{20}$$

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 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

$$t_{ES} = \frac{\overline{H(q)} - q/2}{\sqrt{v_{ES}(q)/P}}, \tag{21}$$

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

Table 1: S&P 500 constituents

This table lists ticker symbols of 100 companies listed at the S&P 500 index during the period January 2, 2001 until December 31, 2014. All Tickers are grouped per industry.

Ind Nr.	Industry	# Comp.	Tickers
1	Capital Goods	10	AA,BA,CAT,HON,F,NOC,UTX, A,IR,GD
2	Finance	19	AXP,JPM,AIG,BAC,C,KEY,MTB, COF,USB,BBT,STI,WFC,GS,MS, MMC,HIG,PNC,XL,MCO
3	Energy	12	GE,XOM,BHI,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	9	KO,MO,SYY,PEP,CL,AVP,GIS, CPB,EL
6	Health Care	11	PFE,ABT,BAX,JNJ,LLY,THC,MMM, MRK,BMY,MDT,CI
7	Public Utilities	7	AEP,AEE,DUK,SO,WMB,VZ,EXC
8	Technology	5	IBM,DOV,HPQ,TSM,CSC
9	Basic Industries	9	PG,DD,FLR,DOW,AES,IP,ATI, LPX,POT
10	Transportation	4	LUV,UPS,NSC,FDX

4 Empirical application

4.1 Data

The data consist of daily open-to-close returns and daily realized variances for 100 S&P 500 constituents. The 100 stocks are randomly chosen from different industries, such as financials, materials etc. Table 1 provides an overview of the data. The data span the period January 2, 2001 until December 31, 2014. We have $T = 3521$ trading days. The Financial industry covers most companies (i.e. 19), followed by Consumer Services and Energy, respectively.

We retrieve consolidated trades (transaction prices) from the Trade and Quote (TAQ) database from 9:30 until 16:00 with a time-stamp precision of one second. 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 5-minute returns.

4.2 Competing models

We benchmark our model against two alternatives. The first model is the HEAVY model of Shephard and Sheppard (2010), which considers the total daily return r_t, and which uses the lagged realized variance as an explanatory variable for the variance h_t. This model is defined as

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

The HEAVY model does not take into account the possible difference in the overnight and daily volatility processes.

Our second benchmark models the three components of (2) separately; compare Ahoniemi *et al.* (2015). Similar as in our own score-driven GAS model, the first component $h_{d,t}$ is modeled by the GAS F model (7). 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:

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

where $\varepsilon_{o,t}$ denote the demeaned overnight return at time t. We follow Ahoniemi *et al.* (2015) and incorporate the squared daily return $\varepsilon^2_{D,t}$ 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 and 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. In our case, where we use individual stocks instead of a stock index, we do not find any empirical evidence for the value-added of a DCC model during this step. We label this second benchmark model as 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, GAS fix-c, GAS F, GAS t ON, and HEAVY model by Maximum Likelihood using all observations. Table 2 and Figure 1 present the results.

Table 2 shows three interesting findings. First, the average degrees of freedom parameter ν of the Student's t distribution across all 100 assests considered is around 7, indicating that the conditional returns are fat-tailed. This becomes extreme for the overnight returns, as indicated by an average value of around 3 for the GAS t ON model. This result confirms for example Linton and Wu (2017) and shows that overnight returns are considerably fat-tailed and fourth order moments may not exist. Second, the average persistence parameter β of the GAS tv-c model is rather low (0.565). This seems due, however, to some stocks that do not exhibit a time-varying ratio c_t of total to daily volatility. For instance, if we only consider the 19 financial stocks, the average β jumps upward to 0.862, suggesting that for financial stocks there is persistent time-variation in the ratio c_t.

Figure 1 shows the difference in maximized log-likelihood between the GAS tv-c and the HEAVY t model (upper panels), and between the GAS tv-c and GAS fix-c model (lower panels). A positive value means that the GAS tv-c model has a higher log-likelihood. The likelihood of the GAS tv-c model is higher for quite a number of stocks compared to the HEAVY t model. This is true in particular for the financial stocks, as indicated in the upper right panel. In addition, the lower panel show that modeling c_t as time-varying rather than static also improves the likelihood. In particular, for most financial stocks, twice the difference in the log-likelihood is substantial, indicating that a likelihood ratio test for the null of a static c will be strongly rejected.

Figure 2 shows the fitted process c_t of the GAS tv-c model for four financials, AIG, BAC, C, and WFC, along with the estimated c of the GAS fix-c model. The figure shows that indeed the ratio varies through time, with huge peaks in the heat of the Global Fi-

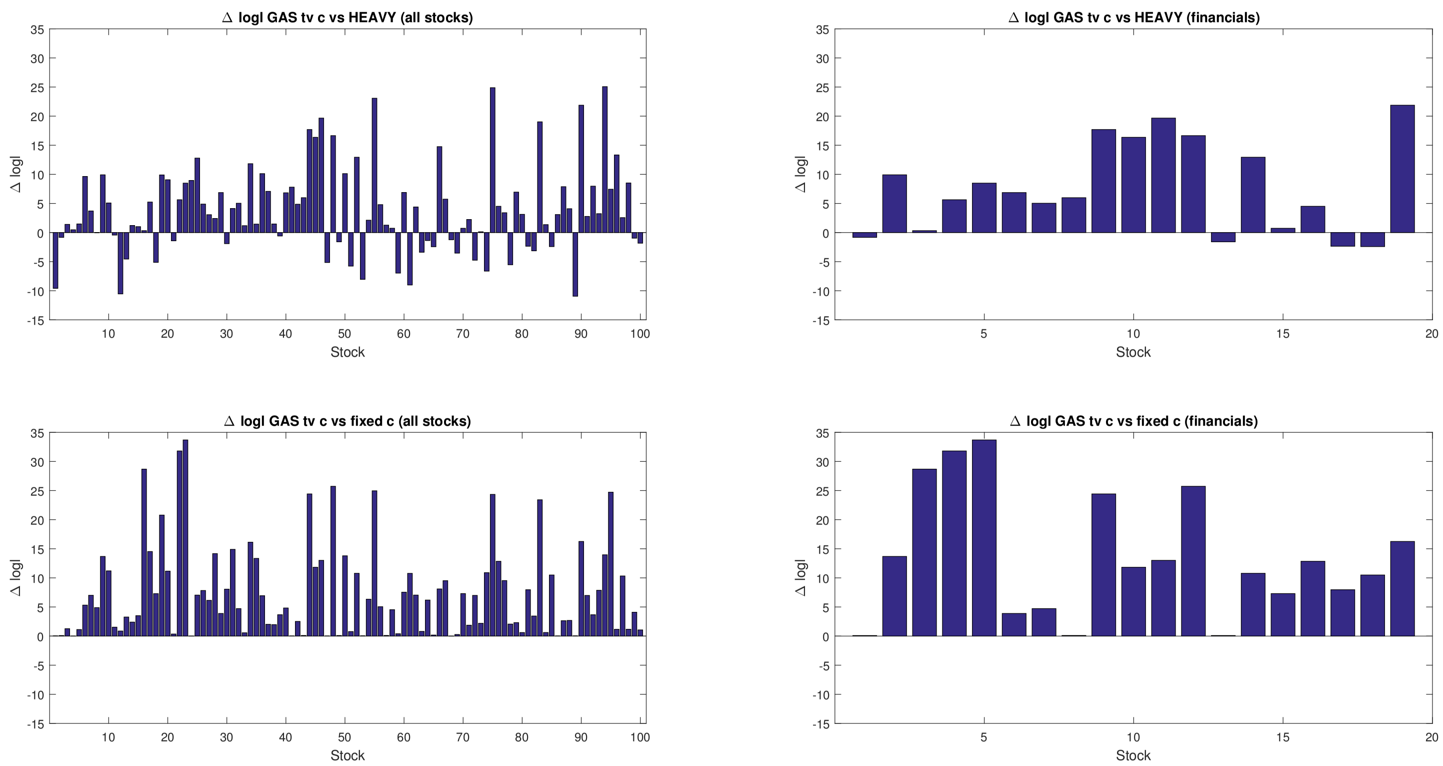

Figure 1: Log-likelihood differences vis-a-vis the GAS tv c model
This figure shows the difference in the maximized log-likelihood between the GAS tv-c model and the HEAVY t model (upper panel) and GAS tv-c model and the GAS fix-c model (lower panel). The left figures depict the case for 100 listed at the S&P 500 index, while the right figures show the differences for 19 financial companies, listed in Table 1. A positive number means a better fit of the GAS model with time-varying ratio of total and daily volatility.

Table 2: Full sample average parameter estimates of all stocks

This table reports maximum likelihood parameter estimates of the HEAVY t, GAS F, GAS t ON, GAS tv-c and GAS fix-c models, applied to daily equity returns, realized variances or overnight returns of 100 assets listed at the S&P 500 index. We list the mean of the estimated parameters over all 100 stocks, with the standard deviation (SD) in parentheses. In case of the GAS tv-c model, we list the mean and standard deviation for all stocks, as well as for the 19 Financial companies. The sample covers January 2, 2001 until December 31, 2014 (3521 observations).

			μ	ω	α	β	ν	ν_1	ν_2
GAS t tv-c	All stocks	mean	0.033	0.560	0.038	0.565	7.022		
		sd	(0.02)	(0.50)	(0.04)	(0.38)	(1.75)		
	Financials	mean	0.023	0.180	0.024	0.862	7.156		
		sd	(0.02)	(0.31)	(0.03)	(0.25)	(1.38)		
GAS t fix-c	All stocks	mean	0.031	1.306			6.862		
		sd	(0.02)	(0.16)			(1.71)		
GAS t ON	All stocks	mean	0.030	0.011	0.116	0.992	3.214		
		sd	(0.04)	(0.01)	(0.02)	(0.01)	(0.67)		
GAS F	All stocks	mean		0.053	0.877	0.983		21.832	14.711
		sd		(0.03)	(0.10)	(0.01)		(3.61)	(2.59)
HEAVY t	All stocks	mean	0.031	0.082	0.489	0.597	7.017		
		sd	(0.02)	(0.09)	(0.15)	(0.14)	(1.76)		

nancial Crisis (2008/2009) and the Sovereign Debt crisis of 2012. Interestingly, the value of c become lower than one during the years 2005/2006, hence correcting the fitted daily volatility downwards. Hence the total information outside trading hours with respect to the information during the day varies heavily through time. We conclude that statistically, it pays off for a large number of stocks to model open-to-close and close-to-close volatility by two processes using a time-varying scaling factor.

Comparing the GAS sep model with the GAS tv-c model in terms of likelihoods is not straightforward: the GAS tv-c model describes daily returns and daytime realized volatilities, while the GAS sep model also models the overnight returns separately and thus has an incomparable likelihood. To compare the two models, Figure 3 plots the difference of the filtered log-volatilities for the same four financial companies shown earlier. The figure shows an interesting pattern that is common among all four financial assets: during the tranquil years 2003-2006 (except 2005 for AIG), the difference in (log) volatility is negative, while during more turmoil period such as 2008/2009 and 2011, the difference is positive. Hence

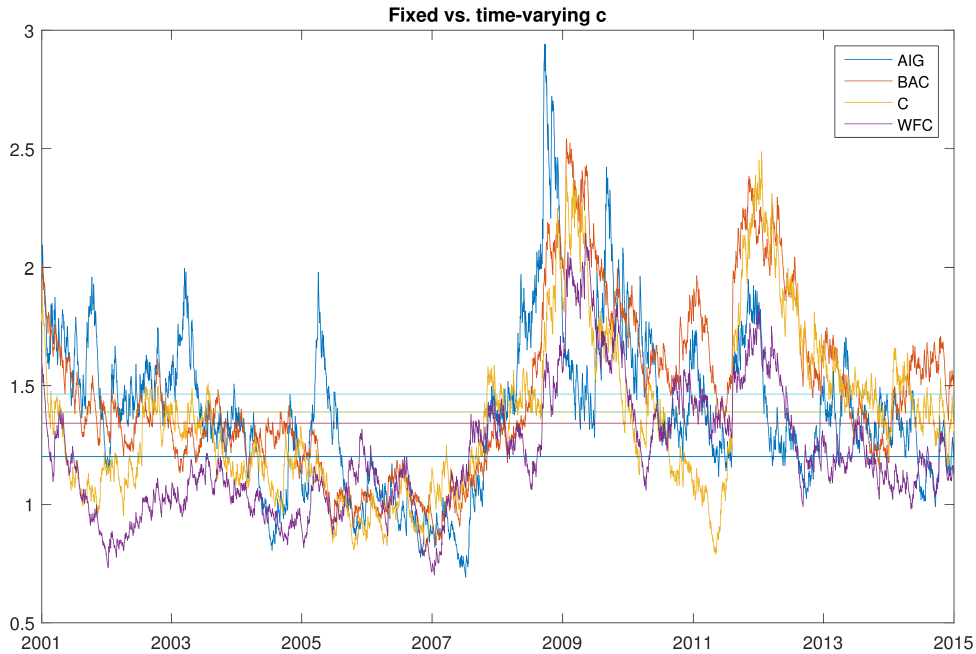

Figure 2: Time-varying volatility ratios
This figure shows the fitted ratio of the total daily volatility and the open-to-close volatility from the GAS tv-c model for AIG, BAC, WFC and C. In addition, the horizontal lines represents the estimated $\hat{c}$ from the GAS fix-c model for four financial companies, which equals 1.39, 1.47, 1.34 and 1.20 respectively. The sample covers January 2, 2001 until December 31, 2014 (3521 observations).

modeling overnight and daily returns separately results in a somewhat higher volatility during calm periods, while during turmoil periods the volatility seems lower. This suggests that modeling the two components separately smooths out some of the time-variation found with our new time-varying ratio c_t model. In the next section we investigate whether these differences matter when forecasting a 1-step ahead VaR and ES.

4.4 Out-of-sample results

To assess the short-term forecasting performance of the different models, we consider 1-step ahead 99% and 95% Value-at-Risk and 97.5% and 95% Expected Shortfall predictions. We use a moving-window of 1000 observations, starting 2001–2004, such that the Global Financial Crisis and the European sovereign debt crisis are part of the out-of-sample period. We re-estimate our model every 50 observations, or about every two months.

Panels A.1 and A.2 of Table 3 list the results for the 1-step ahead VaR forecasts based

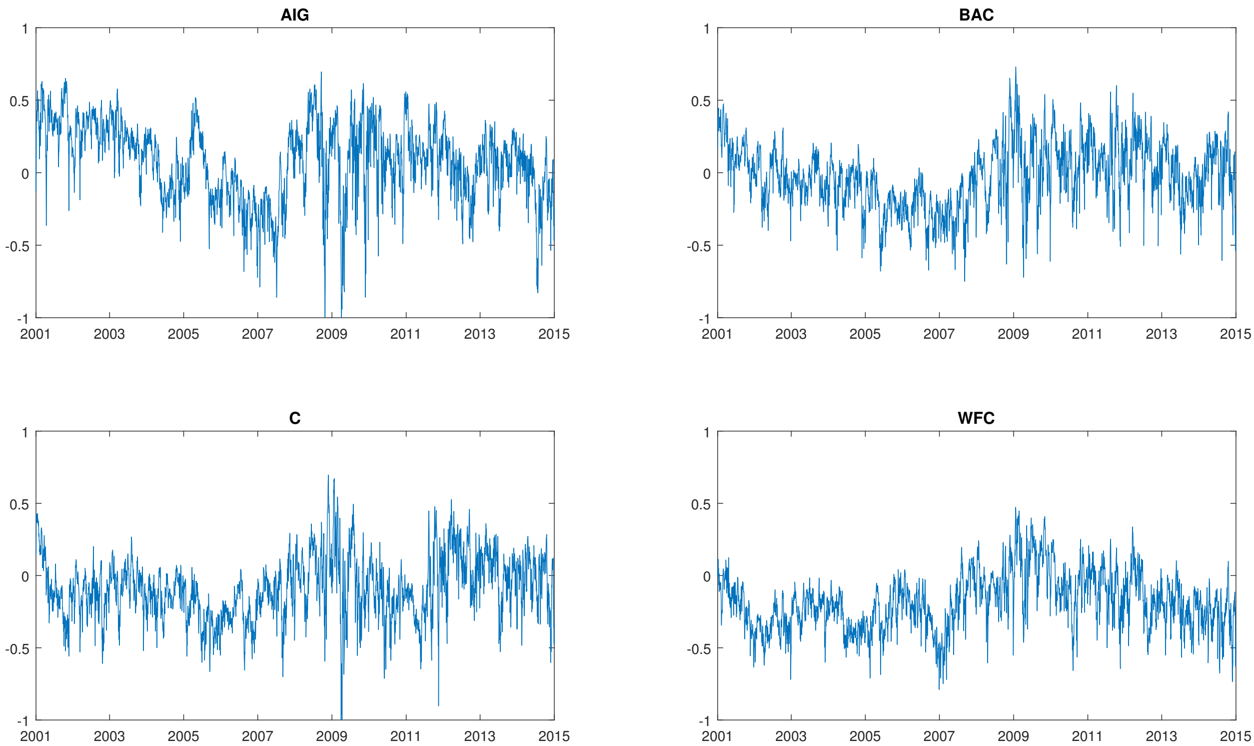

Figure 3: Fitted total daily log-volatilities

This figure shows the differences in fitted log-volatilities of AIG, BAC, C and WFC according to the GAS tv-c model and the GAS sep model. A negative number means that the fitted volatilities of the former model is lower than the volatilities of the latter model. The sample covers January 2, 2001 until December 31, 2014 (3521 observations).

Table 3: 1-step ahead VaR and ES predictions

This table reports results on 1-step ahead 99% and 95% Value-at-Risk (VaR) and 97.5% and 95% Expected-Shortfall (ES) predictions, using the GAS tv-c, GAS fix-c, Gas sep and HEAVY t models, applied to daily equity returns of 100 assets listed at the S&P 500 index. We use an moving window of 1000 observations. Panel A lists results on backtesting the VaR. More specifically, Panel A.1 and A.2 show results of the unconditional and conditional coverage tests of Christoffersen (1998). Panel B shows results of the unconditional backtest of Du and Escanciano (2016). For all tests we list the number of cases the p-value falls below 10, 5 or 1 %. The out-of-sample covers December 20, 2004 until December 31 2014 and contains 2521 observations.

	GAS tv-c	GAS fix-c	GAS sep	HEAVY	GAS tv-c	GAS fix-c	GAS sep	HEAVY
		99% VaR				95% VaR		
Panel A.1: Unconditional Coverage test								
$\sharp$ p-val < 0.10	15	14	27	26	17	19	47	27
$\sharp$ p-val < 0.05	9	7	19	18	12	13	38	20
$\sharp$ p-val < 0.01	5	1	5	7	2	5	24	9
Panel A.2: Conditional Coverage test								
$\sharp$ p-val < 0.10	16	15	21	23	17	28	48	25
$\sharp$ p-val < 0.05	11	9	14	16	9	18	35	17
$\sharp$ p-val < 0.01	3	3	5	7	2	4	21	6
		97.5% ES				95% ES		
Panel B: Du & Escanciano backtest (Unconditional)								
$\sharp$ p-val < 0.10	15	16	36	25	17	16	47	21
$\sharp$ p-val < 0.05	9	7	26	17	9	10	38	19
$\sharp$ p-val < 0.01	4	2	10	7	2	3	27	7

on the unconditional (UC) and conditional coverage (CC) test of Christoffersen (1998). For each test, we report the number rejections (across all 100 stocks) of the null-hypothesis using a 10%, 5%, or 1% level. The main result is that our new model with time-varying c_t clearly outperforms both the HEAVY t model (which ignores the differences between daytime and overnight volatility), and the GAS sep model (which models daytime and overnight returns as separate processes). The number of violations for the UC and CC test are considerably lower in case for the GAS tv-c and fix-c models than for the HEAVY t and GAS sep models. The performance of the GAS sep model is particularly dramatic for the 95% VaR predictions: in 47, 38 and 21 cases out of 100 the null hypothesis is rejected using different significance levels. This finding contrasts with the findings of Ahoniemi *et al.* (2015). A possible explanation is that we consider *individual* stocks, while Ahoniemi *et al.* (2015) use index return only.[1] Overnight index returns may be less prone to he substantial fat-tailedness we find for individual stocks.

The performance of the 99% VaR is similar for both the static and time-varying c_t specification. However, there is a striking difference between the two models in terms of the conditional coverage (CC) test for the 95% VaR predictions. The null-hypothesis of a correct conditional coverage is rejected 17, 9 and 2 times when allowing for a time-varying ratio c_t. For a static c, these numbers are substantially higher at 28, 18 and 4 rejections. Hence the GAS fix-c model fails with respect to the independence of the violations, while the GAS tv-c model does not.

Panel B confirms our results on the VaR predictions for the expected shortfall measures. The GAS models with a fixed or time-varying total-over-daytime volatility ratio are superior to the HEAVY t and GAS sep models. Again, modeling the overnight and daytime volatility separately results in very bad ES predictions. The fixed and time-varying c_t specifications behave quite similarly. However, recall that the 'fixed' $\hat{c}$ is updated every 50 observations, and therefore is effectively much less fixed than the name suggests. Effectively, the rolling-

[1] We also apply our models on the S&P 500 index (overnight) returns (and realized variance) and find that backtests on the VaR and ES forecasts do not provide a clear winner. This confirms Ahoniemi *et al.* (2015), as their result is in particular strong for the Russell 2000 index while for the S&P 500 index their main result is less convincing (see Table 3 and 4 of Ahoniemi *et al.* (2015)).

window estimation of the fixed c specification turns it into a time-varying (per 50 days) c_t specification. Overall, we conclude from the table that the GAS tv-c model produces the best VaR and ES predictions.

5 Conclusions

We introduced a new dynamic score driven model for the ratio of total close-to-close volatility and daytime volatility, where the latter depends on the realized variance. The model differs from the existing literature by considering a time-varying ratio instead of a fixed ratio, as suggested by Hansen and Lunde (2005) and Ahoniemi *et al.* (2015). In addition, we scale the *filtered* volatility, taking into account the fat-tailedness of the realized variance using the (univariate analogue of the) GAS F model of Opschoor *et al.* (2018), instead of the realized variance itself.

Using daily returns and realized variances of 100 U.S. stocks over 2001–2014, the new model improves in-sample upon the alternatives, particularly for financial stocks. We showed that the ratio of total to daytime volatility varies through time. Out-of-sample, 1-step ahead VaR and ES predictions improve as well using the specification with a time-varying volatility ratio. Finally, we found that our model dramatically improves upon a model that models the daytime and overnight volatility separately.

References

Ahoniemi, K., A.M. Fuertes and J. Olmo (2015), Overnight news and daily equity trading risk limits, *Journal of Financial Econometrics* **14**, 525–551.

Ahoniemi, K. and M. Lanne (2013), Overnight stock returns and realized volatility, *International Journal of Forecasting* **29**, 592–604.

Andersen, T., T. Bollerslev, F.X. Diebold and P. Labys (2003), Modeling and forecasting realized volatility, *Econometrica* **71**, 529–626.

Andersen, T.G. and T. Bollerslev (1998), Answering the skeptics: Yes, standard volatility models do provide accurate forecasts, *International economic review* pp. 885–905.

Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde and N. Shephard (2009), Realized kernels in practice: trades and quotes, *Econometrics Journal* **12**, 1–32.

Berkman, H., P.D. Koch, L. Tuttle and Y.J. Zhang (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., S.J. Koopman and A. Lucas (2015), Information theoretic optimality of observation driven time series models for continuous responses, *Biometrika* **102**, 325–343.

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

Brownlees, C.T. and G.M. Gallo (2006), Financial econometric analysis at ultra-high frequency: Data handling concerns, *Computational Statistics and 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 Econometrics* **7**, 174–196.

Creal, D., S.J. Koopman and A. Lucas (2013), Generalized Autoregressive Score Models with Applications, *Journal of Applied Econometrics* **28**, 777–795.

Du, Z. and J.C. Escanciano (2016), Backtesting expected shortfall: accounting for tail risk, *Management Science* **63**, 940–958.

Engle, R (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. and G.M. Gallo (2006), A multiple indicators model for volatility using intra-daily data, *Journal of Econometrics* **131**, 3–27.

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

Hansen, P.R. and A. Lunde (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. and A. Luati (2014), Filtering with heavy tails, *Journal of the American Statistical Association* **109**, 1112–1122.

Koopman, S.J., B. Jungbacker and E. Hol (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.B. and J. Wu (2017), A coupled component GARCH model for intraday and overnight volatility, Working Paper.

Lucas, A., B. Schwaab and X. Zhang (2014), Conditional Euro Area Sovereign Default Risk, *Journal of Business & Economic Statistics* **32**, 271–284.

Noureldin, D., N. Shephard and K. Sheppard (2012), Multivariate high-frequency-based volatility (HEAVY) models, *Journal of Applied Econometrics* **27**, 907–933.

Oh, D.H. and A.J. Patton (2018), Time-varying systemic risk: Evidence from a dynamic copula model of CDS spreads, *Journal of Business & Economic Statistics* **36**, 181–195.

Opschoor, A., P. Janus, A. Lucas, and D. van Dijk (2018), New HEAVY models for fat-tailed realized covariances and returns, *Journal of Business & Economic Statistics* **36**, 642–657.

Shephard, N. and K. Sheppard (2010), Realising the future: forecasting with high-frequency-based volatility (heavy) models, *Journal of Applied Econometrics* **25**, 197–231.

Sinclair, Euan (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. and M. Souček (2014), Overnight information flow and realized volatility forecasting, *Finance Research Letters* **11**, 420–428.