Erasmus University Rotterdam

Erasmus School of Economics

Improving Time-Varying Volatility Estimates Using ProPar: a Score-Driven Application on High-Frequency Data

Author:
Sjamiel Bagirov (536308)

Supervisor:
dr. R. Lange
Second Assessor:
B. van Os

June 19, 2023

Abstract

We introduce new (univariate and multivariate) score-driven models based on the ProPar filter, an implicit stochastic updating model. These models are aimed at accurately capturing the joint dynamics of leptokurtic stock returns and time-varying volatilities. We investigate the use of a Student's t distribution instead of the traditional conditional normal distribution for returns and consider high-frequency data to modify existing score-driven time series models. Through simulations that encompass both correct and incorrect return and volatility specifications, as well as empirical applications involving up to 15 dimensions, our findings demonstrate that the implicit models consistently yield more favorable results compared to their explicit counterparts. Moreover, the adoption of the Student's t distribution was found to lead to significant improvements. Our univariate and multivariate score-driven models, ProPar-HF and HEAVY-ProPar-tF, respectively, outperform explicit models in estimating and predicting time-varying volatilities and covariance matrices.

Contents

1 Introduction and Literature

The behavior of the stock market is affected by a variety of factors such as interest rates, corporate earnings, and geopolitical events. It is therefore widely acknowledged that the volatility of this type of financial time-series data is generally not constant. Here, the time-varying nature of volatility is a stylized fact that stems from various sources, such as volatility trends, regimes, volatility clustering, structural breaks, and jumps. Consequently, numerous studies in the literature have proposed diverse techniques to model and forecast this volatility. Obtaining such time-varying volatility estimates and predictions is highly critical in financial econometrics and option pricing, as it allows for better risk assessment and derivative pricing. Accurate models enable financial market agents to make informed decisions, thereby improving market efficiency and stability. However, the process of accurately estimating time-varying parameters presents a significant challenge, mainly because time-series observations are often accompanied by considerable amounts of noise that contaminate the signals.

To obtain a clean time-varying volatility estimate for financial time-series data, various approaches have been proposed. Among these methods, some rely on modeling the noise and volatility itself through filtering techniques such as the popular autoregressive models like the (generalized) autoregressive conditional heteroskedasticity ((G)ARCH) models of Engle (1982), Bollerslev (1986) and Engle & Bollerslev (1986) along with their extensions. In this study, we investigate the performance of a novel approach to tackle this problem, namely the proximal-parameter (ProPar) updating algorithm proposed by Lange et al. (2022). Time-varying parameter models have become increasingly popular in applied econometrics and finance, where capturing the dynamic behavior of time series processes is essential. These models can be broadly categorized into two classes (Cox et al., 1981): observation-driven and parameter-driven models. The ProPar filter joins the scientific strand in the former class. In this class, observations are assumed to follow a certain distribution with time-varying parameters, while parameter-driven models take parameter processes as being driven by their own sources of uncertainty. Common examples of observation-driven and parameter-driven models are GARCH models and stochastic volatility (SV) models, respectively (see Shephard (2005) for an overview of SV models and Heston (1993) for the commonly used SV Heston model).

ProPar's updating algorithm builds upon the literature regarding dynamic conditional score (DCS, Harvey (2013)) and generalized autoregressive score (GAS,Creal et al. (2013)) models, both of which are observation-driven models. The ProPar updating filter extends these explicit stochastic gradient methods, by creating an implicit stochastic gradient method. The main concept of this implicit part is that the score (gradient of the log-likelihood) is evaluated at the

updated parameter value (rather than the predicted parameter value as in the explicit case), which, in turn, uses the observation at that moment to update its parameter value. As such, the ProPar filter also joins the scientific strand of proximal-point methods. Owing to its stochastic modeling, the ProPar filter is related to implicit stochastic gradient methods, as discussed in e.g. Toulis & Airoldi (2017).

The ProPar filter has been shown to exhibit some crucial advantages. It has strong invertibility and optimality properties, even under model misspecification. The filter is also highly robust to large shocks and poorly specified learning rates and converges globally towards a pseudo-truth. ProPar has successful and improved (compared to traditional time-varying estimators) practical applications in time-varying regressions, volatility, and growth-at-risk estimates.

This research aims to provide a comprehensive analysis of the ProPar filter's performance, as well as its potential application in high-frequency financial data. Furthermore, we explore modeling alternatives of the ProPar filter to potentially improve its performance in estimating and predicting volatility. The research is divided into three main branches.

1. We start by conducting simulation studies to investigate the performance of the ProPar filter. Specifically, we will focus on three key aspects. Firstly, we will evaluate the efficiency of the ProPar filter (using maximum likelihood) under a correctly and incorrectly specified stochastic volatility model. Besides, we will examine ProPar's ability to capture simulated jumps following Bates (1996). Lastly, using the simulation study, we test whether there is statistical outperformance of the implicit ProPar framework compared to its explicit DCS/GAS counterpart by performing Diebold-Mariano tests (Diebold & Mariano, 2002).

2. The second branch of interest is the empirical performance of the ProPar filter, regarding time-varying volatility. Here, we will establish whether the implicit part of this proximal-point algorithm is superior to its explicit counterpart by conducting statistical parameter tests and by studying the interaction of the exponential GARCH (EGARCH) model of Nelson (1991) and the ProPar filter. The EGARCH model was shown in Lange et al. (2022) to be similar to the ProPar filter to some degree, but also performed somewhat worse in terms of predictions. This model may also be regarded as explicit as it uses the past volatility prediction instead of the updated volatility estimate for its current prediction. Therefore, a hybrid version of the two filters may potentially display superior performance and provide additional insights into the relative importance of this implicit part. Secondly, we will study alternatives to the modeling assumptions when creating the framework behind the ProPar filter. Specifically, instead of assuming a conditional normal distribution for returns, assuming Student's t distribution will be discovered, causing ProPar's volatility

estimator to change. The Student's t distribution allows for fatter tails, thereby accounting for a greater probability of extreme events or outliers. This assumption may be more valid when considering the leptokurtic stock returns (see e.g Hansen (1994) and Alberg et al. (2008)). Alberg et al. (2008) even found that the EGARCH model in combination with assuming a skewed Student's t model outperformed other GARCH models.

3. The third and final branch of our research is concerned with the performance of the ProPar filter on high-frequency financial data. Research on the use of intraday financial data is not new. Taylor & Xu (1997) for example made use of the intraday standard deviation of returns to better estimate volatility. And in theory, the sum of squares intraday returns should provide a good estimate for the daily realized variance. As such, Andersen et al. (1999) found favorable results when forecasting volatility on multiple horizons, using intraday returns. However, in a more recent study, Ait-Sahalia et al. (2005) argue that this intraday information could suffer heavily from the so-called microstructure noise often found in high-frequency data. Other approaches involve simply down-sampling the data or constructing efficient estimators using smart subsample averaging, such as the two scales realized volatility (TSRV) estimator of Aït-Sahalia et al. (2011). However, the former approach may be statistically non-trivial and inefficient, as it may neglect valuable data points with informative signals, while the latter approach can suffer from model misspecification and lead to spurious limiting behaviors following invalid assumptions. Therefore, the ProPar filter, which makes a tradeoff between the optimal estimate and the closeness to the previous estimate (due to the proximal feature) could be more applicable. Specifically, using sub-batches of observations, 5-minute returns will be calculated to update daily volatilities. In our high-frequency application, we will both present a 1-dimensional modeling framework using ProPar and a multivariate ProPar model (using a framework similar to Opschoor et al. (2018)). The relevance surrounding the use of high-frequency financial data is quite trivial. Namely, the ProPar filter could provide a substantial estimator for trading companies that seek to manage their volatility estimates continuously throughout the day, e.g. to adjust their option pricing estimates.

During the simulations and empirical results, we found that maximum likelihood generally estimates parameters efficiently, under both an incorrectly and correctly specified model for volatility. Here, ProPar was found to generally outperform the explicit GAS model, with an even stronger effect at the correctly specified case. Therefore, implicitly (instead of explicitly) evaluating the score can be regarded as superior information. This finding is in line with our

empirical results, where implicit models generally outperformed GAS, and implicit parameters yielded significantly larger (in magnitude) estimates than explicit parameters. Depending on the loss function and prediction horizon, our implicit model that instead assumes a conditional Student's t distribution for returns was found to provide a more accurate volatility predictor than other (implicit and explicit) dynamical conditional score models. Lastly, using our high-frequency dataset of intraday stock returns, both the univariate and multivariate implicit models were found to significantly outperform their explicit counterparts, thereby providing relatively accurate models for updating and predicting the covariance matrix and volatility estimates.

The rest of the paper is structured as follows. We extensively describe our methodology in Section 2, followed by a simulation study in Section 3. In Section 4, we apply the models to our panel of 15 daily equity returns from the Dow Jones Industrial Average index. Additionally, we introduce our high-frequency models in Section 5, wherein we proceed to implement these models utilizing our high-frequency dataset. Finally, Section 6 contains the conclusion of our research, including suggestions for further research.

2 Modeling Framework

In this section, we present an overview of the ProPar methodology developed by Lange et al. (2022). We subsequently introduce novel model extensions, consisting of assuming a Student's t distribution for returns and a hybrid filter combining the EGARCH model and the ProPar filter.

2.1 ProPar

The implicit updating algorithm of Lange et al. (2022) is based on the following structure. Let K denote the number of dimensions under consideration, such as the number of stocks. Let $\boldsymbol{y}_t$ denote the multivariate returns and let the time-varying parameters be denoted as $\boldsymbol{\theta}_{t|t-1}$ and $\boldsymbol{\theta}_{t|t}$ (all $K \times 1$ vectors). This yields their considered optimization setup:

$$\boldsymbol{\theta}_{t|t} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[\log p(\boldsymbol{y}_t \mid \boldsymbol{\theta}) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_{t|t-1})' \boldsymbol{P}_t (\boldsymbol{\theta} - \boldsymbol{\theta}_{t|t-1}) \right], \tag{1}$$

where $p(\boldsymbol{y}_t \mid \boldsymbol{\theta})$ denotes the multivariate observation density of $\boldsymbol{y}_t$ and $\boldsymbol{P}_t$ is a symmetric and positive-definite penalty matrix (to be estimated using e.g. maximum likelihood), possibly time-varying itself. Conceptually, this means that we are optimizing time-varying parameters while at the same time penalizing deviations from the predicted parameters. Therefore, their model is called a proximal-parameter (ProPar) updating algorithm. What is unique about this filter is that it is an implicit stochastic updating algorithm. Namely, in Equation 1, the gradient of the

log-likelihood function with respect to the parameter vector (also called the score) is evaluated at the updated parameter vector instead of at the one-step ahead predicted parameter vector as in explicit stochastic gradient methods. Mathematically, when computing the first-order condition of Equation 1 with respect to the updated parameter vector, it yields:

$$\begin{aligned} \nabla(\boldsymbol{y}_t \mid \boldsymbol{\theta}_{t|t}) - \boldsymbol{P}_t(\boldsymbol{\theta}_{t|t} - \boldsymbol{\theta}_{t|t-1}) = 0 \\ \implies \boldsymbol{\theta}_{t|t} = \boldsymbol{\theta}_{t|t-1} + \boldsymbol{P}_t^{-1}\nabla(\boldsymbol{y}_t \mid \boldsymbol{\theta}_{t|t}), \end{aligned}$$

where $\nabla(\boldsymbol{y}_t \mid \boldsymbol{\theta}) = \frac{\partial \log p(\boldsymbol{y}_t|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is the score vector which is evaluated at the updated parameter vector $\boldsymbol{\theta}_{t|t}$ rather than at the predicted parameter $\boldsymbol{\theta}_{t|t-1}$ (as in an explicit stochastic gradient method).

After updating their parameter estimate, the authors proposed the following linear prediction step:

$$\boldsymbol{\theta}_{t+1|t} = \boldsymbol{\omega} + \boldsymbol{\Phi}\boldsymbol{\theta}_{t|t}, \tag{2}$$

where $\boldsymbol{\omega}$ is a $K \times 1$ vector and $\boldsymbol{\Phi}$ is a $K \times K$ matrix.

When considering volatility, we first turn to a 1-dimensional parameter space, namely the time-varying volatility for a certain stock. For this, the authors consider the following model specifications. Let stock returns be defined as y_t under the following model:

$$y_t = \mu + \sigma_t z_t \qquad \sigma_t = \exp(h_t) \qquad z_t \overset{i.i.d}{\sim} \mathrm{N}(0,1),$$

where μ, σ_t, and z_t denote the average return of the time-series, time-varying exponential volatility, and a random, standard normal distributed variable, respectively. This implies that y_t is also conditionally normally distributed as $\mathrm{N}(\mu, \exp(2h_t))$. With such a conditional normal density, we can solve for $h_{t|t}$ in a similar manner as the general, multivariate case. Namely, using a constant penalty term, we have the following optimization setup:

$$\begin{aligned} h_{t|t} &= \underset{h \in R}{\operatorname{argmax}} \left[\log p(y_t \mid \exp(h)) - \frac{1}{2}(h - h_{t|t-1})^2 P \right] \\ p(y_t \mid h_t) &= \frac{1}{\exp(h_t)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_t - \mu}{\exp(h_t)}\right)^2}. \end{aligned} \tag{3}$$

Similar to the general setup, when taking the first-order condition with respect to the updating parameter, h_t, and rewrite[1], it yields:

[1] This can be found in Appendix Section 7.2

$$h_{t|t} = h_{t|t-1} + \eta \left[\left(\frac{y_t - \mu}{\exp\left(h_{t|t}\right)} \right)^2 - 1 \right], \tag{4}$$

where $\eta = \frac{1}{P} > 0$ and thus may be regarded as the time-invariant learning rate. Notice that the updated volatility appears both on the left-hand side and the right-hand side. To solve it, we can use a Lambert W function, exact details can be found in Appendix Section 7.2.2. This yields

$$h_{t|t} = h_{t|t-1} - \eta + \frac{1}{2} W_0 \left(2\eta (y_t - \mu)^2 \exp\left(-2(h_{t|t-1} - \eta) \right) \right), \tag{5}$$

where $W_0(x)$ is the Lambert W function for $x \in \mathbb{R}^+$. Furthermore, the ProPar prediction step is modeled as

$$h_{t+1|t} = \omega + \phi h_{t|t}. \tag{6}$$

In the proposed ProPar filter, the parameters (η, ω, and ϕ) can be estimated by maximum likelihood, using the standard prediction-error decomposition. Following Blasques et al. (2022), the resulting parameter set under a one-dimensional explicit stochastic gradient method follows a normal distribution, as supported by the works of Newey & McFadden (1994) and Creal et al. (2013). Under correct specification, this yields

$$\sqrt{n} \left(\widehat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}_0 \right) \overset{d}{\sim} N\left(0, \boldsymbol{J}^{-1}\right), \tag{7}$$

where n denotes the length of the time series, $\widehat{\boldsymbol{\theta}}_{ML}$ denotes the estimated optimal parameter vector under maximum likelihood, $\boldsymbol{\theta}_0$ denotes the true parameter vector, and $\boldsymbol{J} = -\operatorname{E}_t[\boldsymbol{H}_t]$ denotes the (Fisher) information matrix. Here, $\boldsymbol{H}_t := \ell_t''(\boldsymbol{\theta})$ denotes the Hessian matrix, where $\ell(\boldsymbol{\theta})$ represents the observation log-likelihood, evaluated at the parameter set $\boldsymbol{\theta}$. To compute standard errors, we use Equation 7 and apply a similar approach as Creal et al. (2013) by computing the finite difference Hessian, as an approximation for the true Hessian matrix. Throughout this report, we apply a 5% significance level to determine the significance of each parameter.

In Lange et al. (2022), the authors display the relation of their ProPar filter with explicit score-driven models. In fact, the explicit score-driven models can be obtained within the ProPar framework by linearly approximating the logarithmic density target around the prediction. Here, the same normal distribution for returns is assumed and the following optimization problem is considered:

$$h_{t|t} = \underset{h \in R}{\operatorname{argmax}} \left[\log p(y_t \mid \exp(h_{t|t-1})) + \nabla(y_t \mid \exp(h_{t|t-1}))(h - h_{t|t-1}) - \frac{1}{2}(h - h_{t|t-1})^2 P \right] \tag{8}$$

Using similar steps as in the ProPar framework (so taking the first order condition and solving for $h_{t|t}$) yields the following DCS/GAS model:

$$h_{t|t} = h_{t|t-1} + \eta \left[\left(\frac{y_t - \mu}{\exp\left(h_{t|t-1}\right)} \right)^2 - 1 \right]. \tag{9}$$

Now, the right-hand side is immediately computable. This simplifies the execution of the update, but may also lead to an efficiency loss. Equation 6 can still be used for the 1-step ahead volatility predictions. Namely, the prediction can be rewritten into

$$\begin{aligned} h_{t+1|t} &= \omega + \phi h_{t|t} \\ &= \omega + \phi \left(h_{t|t-1} + \eta \left[\left(\frac{y_t - \mu}{\exp\left(h_{t|t-1}\right)} \right)^2 - 1 \right] \right) \\ &= \omega + \phi h_{t|t-1} + \phi \eta \nabla(y_t \mid \exp(h_{t|t-1})), \end{aligned} \tag{10}$$

which exactly formulates a GAS model (Creal et al., 2013) under the given modeling framework for returns.

2.2 Student's t distribution

In the modeling of returns and volatility, we made use of the assumption that returns are conditionally normally distributed. Following the literature, this may not always be a reasonable assumption. However, the advantage of using this distribution is that it enables us to rewrite the volatility optimization into Equation 4, for which a closed-form solution can be uniquely found using the Lambert W function. Now, when assuming a Student's t distribution for the residuals z_t, the returns are also Student's t distributed with $\nu > 4$ degrees of freedom and, consequently, the volatility optimization does not turn into Equation 4.

Hence alternatively, we can modify our optimization setup following Harvey & Lange (2017). The authors consider a generalized Student's t-distribution (McDonald & Newey, 1988) for returns, yielding what Harvey & Lange (2017) call the Beta-Gen-t-EGARCH model. Mathematically, we apply the following modeling assumptions:

$$\begin{aligned} y_t &= \mu + \exp(h_t) z_t, \text{where } z_t \overset{i.i.d}{\sim} \mathrm{t}(\xi, \nu) \\ h_{t|t} &= \underset{h \in \mathbb{R}}{\operatorname{argmax}} \left[\log p(y_t \mid \exp(h)) - \frac{1}{2}(h - h_{t|t-1})^2 P \right] \\ p(y_t \mid h_t) &= \frac{\xi}{2\nu^{1/\xi}} \frac{1}{B(1/\xi, \nu/\xi)} \frac{1}{\left(1 + \left|(y_t - \mu) / \exp(h_t)\right|^{\xi} / \nu\right)^{(\nu+1)/\xi}}, \end{aligned} \tag{11}$$

where the non-standardized Student's t-distribution with positive shape parameters $\xi = 2$ and

ν degrees of freedom is a special case of the provided observation density and yields the known Beta-t-EGARCH model. Here, $B(.,.)$ denotes the beta function and $\exp(h_t)$ sets the scaling of the distribution. Moreover, μ is a constant return average, z_t denotes the standardized i.i.d. Student's t distributed shock, and h_t is the time-varying conditional scaling factor (which should be multiplied by $\sqrt{\frac{\nu}{\nu-2}}$ for time-varying volatility (when $\xi = 2$). The authors formulated the score function evaluated at the predicted parameter $h_{t|t-1}$ as:

$$\nabla(y_t \mid h_{t|t-1}) = \frac{d}{dh_t} \log p\left(y_t \mid \exp(h_t)\right) \Big|_{h_t = h_{t|t-1}} = (\nu + 1) b_t, \tag{12}$$

where $b_t = \frac{\left(|y_t - \mu| e^{-h_{t|t-1}}\right)^{\xi} / \nu}{\left(|y_t - \mu| e^{-h_{t|t-1}}\right)^{\xi} / \nu + 1} = \frac{|\varepsilon_t|^{\xi}}{|\varepsilon_t|^{\xi} + \nu}$. In this research, we simply assume $\xi = 2$ corresponding to a Student's t-distribution and implicitly evaluate the score at the updated parameter $h_{t|t}$ itself, yielding $\nabla(y_t \mid h_{t|t})$. Similar to the ProPar framework, we take the first-order condition of the maximization formula, yielding: $h_{t|t} = h_{t|t-1} + P^{-1} \nabla(y_t \mid h_{t|t}) = h_{t|t-1} + \eta \nabla(y_t \mid h_{t|t})$.

Instead of solving the optimization problem mathematically, we apply a more numerical analysis. That is, since both terms in the maximization formula are continuous and differentiable with respect to the real-valued $h_{t|t}$, we can apply the Newton-Raphson (NR) root-finding algorithm on the first-order condition of the maximization formula. The general Newton-Raphson search algorithm works as follows, following the analytical steps of e.g. Nocedal & Wright (2006). Here, the function to be evaluated is the gradient of the optimization formula, thereby providing an approximation for the optimal value for $h_{t|t}$.

Algorithm 1 Newton-Raphson algorithm

Input: Function $f(x)$, derivative $f'(x)$, initial point x_0

Output: $\hat{x}$ s.t. $f(\hat{x}) \approx 0$

$y \leftarrow f(x_0)$

$y' \leftarrow f'(x_0)$

$x_1 \leftarrow x_0 - y/y'$ ▷ Do Newton's computation

$x_0 \leftarrow x_1$ ▷ Update x_0 to start the process again

The optimization scheme is repeated iteratively until one of the following stopping criteria is met:

1. The maximum number of iterations $i_{\max}$, which is set to $2{\times}10^3$, is reached. Note, in our research, the average number of iterations only equals 2.403, indicating relatively fast convergence.

2. The changes in the state become smaller than the tolerance value $\epsilon = 10^{-4}$, or the derivative value of the function becomes smaller than the threshold value δ. This is set to 10^{-3},

to prevent the denominator from becoming too small and causing explosive behavior.

Hence, to solve for $h_{t|t}$, the function to be evaluated is the gradient of the optimization formula, yielding $f(h_{t|t}) = \nabla(y_t \mid h_{t|t}) - P(h_{t|t} - h_{t|t-1})$. We have derivative[2]:

$$\begin{aligned} f'(h_{t|t}) &= H(y_t \mid h_{t|t}) - P, \\ H(y_t \mid h_{t|t}) &= (\nu+1)\frac{-2\nu(y_t-\mu)^2\exp(-2h_{t|t})}{\left(\left(\frac{y_t-\mu}{\exp(h_{t|t})}\right)^2+\nu\right)^2}. \end{aligned} \tag{13}$$

It is crucial to acknowledge the possible challenges that may arise when applying the Newton-Raphson algorithm. One of these challenges is convergence issues that could result from e.g. a weak initial guess. In this study, we will address this by using $h_{t|t-1}$ as a starting point, which can be conjectured to be close to the $h_{t|t}$ solution due to the fact that the ProPar framework possesses the 'proximal' feature. Another issue would be that there could be a non-zero probability that the NR algorithm converges to a local optimum instead of a global optimum. Given that $f'(h_{t|t}) = \mathrm{H}(y_t \mid h_{t|t}) - P$ is actually the second derivative of the maximization formula in Equation 11, we can state that the NR algorithm will find the global (rather than local) optimum whenever $f'(h_{t|t})$ is always negative for all values of $h_{t|t}$. A simple similar example would be $f(x) = -x^2$, where the second derivative, -2, is always negative for all values of $x \in \mathbb{R}$. Hence, this would ensure that only one maximum (= global optimum) can be attained. Taking this into consideration, notice that in Equation 13, the sign of the derivative is consistently negative for all values of $\nu > 0$ (taking into account that the penalty term P is also greater than zero) it follows that the first derivative (which is equivalent to the second derivative of the maximization formula) is strictly negative for $t = 1, \ldots, T$. As a result, the risk of encountering a local optimum, as opposed to a global one, is rendered irrelevant. In total, the algorithm yields optimal updated scales $\exp(h_{t|t})$, $t = 1, \ldots, T$ (transformable into time-varying volatility by factor $\sqrt{\frac{\nu}{\nu-2}}$).

For the 1-day ahead predictions, we again apply the prediction step in Equation 6. Let us denote this modified version as the Implicit-Beta-t-EGARCH (IBT-EGARCH) model.

2.3 Implicit EGARCH

In this section, we provide a hybrid modeling framework (based on the exponential GARCH (EGARCH) model developed by Nelson (1991) and the ProPar filter) to also assess the importance of the implicit aspect of the proximal-point algorithm. We adopt a similar approach to

[2]Derivation can be seen in Appendix subsection 7.2

that of Lange et al. (2022) in specifying the EGARCH volatility filter, namely

$$\log\left(h_{t+1|t}\right) = \omega + \alpha\left(|z_t| - \sqrt{\frac{2}{\pi}}\right) + \beta\log\left(h_{t|t-1}\right), \tag{14}$$

where ω, α, β are parameters to be estimated by maximum likelihood and $z_t = \frac{y_t - \mu}{h_{t|t-1}}$. Notice, the right-hand side can be regarded as explicit in the sense that the 1-step ahead volatility prediction is evaluated at the past prediction $h_{t|t-1}$ instead of the updated volatility $h_{t|t}$. To that extent, we provide two hybrid implicit EGARCH models. In the first one (denoted I-EGARCH-1), we simply add an extra parameter in the EGARCH filter:

$$\log\left(h_{t+1|t}\right) = \omega + \alpha\left(|z_t| - \sqrt{\frac{2}{\pi}}\right) + \beta\log\left(h_{t|t-1}\right) + \gamma\log\left(h_{t|t}\right), \tag{15}$$

where $h_{t|t}$ first needs to be computed using the ProPar updating step, γ also is to be computed using maximum likelihood, and z_t is an explicit shock as defined before. Notice that in the ProPar framework, α and β are equal to zero. So by comparing the magnitudes of β and γ, we can establish whether the extra implicit step under γ is of greater importance than the explicit step under β. This follows from the fact that these two parameters capture the effect of the latest volatility estimations with respect to the current volatility prediction (contrary to the parameter α which captures the effect of the current shock). Given a null hypothesis stating equal importance between the two variables, we statistically test for it by performing the Wald test for each stock. Specifically, under the null hypothesis of equal coefficients (between β and γ), the test statistic takes on the following form:

$$\mathrm{W} = \frac{|\hat{\beta}| - |\hat{\gamma}|}{\sqrt{\hat{\sigma}_\beta^2 + \hat{\sigma}_\gamma^2 - 2\hat{\sigma}_{\beta\gamma}}} \overset{H_0}{\sim} N(0,1), \tag{16}$$

where $\hat{\sigma}_\beta^2$, $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_{\beta\gamma}$ are the estimated (co)variances of the (maximum-likelihood generated) coefficients. The Likelihood-Ratio (LR) test is also applied in this research to determine whether the addition of this implicit parameter yields a significantly better fit. Namely, let us denote $L_0(\omega, \alpha, \beta)$ as the log-likelihood under the restriction of $\gamma = 0$, which yields a model equivalent to the standard EGARCH filter in Equation 14. Furthermore, let $\hat{L}(\omega, \alpha, \beta, \gamma)$ be the log-likelihood of the unrestricted I-EGARCH-1 model. Then, under the null hypothesis that the restricted model is the correct specification, the LR test statistic takes on the following form:

$$\mathrm{LR} = -2(L_0(\omega, \alpha, \beta) - \hat{L}(\omega, \alpha, \beta, \gamma)) \overset{H_0}{\sim} \chi_1^2, \tag{17}$$

where chi-squared distribution is assumed with one degree of freedom, following from the addition of a single parameter in the unrestricted model.

In the second hybrid model (denoted I-EGARCH-2), we directly replace the explicit volatility step with the implicit volatility update (i.e. enforce $\beta = 0$). Mathematically,

$$\log\left(h_{t+1|t}\right) = \omega + \alpha\left(|z_t| - \sqrt{\frac{2}{\pi}}\right) + \gamma \log\left(h_{t|t}\right), \tag{18}$$

where $h_{t|t}$ again first needs to be computed using the ProPar updating step and z_t (shock variable) is defined as before. This hybrid model especially holds promising potential for improved out-of-sample predictions since it leverages the advantages of both the EGARCH model and the ProPar filter while maintaining a parsimonious set of parameters.

2.4 Forecasting analysis

When making h-step ahead predictions, we turn to the predicted volatility (PVol) over a d-day horizon as the square root of the predicted variance(PVar) over a d-day horizon. This, in turn, could be modeled by the conditional expectation of the sum of squared (forward) returns, also seen in e.g. Andersen et al. (1999). For the ProPar, GAS and IBT-EGARCH model, this mathematically yields[3]

$$\text{PVar}_t := \text{E}_t\left[\sum_{i=1}^{d} y_{t+i}^2\right] \approx d\mu^2 + \sum_{i=1}^{d} \exp\left(2\omega\frac{1-\phi^i}{1-\phi} + 2\phi^d h_{t|t}\right). \tag{19}$$

In our research, we decide to follow Andersen et al. (1999) and investigate $d \in \{1, 5, 20\}$, corresponding to a prediction horizon of one day, one calendar week, and one calendar month respectively. Using the predicted volatility, we first turn to the mean squared error (MSE) as a loss function. This takes the following formula: $\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}(\text{PVar}_t - H_t)^2$, where T denotes the length of the time-series and H_t is the proxy for the variance at time t. One trivial proxy for the variance over a d-day horizon is the (forward-looking) sum of squared daily returns. Mathematically,

$$H_t = \sum_{i=1}^{d} y_{t+i}^2. \tag{20}$$

Note, the choice of loss function (here, the MSE) is important. Patton (2011) showed that well-known loss functions could lead to inaccurate comparisons between volatility forecasts, stemming from a noisy and perhaps false proxy for the variance. The MSE loss function furthermore is heavily influenced by a small number of large observations, making it less robust to outliers. For our research, this is of considerable concern given that volatility processes in

[3] Derivations can be seen in Appendix subsection 7.2

stock market returns are generally dominated by a small number of such extreme events. Under a simulation analysis, Patton & Sheppard (2009) argued for the use of the QLIKE loss function due to its superior power in DM tests. Therefore, in this research, we also compare the forecasting ability of two models using the QLIKE function, which takes on the following form:

$$QLIKE = \log(\text{PVar}_\text{t}) - \frac{H_t}{\text{PVar}_\text{t}}. \tag{21}$$

This loss function can be argued to be more robust to outliers compared to the MSE. However, unlike MSE (a symmetrical loss function), QLIKE is asymmetrical and treats positive and negative losses differently. Consequently, predictive volatility processes with a known (especially positive) bias should be carefully studied, since the QLIKE function may display favorable results for such a model.

To determine whether potential outperformance is statistically significant, we will use the Diebold-Mariano (DM) test of Diebold & Mariano (2002) where the null hypothesis states equal predictive ability. The test is constructed as follows. Let $d_{t,p}$ be the loss function at time t for model p (for simplicity, either the ProPar or GAS model but it can be any other model comparison). Then, denote e_t as the difference between the loss functions of the two models at time t, that is, $e_t = d_{t,ProPar} - d_{t,GAS}$. The DM test statistic is modeled as:

$$\text{DM} = \frac{\hat{e}}{\sqrt{\left[\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k\right]/T}} \overset{H_0}{\sim} N(0,1), \tag{22}$$

where $\hat{e}$ is the average of all e_t values ($t = 1, \ldots, T$), $\gamma_k = \frac{1}{T}\sum_{t=k+1}^{T}(e_t - \hat{e})(e_{t-k} - \hat{e})$ and T denotes the length of the time-series. The choice for h can be quite complex but a well-known rule-of-thumb is to take $h = T^{\frac{1}{3}} + 1$, which is implemented in our research. A common alternative is to determine the number of significant autocorrelations in $\hat{e}$ a priori. Using a significance level of 5%, the DM test statistics and their p-values can be estimated in order to test the null hypothesis stating that the two models achieve equal predictive ability. If rejected, the model with the lower average loss function is assumed to have statistical outperformance.

3 Simulation

In simulation studies, we start a comparative analysis between the ProPar filter and its DCS/GAS counterpart to evaluate their respective capabilities in estimating volatility. This comparison will help us determine which method performs better under the given conditions and assumptions. This way, we can establish the importance and usefulness of this implicit part (in implicit stochastic proximal-point algorithm). This will be done by generating a simulated sample of

volatilities within a specified stochastic volatility (SV) model, and comparing the time-varying volatility estimates under the ProPar framework as in Equation 4 and the DCS/GAS counterpart as in Equation 9. We implement this by generating volatility and stock price observations using multiple time-series lengths. Our research studies the lengths $T \in \{252, 504, 1009, 4036\}$, corresponding to 1, 2, 4, and 16 years with roughly 252 trading days. We simulate such time series one thousand times for each simulation scenario. We extract the data using both a relatively straightforward correctly-specified SV model, and an incorrectly specified SV model. This way, we can determine the performance of both filters using a broader data sample, which ultimately is more informative.

We start by utilizing the popular discrete-Heston SV model (Heston, 1993) for our incorrectly specified framework. Specifically, the Heston model takes the following stochastic differential equations (SDE):

$$\begin{aligned} dS_t &= \mu S_t dt + \sqrt{v_t} S_t dW_t^s \\ dv_t &= \kappa\left(\theta - v_t\right) dt + \sigma\sqrt{v_t} dW_t^v, \end{aligned} \tag{23}$$

where μ is a constant drift for the price process dS_t, dW_t denotes the increment of a standard Wiener process with zero mean and variance 1, and dv_t is the instantaneous change in the volatility process. Here, $\theta > 0$ is the long-term mean of variance under risk-neutral dynamics and $\kappa > 0$ is the rate of mean reversion of variance under risk-neutral dynamics. Here, the processes dW_t^s and dW_t^v are related by correlation ρ, that is, $dW_t^s dW_t^v = \rho dt$. To make it discrete, we consider the following discrete-Heston SV model:

$$\begin{aligned} S_{t+\Delta t} &= S_t \exp\left(\left(\mu - \frac{1}{2} v_t\right)\Delta t + \sqrt{v_t \Delta t} z_t^s\right) \\ v_{t+\Delta t} &= v_t + \kappa\left(\theta - v_t\right)\Delta t + \sigma\sqrt{v_t \Delta t} z_t^v, \end{aligned} \tag{24}$$

where z_t^s and z_t^v $(t = 1, \ldots, T)$ are identically distributed samples from a normal distribution with mean 0, variance 1 and $\rho_{sv} = \rho$ and Δt denotes the length of each time interval.

In Table 1, we display the 6 different simulation scenarios, taken from Lord et al. (2010) and Van Haastrecht & Pelsser (2010).

Next to these six scenarios and four different time-series lengths, we are interested in ProPar's ability to capture (simulated) jumps well. To account for jumps, we follow the methodology of Bates (1996) that includes jumps in its SDE for the price process, the so-called SVJ model. Namely, using a Poisson process dN_t (independent of other motions) for jumps with

Table 1: Test scenarios for the Heston model, with their corresponding parameter values. For all cases, we use S_0 equal to 100 and $\Delta t = \frac{1}{252}$.

	1	2	3	4	5	6
μ	0	0	0	0	0	0.05
V_0	0.04	0.04	0.09	0.09	0.04	0.09
κ	0.5	0.3	1	2	0.5	1
θ	0.04	0.04	0.09	0.09	0.04	0.09
σ	1	0.9	1	1	1	1
ρ	-0.9	-0.5	-0.3	-0.3	0	-0.3

rate/intensity λ, we adjust the Heston SDE as follows:

$$\begin{aligned} dS_t &= (\mu - \lambda\hat{\gamma})S_t dt + \sqrt{v_t}S_t dW_t^s + J_{N_t}S_t dN_t \\ dv_t &= \kappa\left(\theta - v_t\right)dt + \sigma\sqrt{v_t}dW_t^v, \end{aligned} \tag{25}$$

where J_{N_t} is the N_t'th percentage jump size, log-normally distributed as $\log(J_{N_t}+1) \sim N(\log(\hat{\gamma}+1) - \frac{1}{2}\delta^2, \delta^2)$.

In our application, we take the parameter settings of such an SVJ model from Duffie et al. (2000). The authors derived these estimations from a calibration process (by minimizing the MSE pricing error) performed on options of the S&P 500 index. This yields the following (seventh) scenario:

1. Belonging to the jump process: $\lambda = 0.11$ such that the mean of this process (i.e. the expected number of jumps) is equal to λT), $\hat{\gamma} = -0.12$ and $\delta = 0.15$.

2. Other parameter calibrations: $\mu = 0$, $\sqrt{V_0} = 0.094$, $\kappa = 3.99$, $\theta = 0.014$, $\sigma = 0.27$ and $\rho = -0.79$

We apply these scenarios to the ProPar and the GAS counterpart model for volatility to compare an implicit and explicit stochastic gradient method. For this, we also make use of the MSE loss function. Since we are doing 10^3 repetitions for different time-series lengths, the MSE formula is now altered as follows. MSE $= \frac{1}{N\times T}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(\exp(h_{t|t}) - h_t\right)^2$, where $N = 10^3$ denotes the number of repetitions, $T \in \{252, 504, 1009, 4036\}$ denotes the length of the simulated time-series, $h_{t|t}$ denotes the estimated log-volatility and h_t is the simulated volatility at time t. To determine whether potential outperformance is statistically significant, we will also use the Diebold-Mariano test for each scenario and time-series length ($T \in \{252, 504, 1009, 4036\}$). Here, the GAS filter is the benchmark model, such that the DM test statistic is negative (positive) when the ProPar filter, on average, has a lower (higher) MSE.

In Table 2 we display the results for $T = 1009$ (4 years). Based on the reported standard errors, Table 2 shows that the maximum-likelihood optimization for both filters results in quite efficient parameter estimates. While many parameters are estimated to be significantly different from 0, the standard errors for the ProPar filter are generally slightly higher, possibly related due to more precision errors caused by the complex Lambert W function needed to analytically solve for $h_{t|t}$. Furthermore, since the DM test statistics are always negative, the ProPar filter consistently has the lowest MSE for each scenario, suggesting that its implicit updating algorithm is superior to its explicit counterpart. The DM test statistics significantly validate this suggestion for scenarios 4 and 7. Scenario 4 is unique in the sense that it has the highest value for κ, the rate of mean reversion for the volatility process. Consequently, the ProPar filter seems to behave well for a relatively stable time series. Because of this enhanced stability, the estimated learning rate, η, under the ProPar filter is substantially larger, thereby allowing for more sensitivity during low volatility periods. However, since scenario 7 includes jumps in its stochastic differential equation for the asset price, it appears that the ProPar filter is less vulnerable to these sudden shocks, hence also presenting a more suitable model for highly noisy data. While it is true that for scenario 7, the persistence parameter (ϕ) is larger under the ProPar filter (thereby suggesting worse predictions for unstable time series), it also has a substantially larger learning rate (η) compared to its GAS counterpart. The latter relationship thus seems to outweigh the former in importance. Besides, during low-volatility periods, sudden large shocks adversely impacted the GAS filter more, due to its explicit modeling — where the score is evaluated at the predicted volatility — as opposed to the implicit updating strategy of ProPar where the score is evaluated at the updated volatility.

In Table 3, we display the results when considering a substantially shorter time series ($T = 252$). Here, maximum likelihood still estimates the parameters quite efficiently for both filters, following small standard errors. However, some parameters lose their significance. This outcome can be attributed to the reduced amount of information available in a shorter time series, resulting in fewer data points that can be effectively utilized for parameter calibration. Still, the majority of the parameters are significant, indicating that maximum likelihood efficiently estimates the parameters both for a long and short (simulated) time series. Compared to Table 2, the MSEs for the GAS filter are substantially higher for scenarios 2 and 5. This can be the result of a few outliers, caused by large return shocks following extremely low volatility periods such that the explicit score function displays explosive behavior (in combination with the exponential formulation for volatility). As a result, the DM test statistic, using differences in the MSE loss function, also contains a few extreme outliers, creating insignificant accuracy comparisons.

Nevertheless, in the other five scenarios, the ProPar filter statistically outperforms its GAS counterpart. This is a substantial improvement compared to Table 2 where this was the case for only two scenarios. Hence, the ProPar filter seems to achieve a relatively better modeling accuracy than its explicit GAS counterpart for shorter time series. One explanation for this would be that the event of having a large return shock following extremely low volatility periods is considerably lower for shorter time series. Therefore, the DM test statistics for these data samples are expected to be based on fewer outliers, thereby increasing the chance of statistical significance. This is supported by the fact that in scenarios 2 and 5, the difference in MSE is quite high, yet not significant. In Appendix Section 7.3, a similar comparison analysis can be made for the results using $T = 504$ and $T = 4036$.

Table 2: Average parameter estimates and standard errors in parentheses, MSEs of the filtered volatility updates, and their corresponding DM test statistic.

Scenario	1	2	3	4	5	6	7
	Maximum Likelihood Optimization						
$\hat{\eta}$	0.636(0.070)	0.627(0.072)	0.302(0.044)	0.208(0.030)	0.559(0.065)	0.289(0.040)	0.072(0.015)
	0.055(0.005)	0.055(0.005)	0.058(0.006)	0.071(0.007)	0.055(0.005)	0.057(0.006)	0.038(0.006)
$\hat{\omega}$	0.165(0.028)	0.186(0.028)	0.091(0.018)	0.066(0.013)	0.143(0.024)	0.087(0.017)	0.001(0.006)
	-0.093(0.010)	-0.090(0.009)	0.011(0.008)	0.039(0.007)	-0.091(0.008)	-0.008(0.004)	-0.069(0.009)
$\hat{\phi}$	0.925(0.010)	0.901(0.010)	0.949(0.008)	0.964(0.007)	0.919(0.009)	0.954(0.007)	0.972(0.008)
	0.882(0.008)	0.899(0.008)	0.956(0.008)	0.904(0.010)	0.876(0.007)	0.978(0.004)	0.779(0.016)
$\hat{\mu}$	-0.026(0.023)	-0.020(0.012)	-0.015(0.020)	-0.020(0.026)	-0.012(0.012)	0.002(0.022)	-0.005(0.019)
	-0.026(0.023)	-0.021(0.022)	-0.014(0.023)	-0.019(0.030)	-0.012(0.016)	0.002(0.026)	-0.005(0.019)
	Performance Metrics						
MSE_ProPar	0.966	0.143	0.569	1.110	0.176	1.005	0.194
MSE_GAS	1.890	0.562	0.927	1.441	0.681	1.482	0.212
DM	-1.011	-1.159	-1.010	**−13.350**	-1.064	-1.010	**−19.349**

$T = 1009$. For the parameter estimates, the upper and lower element corresponds to those of the ProPar and GAS filter, respectively. The DM test statistics in bold indicate a p-value lower than the significance level of 0.05.

Table 3: Average parameter estimates and standard errors in parentheses, MSEs of the filtered volatility updates and their corresponding DM test statistic.

Scenario	1	2	3	4	5	6	7
	Maximum Likelihood Optimization						
$\hat{\eta}$	0.685(0.143)	0.631(0.143)	0.336(0.122)	0.269(0.134)	0.631(0.138)	0.329(0.089)	0.137(0.239)
	0.055(0.010)	0.053(0.014)	0.052(0.013)	0.050(0.015)	0.055(0.011)	0.052(0.019)	0.044(0.020)
$\hat{\omega}$	0.152(0.055)	0.114(0.055)	0.145(0.070)	0.098(0.044)	0.126(0.051)	0.102(0.039)	-0.049(0.028)
	-0.093(0.015)	-0.081(0.019)	-0.003(0.017)	-0.000(0.027)	-0.101(0.020)	0.011(0.025)	-0.114(0.028)
$\hat{\phi}$	0.910(0.025)	0.897(0.026)	0.887(0.049)	0.903(0.040)	0.895(0.025)	0.930(0.025)	0.838(0.049)
	0.921(0.013)	0.889(0.019)	0.866(0.019)	0.796(0.034)	0.870(0.017)	0.871(0.026)	0.733(0.044)
$\hat{\mu}$	-0.005(0.039)	-0.006(0.030)	-0.008(0.078)	-0.010(0.081)	0.003(0.027)	0.021(0.070)	0.003(0.037)
	-0.005(0.039)	-0.006(0.037)	-0.009(0.082)	-0.010(0.085)	0.003(0.035)	0.021(0.078)	0.002(0.038)
	Performance Metrics						
MSE_ProPar	0.502	0.494	1.333	1.106	0.363	1.148	0.156
MSE_GAS	1.474	5.289	1.606	1.372	6.248	1.372	0.179
DM	**−2.114**	-1.462	**−5.480**	**−6.830**	-1.208	**−9.250**	**−5.924**

$T = 252$. For the parameter estimates, the upper and lower element corresponds to those of the ProPar and GAS filter, respectively. The DM test statistics in bold indicate a p-value lower than the significance level of 0.05.

For our correctly specified case, note that we assumed that returns are conditionally normally distributed with an exponential formulation for volatility, both for the ProPar and GAS filter. Hence, a trivial correctly-specified SV model would be an EGARCH SV model. Mathematically, this has the following framework (see e.g. Harvey et al. (1994)):

$$\begin{aligned} y_t &= \exp\left(\frac{1}{2}v_t\right)\epsilon_t \\ v_t &= \omega + \phi(v_{t-1} - \omega) + \eta_t, \end{aligned} \tag{26}$$

where $\epsilon_t \sim \text{NID}(0,1)$ is independent from $\eta_t \sim \text{NID}(0, \sigma_\eta^2)$, y_t denotes the simulated return at time t, modeled by variance $\exp(v_t)$, using a mean ω and persistence parameter ϕ. In this research, we follow parameter calibrations for ω, ϕ, and σ_η^2 following the literature that performed it using S&P 500 returns (see e.g. Tsay (2005)). This yields $\omega = -0.06$, $\phi = 0.904$ and $\sigma_\eta^2 = 0.135$. The high persistence parameter ϕ closely mirrors stock market volatility by inducing volatility clustering, a commonly observed stylized fact in volatility processes. Therefore, the aforementioned parameter settings are particularly useful as they will align with our empirical dataset, which comprises stocks from the Dow Jones Industrial Average index (more information about this dataset will follow in the next section).

Using this correctly specified framework, we reveal the estimated maximum-likelihood estimators (and standard errors) for both filters and all four time-series lengths in Table 4. Similarly to the incorrectly specified Heston model, the maximum-likelihood optimization for both filters generally yields efficient parameter estimates. For the ProPar filter, the learning parameter

η tends to decrease substantially with an increase in simulated time series length. A reverse effect is found for the persistence parameter ϕ. One explanation for this phenomenon would be that a short time series is relatively more affected by a large shock due to the high persistence parameter setting in Equation 26, whereas for a longer time series, the effects of such shocks may decay better. Namely, with a high persistence parameter, the effects of large volatility shocks are long-lasting, thereby influencing shorter time series for a relatively longer period of time. As a result, the learning parameter should be lower for these time series (equivalent to a higher penalty parameter), and vice versa for a longer time series. Consequently, the maximum-likelihood estimator for the persistence parameter ϕ can be higher for longer time series, since relatively less simulated volatility observations are highly affected by large past shocks, thereby increasing the value of the previous volatility.

Table 4 also displays the accuracy analysis based on the MSE. Now, the ProPar filter is not only significantly outperforming GAS in the shortest time series but also in the longest one. Using the incorrectly specified case, the ProPar filter struggled in some scenarios with finding significant outperformance for long time series.

Table 4: Average parameter estimates and standard errors in parentheses, MSEs of the filtered volatility updates, and their corresponding DM test statistic.

Scenario	T=252	T=504	T=1009	T=4036
	Maximum Likelihood Optimization			
$\hat{\eta}$	0.227(0.157)	0.159(0.062)	0.148(0.033)	0.138(0.015)
	0.053(0.022)	0.055(0.015)	0.057(0.010)	0.059(0.005)
$\hat{\omega}$	0.062(0.034)	0.045(0.019)	0.042(0.012)	0.040(0.006)
	0.027(0.018)	-0.014(0.009)	-0.009(0.006)	0.008(0.002)
$\hat{\phi}$	0.811(0.090)	0.871(0.046)	0.895(0.025)	0.902(0.012)
	0.778(0.075)	0.865(0.043)	0.900(0.026)	0.906(0.011)
$\hat{\mu}$	-0.006(0.063)	-0.000(0.044)	-0.001(0.030)	0.002(0.015)
	-0.006(0.066)	-0.000(0.046)	-0.001(0.032)	0.002(0.016)
	Performance Metrics			
MSE_ProPar	0.415	0.433	0.437	0.436
MSE_GAS	0.535	0.623	0.570	0.650
DM	**-24.785**	**-1.992**	**-4.090**	**-3.104**

The upper and lower elements of parameter estimates correspond to those of the ProPar and GAS filters, respectively. The DM test statistics in bold indicate a p-value lower than the significance level of 0.05.

In summary, parameters in the ProPar and GAS filter can be estimated generally efficiently using maximum likelihood. Besides, ProPar significantly outperforms its explicit counterpart in many scenarios and time series lengths, using an incorrectly specified modeling framework. This effect is stronger when using a correct specification, significantly outperforming GAS in all cases. Consequently, the implicit evaluation of the score, as opposed to the explicit evaluation, can be considered of greater value. This study therefore also provides support for the adoption of the ProPar filter as a relatively more suitable model for empirical financial time series.

4 Empirical Application: Daily U.S. Equity Returns

4.1 Data

For the empirical studies, close-to-close daily stock returns are used. These are extracted using stock price data from Yahoo Finance[4] for 15 Dow Jones Industrial Average stocks from 2000-2022 (5786 observations per stock) [5]. Furthermore, they are transformed into daily log returns by taking the natural logarithm, taking differences, and multiplying by 100 to yield percentage log returns.

As discussed in Section 1, the assumption of normally distributed daily stock returns is often rejected throughout the literature. Instead, in this research we study the effects of assuming a conditional Student's t distribution for returns. To empirically motivate the use of such a distribution, we conduct Kolmogorov-Smirnov tests on both the optimal normal and Student's t distribution for each stock. This way, we can establish statistical validation for the null hypothesis stating that the theoretical distribution being tested (here, normal or Student's t) is identical to the population CDF. The test statistic is as follows:

$$D = \sup_x |F_n(x) - F(x)|, \tag{27}$$

where x is a random daily return observation, $F_n(x)$ is the empirical CDF (eCDF) for the corresponding dataset of returns, and $F(x)$ denotes the CDF of the theoretical distribution being tested. It thus measures the maximum absolute difference between the eCDF and the CDF of the given distribution. For large sample sizes (which could be argued to be the case for our dataset), the distribution of the test statistic under the null hypothesis resembles the Kolmogorov distribution. In Table 5, we list the p-values for this test, for each stock. It indicates for all equities that a normal distribution is significantly rejected at the 5% significance level, whereas we can never state the same for the Student's t distribution. Hence, an unconditional normal distribution for returns is strongly rejected by the data, while we can not reject a Student's t distribution. Note, in our research, conditional distributions are a key aspect as we are modeling time-varying volatility. Hence, these test results (based on unconditional distributions) may not provide a rigorous motivation for the use of a conditional Student's t distribution.

[4] https://finance.yahoo.com/

[5] This list is inspired from Gorgi et al. (2019) and can be found in Appendix Section 7.1

Table 5: p-values of the Kolmogorov–Smirnov test on the distribution of returns. The 15 stocks are denoted by their ticker symbol and can be seen in the first column.

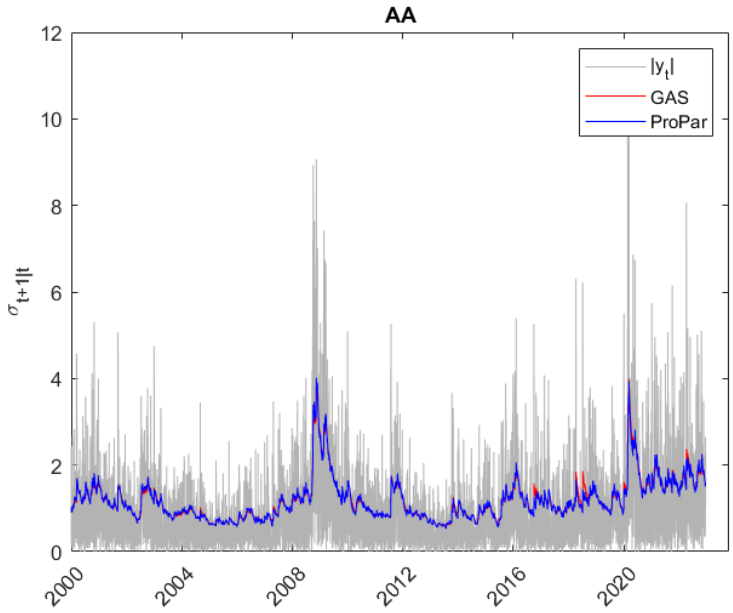
	H0: Normal Dist.	H0: Student's t Dist.
AA	0.000	0.751
AXP	0.000	0.236
BA	0.000	0.946
CAT	0.000	0.639
GE	0.000	0.302
HD	0.000	0.663
HON	0.000	0.546
IBM	0.000	0.916
JPM	0.000	0.430
KO	0.000	0.672
MCD	0.000	0.402
PFE	0.000	0.390
PG	0.000	0.950
WMT	0.000	0.822
XOM	0.000	0.826

4.2 Explicit vs Implicit

We start our comparison between the implicit and explicit stochastic gradient methods — ProPar and GAS respectively — by plotting their 1-day ahead volatility predictions against a proxy for the real volatility on each day. One trivial proxy is the square root of the (forward-looking) daily squared return, which is equivalent to the absolute daily return. In Figure 1, we show such a comparison for the first stock, Alcoa Corporation (AA). Figure 1 reveals that both the ProPar and GAS filter can predict volatility with a somewhat similar accuracy. In Appendix 7.4, we show the comparisons between the ProPar and GAS filter for the 14 other stocks. In some occasions, the GAS filter displays explosive behavior, stemming from the explicit gradient in combination with an exponential formulation for volatility and a sudden shift from a low-to-high volatility period. This finding is in line with Lange et al. (2022) where the authors used daily S&P500 return data. It is also in line with our simulation study where the MSEs for the GAS filter can sometimes be relatively high due to exactly this type of outliers.

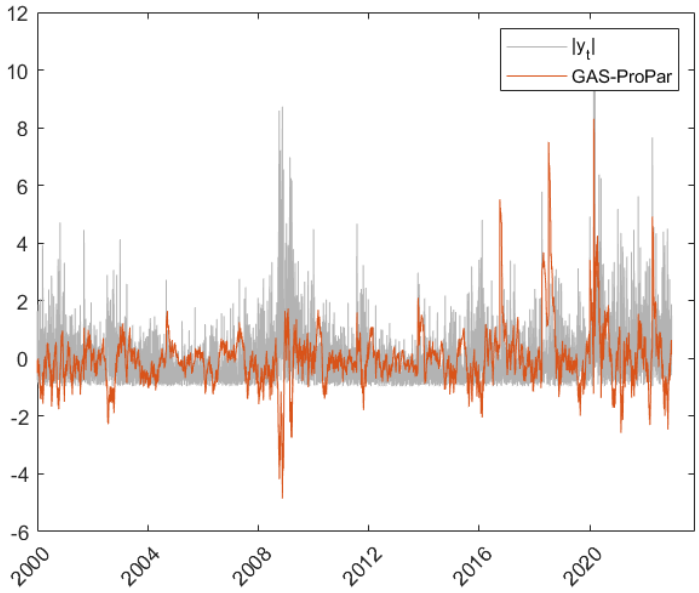
To gain a better view of the relative performance between the two filters, in Figure 2 we display the standardized absolute returns together with the standardized difference between the volatility predictions of the GAS and ProPar filters. From here, we can deduce that the GAS filter underestimated volatility considerably more at the volatility peak around 2008. This year is characterized by a series of events that have significantly increased volatility. Among these events, the most key ones include the downfall of Bears Stearns and Lehman Brothers (following the subprime mortgage crisis), as well as the decision of the House of Representatives to reject the $ 700 Billion Bailout Bill. One explanation for this relative under-estimation is that

Figure 1: Estimated paths for $\sigma_{t+1|t}$ of the ProPar and GAS filter for daily AA returns, from January 2000 until December 2022



the maximum-likelihood estimator for the learning rate η is lower for the GAS filter than for the ProPar filter (0.016 and 0.027 respectively). As a result, the GAS model penalizes volatility updates (and consequently the prediction) more for deviations from the previous volatility prediction. From 2016 to 2020, key macroeconomic developments such as Brexit's referendum, trade wars (especially affecting AA as it operates globally and relies on international trade for its aluminum products), and the COVID-19 pandemic all led to phases of relatively high volatility following low-volatility periods. During this time frame, both the GAS and ProPar filter generally under-estimated volatility, mainly due to their low learning rates. However, from Figure 2, now, the ProPar filter seems to underestimate volatility relatively more. One explanation for this phenomenon could be that the joint effect of an explicit score during these volatility shifts (which causes explosive behavior) and the reducing effect of a low learning rate η led to enhanced predictive abilities. Another, perhaps minor, explanation would be that during these years, the GAS filter benefits more from the information on past volatility shocks due to a higher maximum-likelihood estimator for the persistence ϕ (0.9943 and 0.9926 for the GAS and ProPar filter respectively). Still, throughout the whole data sample for AA, the ProPar filter seems to have a better fit than its explicit counterpart (MSEs of 0.750 and 0.814 respectively).

Figure 2: Standardized volatility prediction differences (GAS minus ProPar) and absolute returns, for daily AA returns, from January 2000 until December 2022



In Table 6, we display the MSEs of both filters' 1, 5, and 20-days ahead predictions for all 15 stocks. Here, we also test for potential significant outperformance using the Diebold-Mariano test. For both filters, it is evident that their accuracy is notably higher when applied to the shortest prediction horizon ($d = 1$) as compared to longer horizons. This is to be expected due to volatility clustering and the consequently diminishing information of the volatility update with respect to the prediction horizon. The ProPar filter furthermore successfully outperforms its explicit counterpart in 35 of the 45 comparisons. Of these 35, 10 are significant meaning that the ProPar filter has a significantly better predictive accuracy for these ten cases. The GAS model, on the other hand, only outperforms 10 times, only two of which are significant. Hence, the ProPar model seems to generally predict volatility better. For the shortest two prediction horizons ($d = 1$ and $d = 5$) this finding holds the most. One explanation for this phenomenon is that the explicit score — used in the GAS filter — is relatively more unstable for short-term dynamics due to more noise, while the implicit score — used in the ProPar filter — gives, per definition, more weight on current information thereby capturing short-term patterns more accurately thus being able to react faster to recent market conditions (in other words, ProPar may be more useful for now-casting, a finding that is consistent with Lange et al. (2022)). The few cases where the GAS filter outperforms are mainly for the longest prediction horizon ($d = 20$). The ProPar filter, in contrast with its simplistic explicit counterpart, requires a more complex solving system, which could ultimately have a worsened effect on these long-term predictions due to perhaps overfitting the updated volatility. This effect is also visible in Table 7, where we display similar results, but now based on the robust QLIKE loss function.

From here, we may also conclude that ProPar outperforms for the shortest prediction horizon, and underperforms for the longest one. For the GAS filter, the cases of outperformance are also mainly significant. This may stem from the fact that the QLIKE loss function heavily favors a positively biased volatility process (compared to a negatively biased (of the same magnitude) volatility process). Namely, the GAS filter is susceptible to yielding explosive volatility behavior due to a large return shock during low-volatility periods, amplified by evaluating the score explicitly.

Table 6: Prediction accuracy analysis of the ProPar and GAS filter: MSEs using $d-$step ahead volatility estimates.

	d=1		d=5		d=20	
	MSE_ProPar	**MSE_GAS**	**MSE_ProPar**	**MSE_GAS**	**MSE_ProPar**	**MSE_GAS**
AA	0.814	0.818	1.458	1.472	3.480	3.477
AXP	0.466	0.512	0.887*	1.154	3.056*	3.696
BA	0.454	0.462	0.975	1.038	3.288	3.517
CAT	0.384	0.395	0.711	0.770	1.813	1.867
GE	0.419	0.427	0.819*	0.874	2.153	2.226
HD	0.364	1.206	0.812	4.142	2.571	8.396
HON	0.357*	0.390	0.808*	0.931	2.655	2.572
IBM	0.270	0.275	0.620	0.623	2.606	1.674*
JPM	0.520	0.530	1.067*	1.118	3.393	3.394
KO	0.164*	0.170	0.356*	0.384	1.079	1.058
MCD	0.213	0.217	0.461	0.476	1.325	1.249
PFE	0.237	0.273	0.496	0.621	1.745	1.545
PG	0.211	2.235	0.561	8.044	1.677	12.386
WMT	0.225*	0.233	0.486*	0.528	1.364	1.304
XOM	0.243	0.242	0.463	0.453	1.497	1.393*

A MSE with an asterisk (*) indicates that the corresponding filter for that stock and prediction horizon statistically outperforms its counterpart, using a 5% significance level.

In summary, in terms of fit, the ProPar filter seems to estimate volatility better. It also provides a relatively more adequate model for volatility predictions under a short prediction horizon. In longer prediction horizons, this outperformance generally no longer holds (and the reverse effect, in favor of the GAS filter is observed). Hence, the implicit evaluation of the score (in the ProPar filter), as opposed to the explicit evaluation (in the GAS filter) can be considered of greater value for volatility estimation and short-term volatility predictions.

Table 7: Prediction accuracy analysis of the ProPar and GAS filter: QLIKEs using $d-$step ahead volatility estimates.

	d=1		d=5		d=20	
	QLIKE_ProPar	QLIKE_GAS	QLIKE_ProPar	QLIKE_GAS	QLIKE_ProPar	QLIKE_GAS
AA	-0.793	-0.792	0.826	0.801*	2.242	2.127*
AXP	-1.597	-1.551	0.119	0.053*	1.832	1.410*
BA	-1.461	-1.445	0.223	0.167*	1.832	1.534*
CAT	-1.432	-1.412	0.252	0.209*	1.856	1.632*
GE	-1.632	-1.613	0.018	-0.014*	1.593	1.373*
HD	-1.756	-1.712	-0.019	-0.103*	1.765	1.274*
HON	-1.781	-1.755	-0.036	-0.157*	1.807	1.214*
IBM	-1.949	-1.913	-0.040	-0.277*	1.958	1.200*
JPM	-1.556	-1.545	0.127	0.041*	1.793	1.389*
KO	-2.467	-2.443	-0.744	-0.799*	1.008	0.716*
MCD	-2.218	-2.202	-0.488	-0.555*	1.255	0.949*
PFE	-1.991	-1.962	-0.182	-0.288*	1.756	1.292*
PG	-2.449	-2.393	-0.729	-0.748	1.041	0.760*
WMT	-2.164	-2.138	-0.430	-0.513*	1.346	0.944*
XOM	-2.010	-2.007	-0.303	-0.367*	1.391	1.117*

A QLIKE with an asterisk (*) indicates that the corresponding filter for that stock and prediction horizon statistically outperforms its counterpart, using a 5% significance level.

Additionally, an alternative means of determining the viability of implicit score evaluation is researched in this paper. Namely, in our I-EGARCH-1 model in Equation 15, we extended the regular EGARCH volatility filter by including an extra parameter (γ), accounting for the effects of the updated volatility on the 1-step ahead predicted volatility. Since this extra layer is constructed using ProPar's implicit stochastic gradient method, comparing the magnitudes of γ and β would yield another insight into the relative importance of this implicit score. Namely, the latter parameter accounts for the effects of the previous volatility prediction, indicating explicit short-term dynamics. To statistically test for claiming significant importance, the Wald test for equal coefficients is applied for each stock. Furthermore, to determine whether the extra parameter yields a model with a significantly better fit, Likelihood-Ratio (LR) tests are also applied. Lastly, Diebold-Mariano tests are applied to establish whether the extra parameter also holds importance in out-of-sample analyses. Table 8 shows that for most stocks, the γ parameter is statistically significant different from β in magnitude. Since the numerator of the Wald test statistic is modeled as $|\beta| - |\gamma|$, the many negative and significant Wald test statistics furthermore show that the implicit persistence parameter γ generally carries more importance. Hence, again, implicitly solving these dynamic conditional score systems yields more favorable results. In Table 8, the LR test statistics are also mostly significant in favor of the implicit-EGARCH model. Hence, at least in terms of fit, the addition of the implicit parameter γ is of statistically significant importance. However, most MSEs are still quite similar, with fourteen (of in total fifteen) MSEs not being significantly different from each other. Still, the only case

where one model is significantly outperforming is the implicit EGARCH model for stock PG. The exact same conclusion can be made using the QLIKE function as a loss function for the DM tests instead.

Table 8: Wald test-statistic, Likelihood-Ratio test-statistic, and loss functions of the 1-step ahead volatility predictions of the I-EGARCH-1 and EGARCH model.

	Wald	**LR**	**MSE_I-EGARCH-1**	**MSE_EGARCH**	**QLIKE_I-EGARCH-1**	**QLIKE_EGARCH**
AA	-439.747*	7.016*	0.813	0.813	-0.792	-0.791
AXP	-42.245*	14.393*	0.465	0.466	-1.598	-1.596
BA	-83.957*	26.664*	0.455	0.456	-1.461	-1.456
CAT	-0.754	22.298*	0.383	0.382	-1.440	-1.437
GE	-18.435*	9.584*	0.418	0.419	-1.633	-1.631
HD	-19.664*	28.677*	0.358	0.365	-1.763	-1.757
HON	-6.719*	8.206*	0.356	0.357	-1.784	-1.782
IBM	1.036	64.295*	0.265	0.266	-1.962	-1.951
JPM	-0.945	4.462*	0.519	0.519	-1.555	-1.554
KO	5.648*	6.947*	0.166	0.166	-2.461	-2.460
MCD	-52.860*	18.008*	0.212	0.213	-2.225	-2.221
PFE	-21.270*	8.462*	0.235	0.236	-1.991	-1.989
PG	160.414*	0.051	0.209*	0.211	-2.503*	-2.471
WMT	2.004*	36.589*	0.223	0.223	-2.174	-2.168
XOM	-372.794*	8.783*	0.243	0.244	-2.006	-2.004

An asterisk (*) indicates statistical significance, using a 5% significance level. For the loss functions, Diebold-Mariano tests are used.

4.3 Student's t distribution

During the ProPar and GAS formulation, a normal distribution for returns is assumed. However, a Student's t-distribution might be more applicable, as discussed in Section 1 and 4.1. In Section 2.2, we provided an implicit model where such a distribution is assumed, IBT-EGARCH. In Table 9, we provide a prediction accuracy analysis for this model, with the explicit GAS model being the benchmark for Diebold-Mariano tests. First of all, it reveals that IBT-EGARCH can make more accurate predictions for a shorter prediction horizon, consistent with Table 6. Based on the QLIKE loss function, IBT-EGARCH significantly outperforms mainly for the medium-to-long-term prediction horizon. Hence, this implicit filter may provide a significantly better framework for predicting volatility at these prediction horizons. For the shortest prediction horizon, it is somewhat harder to state the same, since IBT-EGARCH only significantly outperforms with 3 out of the 15 stocks, using the QLIKE loss function. Still, the results do show that assuming a Student's t-distribution can broadly yield superior results. Nevertheless, it is important to note that not all of the preceding conclusions can be inferred when using the MSE loss function instead. Now, IBT-EGARCH does not outperform GAS at the longest prediction horizon anymore. One explanation for this could be that the GAS filter displays explosive behavior for some observations, driven by large return shocks following low-volatility periods.

Therefore, it may not be a good benchmark for MSE calculation as the loss function difference, which is being calculated during DM tests, consequently also displays explosive behavior (returning high (auto)covariances). Still, for the medium-term volatility predictions ($d = 5$), the implicit IBT-EGARCH model provides a significantly better model than the explicit GAS filter.

Table 9: Prediction accuracy analysis of the IBT-EGARCH filter: QLIKEs and MSEs using $d-$step ahead volatility estimates.

	d=1		**d=5**		**d=20**	
	QLIKE	MSE	QLIKE	MSE	QLIKE	MSE
AA	-0.802	0.823	0.506*	1.412	0.707*	5.188
AXP	-1.637*	0.475	-0.471*	0.850*	-0.811*	4.418
BA	-1.481	0.466	-0.287*	0.976	-0.453*	4.546
CAT	-1.432	0.390	-0.125*	0.657*	0.189*	2.224
GE	-1.643	0.425	-0.455*	0.762*	-0.587*	3.262
HD	-1.788	0.357	-0.614*	0.743	-0.812*	3.347
HON	-1.816	0.358	-0.682*	0.757*	-0.938*	3.393
IBM	-2.033*	0.270	-0.836*	0.515*	-0.786*	2.059
JPM	-1.550	0.530	-0.408*	1.033*	-0.796*	5.297
KO	-2.507	0.165	-1.243*	0.332*	-0.961*	1.241
MCD	-2.240	0.214	-0.973*	0.432*	-0.763*	1.510
PFE	-2.021	0.237	-0.819*	0.447	-0.669*	1.779
PG	-2.525*	0.201	-1.268*	0.509	-0.839*	1.809
WMT	-2.224	0.223	-0.951*	0.432*	-0.679*	1.469
XOM	-2.008	0.247	-0.761*	0.458	-0.639*	2.079

A loss function with an asterisk (*) indicates that the filter for that stock and prediction horizon significantly outperforms the explicit GAS benchmark, using a 5% significance level.

4.4 Other Forecasting Results

In the previous sections, more favorable results were found when solving these dynamical conditional score filters using implicit stochastic gradient methods. However, it remains unknown which of the four implicit filters yields significantly better results. In Table 10 and Table 11, we reveal each ranking (and average ranking) based on the MSE and QLIKE loss function respectively. For the MSE loss function, the two hybrid EGARCH models appear to yield the best 1-step ahead predictions. Among the two models, it appears that I-EGARCH-1 (where both the implicit updating step and the previous prediction term are used) is the superior one. While acknowledging the paramount significance of the implicit stochastic updating step (as discussed in the previous section) it seems that also the previous volatility prediction has considerable information regarding current volatility predictions. Perhaps one could regard this effect to be similar to a shrinkage/regularization term, in the sense that it provides a more stable volatility forecast instead of relying purely on the (possibly overfitting) updated volatility. The model that is based on the Student's t-distribution, IBT-EGARCH, also suggests some outperformance. For certain stocks, it outperforms ProPar and may be regarded as the best model. Still, on average,

the ProPar filter (using a normal distribution for returns) is slightly better. This could stem from the fact that although the use of a normal distribution can be rejected, it does ensure a closed-form unique solution to the volatility update, while in the Student's t-based model, the updated volatility estimates have to be approximated using the numerical Newton-Raphson algorithm.

Nonetheless, for the QLIKE loss function, the IBT-EGARCH dominantly appears to be the overall best model. The rankings for the ProPar and GAS filter stay relatively the same, while the hybrid EGARCH models are now overtaken by the IBT-EGARCH filter. As the QLIKE function is more robust to outliers, this may be an explaining factor why the hybrid models are now somewhat performing worse. Namely, next to the GAS filter, these two models are the only ones where the previous volatility prediction is explicitly used in the current one. For the MSE case, this could be regarded as a shrinkage term, since MSE is susceptible to outliers. But now that a robust loss function is used, this concern is rendered irrelevant, diminishing the value of this explicit variable. Now, the IBT-EGARCH model, which assumes a Student's t-distribution for returns, seems to be the superior model. Thus, assuming such a distribution for (leptokurtic) returns not only fits the data significantly better than a normal distribution, but it also seems to be the better choice when using score-driven models to predict volatility.

Table 10: Ranking based on MSEs of 1-day ahead volatility predictions. Here, '1' represents the best filter, and '5' denotes the least favorable filter.

	ProPar	**GAS**	**IBT-EGARCH**	**I-EGARCH-1**	**I-EGARCH-2**
AA	3	4	5	2	1
AXP	3	5	4	1	2
BA	1	4	5	2	3
CAT	2	5	4	1	3
GE	3	5	4	1	2
HD	4	5	1	3	2
HON	2	5	4	1	3
IBM	3	5	4	2	1
JPM	3	5	4	1	2
KO	1	5	2	4	3
MCD	3	5	4	1	2
PFE	3	5	4	2	1
PG	3	5	1	2	4
WMT	4	5	2	1	3
XOM	3	1	5	2	4
Average	**2.733**	**4.600**	**3.533**	**1.733**	**2.400**

Table 11: Ranking based on QLIKEs of 1-day ahead volatility predictions. Here, '1' represents the best filter, and '5' denotes the least favorable filter.

	ProPar	**GAS**	**IBT-EGARCH**	**-EGARCH-1**	**I-EGARCH-2**
AA	2	5	1	4	3
AXP	4	5	1	2	3
BA	2	5	1	3	4
CAT	4	5	1	2	3
GE	2	5	1	3	4
HD	4	5	1	2	3
HON	4	5	1	2	3
IBM	4	5	1	2	3
JPM	2	5	1	3	4
KO	2	5	1	3	4
MCD	4	5	1	2	3
PFE	2	5	1	3	4
PG	4	5	1	2	3
WMT	4	5	1	2	3
XOM	2	3	1	4	5
Average	**3.067**	**4.867**	**1**	**2.600**	**3.467**

5 High-Frequency

5.1 Data

For the high-frequency case, we also extract bid and ask-price quotes data from the (monthly updated) Trades and Quotes (TAQ) database via Wharton Research Data Services (WRDS). The time-stamp precision of this database[6] is one second. Using these high-frequent stock price data, we consider three months of data (1 January 2013 - 31 March 2013) for the same 15 stocks. Cleaning the resulting dataset is highly necessary for most research in this field (see e.g. Gorgi et al. (2019), Brownlees & Gallo (2006), Barndorff-Nielsen et al. (2009)). In cleaning the TAQ data, we follow the guidelines of Barndorff-Nielsen et al. (2009) which is implemented in the R package **highfrequency** of Boudt et al. (2022). For example, we only consider data that are recorded at standard market periods (9:30 AM- 4:00 PM), delete entries with quotes equal to 0, retain data from one exchange (the New York Stock Exchange (NSYE) in our research), and remove quotes that generate a negative spread. We compute mid-prices as the average between the best-bid and best-ask price for each second [7]. These mid-prices are consequently used to calculate intraday returns in a similar way as for daily prices: taking the natural logarithm, taking differences, and multiplying by 100 to achieve percentage intraday returns. With 64 trading days considered, in

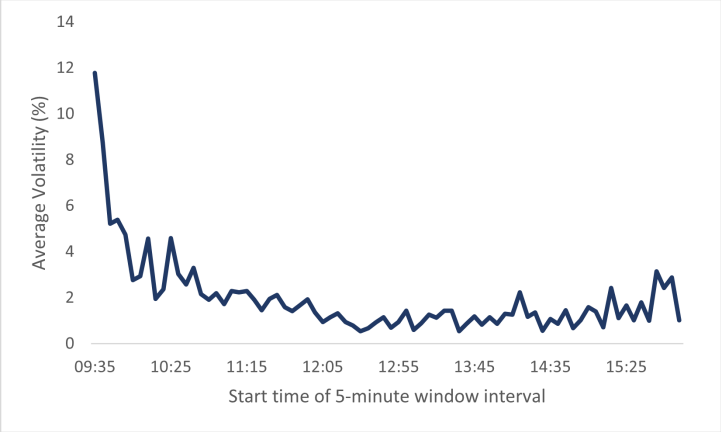
[6]TAQ also provides data with an even higher frequency (milliseconds and even microseconds starting in April 2015), but this is outside the scope of the research as it would cost too much computer memory or would substantially reduce our data-range.

[7]Note, the choice of using quoted mid-prices is not trivial. Instead, an alternative would be to use the last traded price as a proxy for the real price of stocks, which is also available in the TAQ database. However, we deem the quoted mid-price to contain more valuable information as order book quotes are more continuously updated, following more updated activity. The last traded price can be considered to be a more discretization of the stock price process.

total 64 $\times$ 6.5 hours $\times$ 60 minutes/5 minutes=4992 5-minute returns per stock are recorded (so in total 74,880 observations in our cross-section of stocks).

In our research, the realized (co)variance estimator (based on the sum of squared intraday returns) is a key aspect of our high-frequency models. More details about this methodology will follow in this section. To gain insight into these intraday returns, in Figure 3 we display the average volatility of the different 5-minute returns per day for the first stock, Alcoa Corporation (AA). As we are dealing with 6.5 trading hours, this corresponds to 6.5 $\times$ 60/5=78 unique 5-minute windows, each of which we take the average for. For every unique window, we calculate its volatility as the absolute 5-minute return (which is equivalent to the square root of the squared 5-minute return) and annualize it by multiplying it with the square root of 78 x 252 (the number of 5-minute window intervals in a day times the number of trading days in a year). From Figure 3, we observe a skewed U-shape, indicating that, on average, volatility is relatively high in the morning when the NYSE opens but also somewhat higher when it closes in the afternoon. At lunchtime, intraday volatility appears to be substantially lower. These findings are in line with previous research on high-frequency data (see e.g. Wood et al. (1985) and Andersen et al. (2001)).

Figure 3: Average Alcoa Corporation (AA) volatilities across 5-minute time intervals, Jan 1, 2013 - Mar 28, 2013. Average volatility is calculated as the average absolute 5-minute return per unique 5-minute window (starting from the first moment of each trading day).



5.2 ProPar-HF

In this section, we will provide an extension to the ProPar filter by researching the use of high-frequency financial data to predict daily volatility. Namely, by using sub-batches of intraday returns, we update daily volatilities. Let us denote the i^{th} 5-minute return during day t as $y_{t,i}$, for all $i = 1, \ldots, 78$ (# unique 5-minute intervals), $t = 1, \ldots, T$. We consider the following

optimization framework for each sample of 78 i.i.d. normal random intraday returns:

$$\begin{aligned} h_{t|t} &= \underset{h\in\mathbb{R}}{\operatorname{argmax}} \left[\log p\left(y_{t,1}, y_{t,2}, \ldots, y_{t,78} \mid \exp(h)\right) - \frac{1}{2}\left(h - h_{t|t-1}\right)^2 P \right] \\ h_{t|t} &= \underset{h\in\mathbb{R}}{\operatorname{argmax}} \left[\log \prod_{i=1}^{78} p\left(y_{t,i} \mid \exp(h)\right) - \frac{1}{2}\left(h - h_{t|t-1}\right)^2 P \right] \\ &= \underset{h\in\mathbb{R}}{\operatorname{argmax}} \left[\sum_{i=1}^{78} \log p\left(y_{t,i} \mid \exp(h)\right) - \frac{1}{2}\left(h - h_{t|t-1}\right)^2 P \right] \\ p\left(y_{t,i} \mid \exp(h)\right) &= \frac{1}{\exp(h_t)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y_{t,i} - \mu}{\exp(h_t)} \right)^2 \right). \end{aligned} \tag{28}$$

The drawback of this setup is that it implicitly assumes that all 78 intraday returns have the same distribution, including the same volatility. However, we observed in Figure 3 a (skewed) U-shape for intraday volatilities. Therefore, we control for this phenomenon by including the average U-shape in the logarithmic observation densities, yielding scaling factors w_i. We propose to construct w_i to account for the (cross-sectional) average U-shape:

$$w_i = \frac{1}{K}\frac{1}{T}\sum_{k=1}^{K}\sum_{t=1}^{T} |y_{t,i,k}|, \tag{29}$$

where $K = 15$ denotes the number of stocks and $y_{t,i,k}$ is the i'th 5-minute return at time t for stock k. To ensure that the sum of weights sums up to 78 (equivalent to the situation without weights, or, equivalently, $w_i = 1$ for all $i = 1, \ldots, 78$), we scale all weights accordingly. Now, the optimization framework becomes:

$$\begin{aligned} h_{t|t} &= \underset{h\in\mathbb{R}}{\operatorname{argmax}} \left[\sum_{i=1}^{78} \log p\left(y_{t,i} \mid w_i \exp(h)\right) - \frac{1}{2}\left(h - h_{t|t-1}\right)^2 P \right] \\ p\left(y_{t,i} \mid w_i \exp(h)\right) &= \frac{1}{w_i \exp(h)\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y_{t,i} - \mu}{w_i \exp(h)} \right)^2 \right). \end{aligned} \tag{30}$$

Using the optimization setup in Equation 30, we analytically solve the value for $h_{t|t}$ into a closed-form solution similar to the ProPar model. The solution is given by:

$$h_{t|t} = h_{t|t-1} + \eta \left[\sum_{i=1}^{78} \left(\frac{y_{t,i} - \mu}{w_i \exp\left(h_{t|t}\right)} \right)^2 - 1 \right], \tag{31}$$

which, again, can be solved using the Lambert W function. Namely, $h_{t|t} = h_{t|t-1} - \sum_{i=1}^{78}\eta + \frac{1}{2}W_0\left[2\eta \exp\left(-2(h_{t|t-1} - \sum_{i=1}^{78}\eta) \right) \sum_{i=1}^{78}\left(\frac{y_{t,i}-\mu}{w_i} \right)^2 \right]$. Let us denote this model as ProPar-High-Frequency (ProPar-HF). Its explicit GAS counterpart can be denoted as GAS-High-Frequency

(GAS-HF). This is straightforwardly modeled as:

$$h_{t|t} = h_{t|t-1} + \eta \left[\sum_{i=1}^{78} \left(\frac{y_{t,i} - \mu}{w_i \exp\left(h_{t|t-1}\right)} \right)^2 - 1 \right], \tag{32}$$

where the right-hand side is immediately computable and thus requires no analytical solving system. Using Equation 6, 1-step ahead predictions can still be calculated [8].

5.3 HEAVY-ProPar-tF

Since our dataset consists of multiple stocks, we are also interested in the multivariate case. That is, let $\boldsymbol{y}_t$ now be a $(K \times 1)$ vector, consisting K close-to-close stock price returns. In our subsequent model, we modify the HEAVY-GAS-HAR-tF model of Opschoor et al. (2018) to account for implicit information. Among seven other multivariate models, this model was found to be generally the most accurate in providing the best multivariate density forecasts for realized covariance matrices and returns (using 1-day ahead forecasts). There, the authors constructed a score-driven multivariate location/scale model for daily returns and (realized) covariance matrices. To capture the dynamics behind the time-varying covariance matrix $\boldsymbol{V}_t$, the authors use the multivariate GAS model, which is also an explicit stochastic gradient method. Hence, we modify their optimization and formulations by implicitly updating this parameter and by making use of ProPar's proximal-point feature, utilizing a (squared) Frobenius norm for the penalty-weighted covariance matrix difference. Namely, we consider the following optimization framework:

$$\boldsymbol{V}_{t|t} = \underset{\boldsymbol{V} \in R^{K \times K}}{\operatorname{argmax}} \left[\log p(\boldsymbol{y}_t \mid \boldsymbol{V}) + \log p(\boldsymbol{RC}_t \mid \boldsymbol{V}) - \frac{1}{2} \left\| \sqrt{\boldsymbol{P}} (\boldsymbol{V} - \boldsymbol{V}_{t|t-1}) \right\|_F^2 \right], \tag{33}$$

where $\boldsymbol{RC}_t$ is the daily realized covariance matrix (where e.g. the i'th diagonal is calculated as the sum of squared 5-minute returns for the i'th stock). Moreover, the matrix $\boldsymbol{P}$ represents a $K \times K$ weighting matrix, and $\|\boldsymbol{A}\|_F^2$ is the squared Frobenius norm for matrices. The rationale behind the use of the Frobenius norm follows from the fact that, just as for the ℓ_2 norm for vectors, it is positively related to the magnitudes of the elements. As a result, the concept of penalty matrix $\boldsymbol{P}$ slightly changes, as it now weighs the difference between $K \times K$ covariance matrices (whereas in univariate models it weighed the $K \times 1$ volatility vector). Since the Frobenius norm is defined as the square root of the sum of singular values of the matrix, we can rewrite our optimization

[8] Note, we multiply the final updated and predicted volatilities by factor $\sqrt{78}$ to transform the intraday volatility into one daily volatility, which will also be used in the forecasting analysis.

framework into:

$$\boldsymbol{V}_{t|t} = \underset{\boldsymbol{V} \in R^{K \times K}}{\operatorname{argmax}} \left[\log p(\boldsymbol{y}_t \mid \boldsymbol{V}) + \log p(\boldsymbol{RC}_t \mid \boldsymbol{V}) - \frac{1}{2} \sum_{i=1}^{K} \lambda_i \left(\boldsymbol{\Delta}^T \boldsymbol{\Delta} \right) \right], \tag{34}$$

where $\boldsymbol{\Delta} = \sqrt{\boldsymbol{P}}(\boldsymbol{V} - \boldsymbol{V}_{t|t-1})$. Here, the term $\lambda_i(\boldsymbol{A})$ denotes the i'th eigenvalue of matrix $\boldsymbol{A}$. Furthermore, vector observation $\boldsymbol{y}_t$ is assumed to be fat-tailed and conditionally follows a Student's t-distribution with $\nu > 2$ degrees of freedom. Therefore, it has the following conditional observation density:

$$p(\boldsymbol{y}_t \mid \boldsymbol{V}_t) = \frac{\Gamma((\nu + K)/2)}{\Gamma(\nu/2)[(\nu - 2)\pi]^{K/2} |\boldsymbol{V}_t|^{1/2}} \left(1 + \frac{\boldsymbol{y}_t' \boldsymbol{V}_t^{-1} \boldsymbol{y}_t}{\nu - 2} \right)^{-(\nu+K)/2}, \tag{35}$$

where $|\boldsymbol{A}|$ denotes the determinant of $\boldsymbol{A}$. Throughout the literature, $\boldsymbol{RC}_t$ often is assumed to be Wishart distributed (see e.g. Gorgi et al. (2019)). However, in Opschoor et al. (2018), the authors significantly reject such a distribution and advocated for the assumption of a scaled multivariate F-distribution instead. Given that our dataset is a subset of the 30 stocks investigated by the authors (thereby rendering it comparable), we likewise employ the scaled multivariate F-distribution in our analysis. With ν_1 and $\nu_2 > K - 1$ degrees of freedom, it has the following conditional observation density:

$$p(\boldsymbol{RC}_t \mid \boldsymbol{V}_t) = \frac{\Gamma_K((\nu_1 + \nu_2)/2)}{\Gamma_K(\nu_1/2)\Gamma_K(\nu_2/2)} \times \frac{|\frac{\nu_1}{\nu_2 - K - 1} \boldsymbol{V}_t^{-1}|^{\nu_1/2} \times |\boldsymbol{RC}_t|^{(\nu_1 - K - 1)/2}}{|\boldsymbol{I}_K + \frac{\nu_1}{\nu_2 - K - 1} \boldsymbol{V}_t^{-1} \boldsymbol{RC}_t|^{(\nu_1 + \nu_2)/2}}, \tag{36}$$

where $\Gamma_p(a) = \pi^{p(p-1)/4} \prod_{i=1}^{p} \Gamma\left(a + (1 - i)/2\right)$ (the multivariate Gamma function), and $\boldsymbol{I}_K$ is a $(K \times K)$ identity matrix.

To implicitly solve Equation 34, we formulate the first-order condition as:

$$\nabla(\boldsymbol{y}_t \mid \boldsymbol{V}_{t|t}) + \nabla(\boldsymbol{RC}_t \mid \boldsymbol{V}_{t|t}) - \frac{1}{2} \left(\sum_{i=1}^{K} \boldsymbol{u}_i \left(\boldsymbol{\Delta}^T \boldsymbol{\Delta} \right) \boldsymbol{u}_i \left(\boldsymbol{\Delta}^T \boldsymbol{\Delta} \right)^T \right) = \boldsymbol{0}_{K \times K}, \tag{37}$$

where $\boldsymbol{u}_i(\boldsymbol{A})$ denotes the normalized $(K \times 1)$ eigenvector corresponding to the i'th eigenvalue of matrix $\boldsymbol{A}$. The vector calculus step behind the relationship of eigenvectors and the derivative of eigenvalues with respect to its matrix is based on Theorem 1 in Magnus (1985). The score

formulas are derived in Opschoor et al. (2018) and take on the following form:

$$\nabla(\boldsymbol{y}_t \mid \boldsymbol{V}_{t|t}) = \frac{1}{2}\boldsymbol{V}_{t|t}^{-1}\left(\frac{\nu+K}{\nu-2+\boldsymbol{y}_t'\boldsymbol{V}_{t|t}^{-1}\boldsymbol{y}_t}\boldsymbol{y}_t\boldsymbol{y}_t' - \boldsymbol{V}_{t|t}\right)\boldsymbol{V}_{t|t}^{-1}$$
$$\nabla(\boldsymbol{RC}_t \mid \boldsymbol{V}_{t|t}) = \frac{\nu_1}{2}\boldsymbol{V}_{t|t}^{-1}\left[\frac{\nu_1+\nu_2}{\nu_2-K-1}\boldsymbol{RC}_t\left(\boldsymbol{I}_{K\times K}+\frac{\nu_1}{\nu_2-K-1}\boldsymbol{V}_{t|t}^{-1}\boldsymbol{RC}_t\right)^{-1} - \boldsymbol{V}_{t|t}\right]\boldsymbol{V}_{t|t}^{-1}. \tag{38}$$

Since Equation 37 is highly complex with respect to $\boldsymbol{V}_{t|t}$, obtaining a unique closed-form solution may be infeasible. Therefore, we numerically optimize the maximization problem in Equation 34 using Gradient Descent. Namely, since the gradients in Equation 38 are highly complex with respect to $\boldsymbol{V}_{t|t}$, obtaining a closed-form Hessian matrix may prove to be impractical. Also, BFGS (Broyden-Fletcher-Goldfarb-Shanno), a conventional quasi-Newton optimization algorithm may be used as it still uses second-order information by approximating the Hessian.

Below, the algorithm is displayed. Here, we initialize $\boldsymbol{V}_0$ by $\boldsymbol{V}_{t|t-1}$. Also, $f(\boldsymbol{V}_0)$ denotes the maximization formula in Equation 34 evaluated at the initialization, and $f'(\boldsymbol{V}_0)$ is its gradient, formulated in Equation 37 [9]. The step size $\delta > 0$ is updated at each iteration by applying a backtracking line search method with the Armijo condition as a stopping criterion. Here, we start with an initial step size equal to 10, with the decrease parameter and decrease factor both equal to 0.5. Moreover, the optimization scheme is repeated until one of the following stopping criteria is met:

1. The maximum number of iterations $i_{\max}$, which is set to 10^3, is reached. Note, in our research, the average number of iterations only is 1.729, indicating relatively fast convergence.

2. The Frobenius matrix norm of the difference between the previous and updated gradient in Equation 37 becomes smaller than the tolerance value $\epsilon = 10^{-3}$.

Algorithm 2 Gradient Descent algorithm

Input: Gradient $f'(\boldsymbol{V})$, initial point $\boldsymbol{V}_0$, step size δ

Output: $\hat{\boldsymbol{V}}$ s.t. $f'(\hat{\boldsymbol{V}}) \approx \boldsymbol{0}_{K\times K}$

$\boldsymbol{x} \leftarrow -\delta \boldsymbol{I}_{K\times K} f'(\boldsymbol{V}_0)$ ▷ Updating argument

$\boldsymbol{V} \leftarrow \boldsymbol{V}_0 + \boldsymbol{x}$ ▷ Updating Step

$\boldsymbol{V}_0 \leftarrow \boldsymbol{V}$ ▷ Re-initialize $\boldsymbol{V}_0$ to start the process again

[9] Since the algorithm is modeled to find a minimum, we minimize the negative of Equation 34, which is equivalent to maximizing the original formula. Consequently, the first-order condition in Equation 37 should be multiplied by -1 as well.

In the HEAVY-GAS-HAR-tF filter, Opschoor et al. (2018) made use of the following explicit GAS filter:

$$\boldsymbol{V}_{t+1|t} = \boldsymbol{\Omega} + \alpha \boldsymbol{S}_{t|t-1} + \beta \boldsymbol{V}_{t|t-1} + \beta_2 \hat{\boldsymbol{V}}_{t|t-1} + \beta_3 \bar{\boldsymbol{V}}_{t|t-1}, \tag{39}$$

where α, β, β_2, and β_3 are scalars that we optimize by maximum likelihood, $\boldsymbol{\Omega}$ is a $K \times K$ matrix of real-valued intercepts, $\boldsymbol{S}_{t|t-1}$ is an explicitly scaled score (as defined in the following paragraph), $\hat{\boldsymbol{V}}_{t|t-1}$ is the average predicted covariance matrix using the last 5 days, and where $\bar{\boldsymbol{V}}_{t|t-1}$ is the average predicted covariance matrix using the last 20 days.

In our research, we alter this prediction filter by utilizing and focusing on the updated/implicit information. Namely, since we can update our score by using the updated covariance matrix, these may be more informative for the current covariance matrix prediction. Mathematically, the prediction step for the HEAVY-ProPar-tF model becomes:

$$\boldsymbol{V}_{t+1|t} = \boldsymbol{\Omega} + \alpha \boldsymbol{S}_{t|t} + \beta \boldsymbol{V}_{t|t}, \tag{40}$$

where the implicit $\boldsymbol{S}_{t|t}$ (instead of the explicit $\boldsymbol{S}_{t|t-1}$) is a scaled score as defined by Opschoor et al. (2018), and takes on the following form:

$$\begin{aligned}
\text{vec}(\boldsymbol{S}_{t|t}) &= \frac{2}{\nu_1 + 1} \text{vec}\left(\boldsymbol{V}_{t|t}\left(\nabla(\boldsymbol{y}_t \mid \boldsymbol{V}_{t|t}) + \nabla(\boldsymbol{RC}_t \mid \boldsymbol{V}_{t|t})\right)\boldsymbol{V}_{t|t}\right) \\
\Leftrightarrow \boldsymbol{S}_{t|t} &= \frac{\frac{\nu+K}{\nu-2+\boldsymbol{y}_t'\boldsymbol{V}_{t|t}^{-1}\boldsymbol{y}_t}\boldsymbol{y}_t\boldsymbol{y}_t' - \boldsymbol{V}_{t|t}}{\nu_1 + 1} + \frac{\nu_1}{\nu_1 + 1}\left[\frac{\nu_1 + \nu_2}{\nu_2 - K - 1}\boldsymbol{RC}_t\left(\boldsymbol{I}_{K\times K} + \frac{\nu_1}{\nu_2 - K - 1}\boldsymbol{V}_{t|t}^{-1}\boldsymbol{RC}_t\right)^{-1} - \boldsymbol{V}_{t|t}\right].
\end{aligned} \tag{41}$$

For the explicit counterpart (HEAVY-GAS-HAR-tF), the score $\boldsymbol{S}_{t|t-1}$ is modeled explicitly and can be calculated using Equation 41 where the predicted covariance matrix $\boldsymbol{V}_{t|t-1}$ is used instead of the updated covariance matrix $\boldsymbol{V}_{t|t}$.

In total, the maximum likelihood estimation for our novel implicit model will thus search for optimal parameter values for ν, ν_1, ν_2, α, β, $\boldsymbol{\Omega}$ and $\boldsymbol{P}$. Since $\boldsymbol{P}$ is a symmetric $K \times K$ weighting matrix, we have $\frac{K(K+1)}{2}$ unique elements/parameters. And in theory, the $K \times K$ matrix $\boldsymbol{\Omega}$ can also be estimated using K^2 unique elements. However, to reduce the number of parameters in the maximum likelihood estimation, we construct $\boldsymbol{P}$ to be a diagonal positive definite penalty matrix with initial/starting elements taken from the univariate models as an educated guess and follow Opschoor et al. (2018) by replacing $\boldsymbol{\Omega}$ with $(1-\beta)\hat{\boldsymbol{RC}}$, where $\hat{\boldsymbol{RC}}$ is the sample mean of $\boldsymbol{RC}_t$. This is argued to be a valid estimator for $\boldsymbol{\Omega}$ as long as $\beta < 1$, which we impose in our maximum likelihood estimation. Altogether, we label this novel model by the

HEAVY-ProPar-tF model.

Note, in Proposition 2 of Opschoor et al. (2018), the authors have proved that the predicted covariance matrices resulting from Equation 39 are positive definite under some assumptions. This proposition is extremely useful regarding the ability to always compute the likelihood function (evaluated at the predicted covariance matrix), which could be problematic since it is dependent on inverting covariance matrices. First of all, in order for the proof to hold for our implicit model, we need $\beta > \alpha > 0$, which we can easily impose during the maximum likelihood estimation. Secondly, the realized covariance matrices are assumed positive semi-definite. This assumption is per definition valid, given that it is a symmetric covariance matrix. During their proof, the authors also made use of the fact that the previous predicted covariance matrix is positive definite, given an initial positive definite covariance matrix $\boldsymbol{V}_0$. However, since we are evaluating at the updated covariance matrix $\boldsymbol{V}_{t|t}$ instead of $\boldsymbol{V}_{t|t-1}$, this step may not always be valid. When that is the case, we could transform the updated matrix as:

$\boldsymbol{V}_{t|t} \leftarrow \boldsymbol{V}_{t|t} + \psi \boldsymbol{I}_{K\times K}$, where ψ is equal to the smallest negative eigenvalue of $\boldsymbol{V}_{t|t}$ plus a small arbitrary value (e.g. in our research equal to 10^{-6}). The resulting matrix is positive definite (Cailliez, 1983). However, during implementation, this transformation function was found to never be invoked, as it only gets exercised when the updated matrix is not positive definite. Therefore, we conjecture (at least with our dataset) that the resulting sequence of updated covariance matrices are positive definite. To further validate this hypothesis, a formal proof is required. For example, a proof by induction could be applied, specifically demonstrating that each step in the algorithm produces a positive definite matrix. This aspect remains an open question for future research.

5.4 Forecasting

Using high-frequency data, we analyze our forecasts slightly differently. First of all, we compare the 1-day ahead predictions of the ProPar-HF model against its explicit GAS-HF counterpart but also against the standard ProPar model that does not include this intraday information. To test the validity of the null hypothesis that states equal predictive ability between two models, a volatility proxy is needed. A commonly utilized proxy in the literature is the daily squared return (which we use in the aforementioned general empirical study). While this is shown to be an unbiased estimator of ex-post volatility, Andersen & Bollerslev (1998) have argued that this estimator is also prone to high levels of noise. Therefore, in Andersen et al. (1999), the authors make use of the sum of squared intraday returns. Given that we work with 78 5-minute returns per trading day, we can apply that methodology and thus consider the following forward-looking

1-day ahead (realized) volatility proxy[10]:

$$h_{t+1} = \sum_{i=1}^{78} y_{t+1,i}^2. \tag{42}$$

We again conduct Diebold-Mariano tests (using the two loss functions) to establish whether potential outperformance is statistically significant.

Lastly, for our multivariate models, we compare the 1-step ahead covariance matrix predictions using differences in forecasted log observation densities as a loss function for the Diebold-Mariano test (similar to Opschoor et al. (2018) where the authors label the log densities itself as scores). As a result, a significantly higher mean score would now indicate that the corresponding model has a significantly superior forecast performance, based on a DM test for equal predictive ability [11].

5.5 Results

In this section, we investigate the potential power of ProPar on high-frequency data. Specifically, we study the performance of implicitly and explicitly evaluating the score, using both univariate and multivariate models.

As discussed, we can generate two volatility proxies. Based on daily return data, we can estimate the variance proxy as before, namely the squared daily return. Based on intraday returns, we can estimate the realized variance traditionally as the sum of squared intraday returns. In Figure 4, we display the resulting two (annualized) volatility paths for stock AA. What stands out is that the annualized volatility based on daily returns sometimes can be extremely low and high. Particularly, instances of exceptionally low-volatility estimates are observed frequently. This suggests that on certain days, the annualized volatility could approach or even reach zero. Conversely, the annualized realized volatility proxy yields a range of roughly 10-25% which may be considered more reasonable. In Appendix Section 7.4 the volatility paths for the other stocks are shown, and similar conclusions can be drawn. Therefore, we believe that this estimator may provide a more accurate proxy for the true daily volatility. Still, we study the predictive ability of ProPar-HF based on both volatility proxies.

In Table 12, the 1-day ahead forecasting results are revealed for the ProPar model utilizing high-frequency data (ProPar-HF). Regardless of the volatility proxy used, the implicit ProPar-HF significantly outperforms the explicit GAS-HF in many cases. Therefore, when considering intraday returns to update daily volatilities, implicitly (instead of explicitly) evaluating the

[10] Straightforward extensions could be the use a higher frequency to compute squared returns or noise-corrected RV estimators, such as the two scales realized variance of Aït-Sahalia et al. (2011).

[11] This is contrary to the MSE and QLIKE loss function where a superior model is assumed to have the lowest MSE or QLIKE.

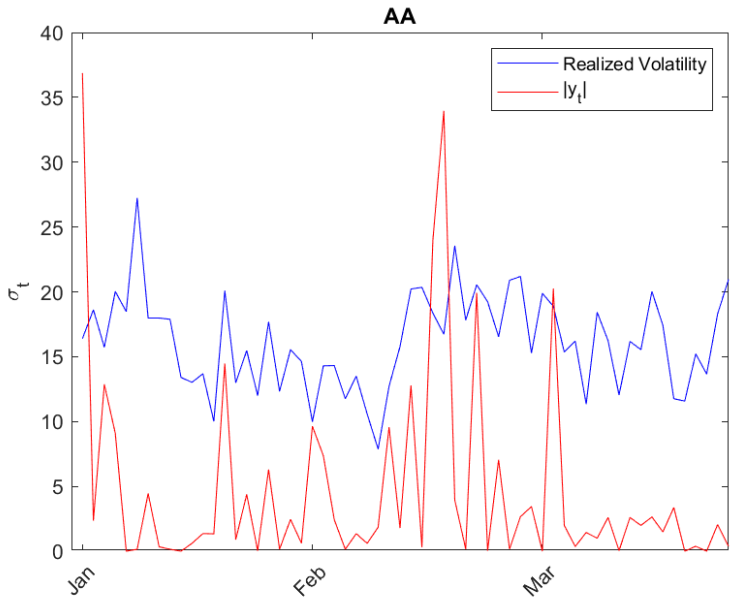


Figure 4: Two daily volatility paths of Alcoa Corporation (AA) using the realized volatility and absolute daily return.

score yields far more favorable results. A conclusion that is repeatedly drawn in this paper. Furthermore, when considering the realized volatility as the proxy for the true volatility, ProPar-HF also significantly outperforms the standard ProPar model when using the MSE as a loss function. Both the use of the MSE as a loss function and the realized volatility as a proxy is more reasonable in this setting. The latter one is already explained using Figure 4. For the former one, note that the QLIKE loss function is asymmetric and consequently heavily penalizes models that yield generally higher volatility estimates, regardless of the volatility proxy. Then, given that the ProPar-HF model is trained on intraday returns (that yield generally higher (realized) daily volatility estimates than daily volatility estimates based on absolute daily returns) it follows that its daily volatility estimates are also generally higher. As a result, the QLIKEs of the ProPar-HF model are higher, despite the fact that its volatility estimations are closer to the pseudo-true volatilities. Also, the ProPar-HF filter (while significantly outperforming its explicit GAS-HF counterpart) fails to outperform the standard ProPar model when considering the absolute daily return as a proxy for the volatility. For the cases using the QLIKE loss function, this can be explained as before. For the cases using the MSE loss function, we believe that this can be attributed to the fact that the ProPar-HF model is trained on intraday returns (which Figure 4 showed to yield a substantially different volatility path) such that its volatility estimates are relatively far from the pseudo-true volatility path.

Table 12: Prediction accuracy analysis of the ProPar-HF filter: MSEs and QLIKEs using 1-step ahead volatility estimates.

	Proxy: Absolute Daily Return		**Proxy: Realized Vol**	
	MSE	QLIKE	MSE	QLIKE
AA	0.533*	-0.267*	$0.051^{*\triangle}$	-1.075*
AXP	0.280*	-0.762*	$0.065^{*\triangle}$	-1.704*
BA	0.346	-0.566*	$0.100^{\triangle}$	-1.297*
CAT	0.337*	-0.614*	$0.085^{\triangle}$	-1.676*
GE	0.322	-0.955*	$0.133^{\triangle}$	-2.024*
HD	0.344*	-0.784*	$0.048^{*\triangle}$	-1.649*
HON	0.252*	-0.935*	$0.040^{\triangle}$	-1.850*
IBM	0.200	-1.313*	$0.030^{\triangle}$	-2.051*
JPM	0.445*	-0.358*	$0.064^{\triangle}$	-1.252*
KO	0.238*	-1.022*	$0.034^{*\triangle}$	-1.838*
MCD	0.131*	-1.579*	$0.041^{*\triangle}$	-2.578*
PFE	0.238*	-0.952*	$0.040^{*\triangle}$	-1.751*
PG	0.182	-1.422*	$0.027^{\triangle}$	-2.234*
WMT	0.345	-0.725*	$0.086^{\triangle}$	-1.548*
XOM	0.228	-1.012*	$0.026^{\triangle}$	-1.861*

A loss function with an asterisk (*) indicates that ProPar-HF significantly outperforms the explicit GAS-HF benchmark, using a 5% significance level. In a similar way, a triangle ($^{\triangle}$) indicates significant outperformance of the ProPar-HF filter against the normal ProPar filter that is trained on daily returns. Here, we use DM tests to statistically test for outperformance and consider both the forward-looking daily absolute returns and realized volatility as a proxy for the true volatility.

In summary, when considering the realized volatility estimate as a proxy for the true volatility (which we believe is more reasonable), ProPar-HF provides a significantly better model compared to its explicit GAS-HF counterpart and the standard ProPar model.

Now, we turn to our multivariate models. Here, we compare the HEAVY-GAS-HAR-tF model of Opschoor et al. (2018) against our proposed implicit counterpart (HEAVY-ProPar-tF). To differentiate between the predictive ability of both multivariate models, we compute 1-day ahead prediction log observation densities (labeled as the score in this context) and take differences to construct the loss function used in the DM test, as explained in the previous section. In Table 13, we display the maximum likelihood estimation output, the optimal total log-likelihoods, and the mean scores using 1-day ahead forecasted densities. Following relatively low standard errors, maximum likelihood thus seems to generate parameters efficiently. This finding is in line with our univariate models. Furthermore, the three degrees of freedom parameters (ν, ν_1, ν_2) for both models are relatively low, suggesting that not only returns (assumed to be conditionally Student's t distributed with ν degrees of freedom) are fat-tailed, but also the realized covariances (that are assumed to be scaled F-distributed with ν_1 and ν_2 degrees of freedom). This is in line with the findings in Section 4.1 where we already found evidence for this fat-tailedness of return data as the unconditional empirical Student's t distribution could never be rejected whereas a normal distribution was always rejected. Note, the values for ν_1

may seem high at first sight (suggesting no evidence for fat-tailed realized covariance matrices), but only when $\nu_1 \to \infty$ do we get that the scaled F-distribution converges to the Wishart distribution. As a result, both models strongly account for the effects of return and realized covariance outliers. Additionally, recall that α solely accounts for information on the explicit score in the HEAVY-GAS-HAR-tf model and $\beta + \beta_1 + \beta_2$ measures the persistence effect of previous predictions on current predictions. Since α is relatively large, it follows that explicit information on the score is essential for explicit models. In our implicit HEAVY-ProPar-tF model, the implicit information on solely the score is less used, following a lower value for α. One explanation for this would be that, in this formulation, β not only captures the effects of the previous predictions but also of implicit information, as it accounts for the effects of the updated (instead of the predicted) covariance matrix. This updated covariance matrix, therefore, carries relatively more information, making it more valuable, and consequently the implicit score on its own less valuable. In terms of fit, our implicit model provides a marginally better model for in-sample analysis, following a slightly higher optimal total log-likelihood and mean score. The results of this latter forecasting evaluation metric also indicate that our proposed HEAVY-ProPar-tF model provides a significantly better model than the explicit HEAVY-GAS-HAR-tF model. This does not only stem from the fact that we use implicit (instead of explicit) information (a conclusion often drawn in this paper). Namely, our model also penalizes for deviations with the previously predicted covariance matrix, weighted by the penalty matrix $\boldsymbol{P}$. In sum, we have developed a novel multivariate score-driven model, proven to yield a relatively more accurate covariance matrix estimator and predictor.

Table 13: Parameter estimates, total likelihoods $\mathcal{L}^*$ and mean scores for the HEAVY-GAS-HAR-tF and HEAVY-ProPar-tF model.

	HEAVY-GAS-HAR-tF	HEAVY-ProPar-tF
ν	2.363(0.004)	2.358(0.003)
ν_1	103.249(53.733)	104.244(2.227)
ν_2	40.764(1.003)	39.964(0.020)
α	0.478(0.002)	0.369(0.001)
β	0.443(0.004)	0.369(0.001)
β_1	0.175(0.022)	\
β_2	0.001(0.151)	\
$\mathcal{L}^*$	6671.289	6714.983
Score	114.195	114.848*

Standard errors are in parentheses. A mean score with an asterisk indicates that the corresponding model significantly outperforms the other using a 5% significance level. Here, we use a DM test to statistically test for potential superior predictive ability using the difference in scores as a loss function.

6 Conclusion

In this research, we provided a broad analysis of the proximal-parameter (ProPar) filter and presented alternative models based on different modeling assumptions. The ProPar filter, constructed by Lange et al. (2022), joined the emerging class of dynamical conditional score (Harvey, 2013, DCS) or generalized autoregressive score (Creal et al., 2013, GAS) models. This field can be characterized by its observation-driven modeling of parameters and its focus on the first-order condition associated with the logarithmic observation density (referred to as the score), to obtain time-varying parameter estimates. By constructing an implicit stochastic-gradient update and penalizing for parameter deviations, the ProPar framework presented this scientific strand with a novel model. Our study leverages these two fundamental aspects to develop univariate and multivariate score-driven models for financial time-series analysis, specifically applicable to high-frequency datasets comprising intraday stock returns.

Throughout our simulations and empirical results, we found more favorable results when using implicit instead of explicit stochastic-gradient updates. Moreover, we have investigated the adoption of a conditional Student's t distribution, rather than the conventional assumption of a conditional normal distribution for returns. Here, we found significant results in favor of the former distribution. Furthermore, our univariate and multivariate score-driven models (ProPar-HF and HEAVY-ProPar-tF respectively), based on high-frequency stock data, were found to possess more accurate estimates and predictions for time-varying volatilities and covariance matrices compared to their explicit counterparts.

Obtaining time-varying volatility estimates is crucial for e.g. financial econometrics, asset management, and option pricing, as it allows for better risk assessment and derivative pricing. As such, our analyses and novel modeling frameworks provide valuable perspectives and adequate score-driven time-series models, not only for this scientific strand but also for market agents seeking the most precise time-varying volatility estimators.

Moving forward, there are several potential paths for future research. For example, during our research, we made use of two relatively established non-linear optimization techniques. However, numerous potentially more efficient algorithms have been proposed throughout the literature, which warrants a comparative study among them. Furthermore, at the cost of more computer memory and enlarged computational times, a larger high-frequency data sample (by expanding dimensions or extending time series) could yield more valuable insights and potentially alternative outcomes, thereby serving as a promising avenue for further investigation.

References

Ait-Sahalia, Y., Mykland, P. A., & Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial Studies*, *18*(2), 351–416.

Aït-Sahalia, Y., Mykland, P. A., & Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, *160*(1), 160–175.

Alberg, D., Shalit, H., & Yosef, R. (2008). Estimating stock market volatility using asymmetric GARCH models. *Applied Financial Economics*, *18*(15), 1201–1208.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*(4), 885–905.

Andersen, T. G., Bollerslev, T., & Das, A. (2001). Variance-ratio statistics and high-frequency data: Testing for changes in intraday volatility patterns. *The Journal of Finance*, *56*(1), 305–327.

Andersen, T., Bollerslev, T., & Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance*, *6*(5), 457–477.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, *12*(3), C1–C32.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche Mark options. *The Review of Financial Studies*, *9*(1), 69–107.

Blasques, F., van Brummelen, J., Koopman, S. J., & Lucas, A. (2022). Maximum likelihood estimation for score-driven models. *Journal of Econometrics*, *227*(2), 325–346.

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

Boudt, K., Kleen, O., & Sjørup, E. (2022). Analyzing intraday financial data in R: The highfrequency package. *Journal of Statistical Software*, *104*(8), 1–36.

Brownlees, C. T., & Gallo, G. M. (2006). Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis*, *51*(4), 2232–2245.

Cailliez, F. (1983). The analytical solution of the additive constant problem. *Psychometrika*, *48*(2), 305–308.

Cox, D., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., & Juselius, K., & Lauritzen, S. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, *8*(2), 93–115.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134-144.

Duffie, D., Pan, J., & Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, *68*(6), 1343–1376.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, *50*(4), 987–1007.

Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, *5*(1), 1–50.

Gorgi, P., Hansen, P. R., Janus, P., & Koopman, S. J. (2019). Realized Wishart-GARCH: A score-driven multi-asset volatility model. *Journal of Financial Econometrics*, *17*(1), 1–32.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, *35*(3), 705–730.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: With applications to financial and economic time series* (Vol. 52). Cambridge University Press.

Harvey, A., & Lange, R.-J. (2017). Volatility modeling with a generalized t distribution. *Journal of Time Series Analysis*, *38*(2), 175–190.

Harvey, A., Ruiz, E., & Shephard, N. (1994). Multivariate Stochastic Variance Models. *The Review of Economic Studies*, *61*(2), 247–264.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, *6*(2), 327–343.

Lange, R.-J., van Os, B., & van Dijk, D. J. (2022). Robust observation-driven models using proximal-parameter updates. *Available at SSRN 4227958*.

Lord, R., Koekkoek, R., & van Dijk, D. J. (2010). A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, *10*(2), 177–194.

Magnus, J. R. (1985). On differentiating eigenvalues and eigenvectors. *Econometric Theory*, *1*(2), 179–191.

McDonald, J. B., & Newey, W. K. (1988). Partially adaptive estimation of regression models via the generalized t distribution. *Econometric Theory*, *4*(3), 428–457.

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

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics* (Vol. 4, pp. 2111–2245). Elsevier.

Nocedal, J., & Wright, S. (2006). *Numerical optimization*. Springer.

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*(4), 643–657.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Patton, A. J., & Sheppard, K. (2009). Evaluating volatility and correlation forecasts. In *Handbook of Financial Time Series* (pp. 801–838). Springer.

Shephard, N. (2005). *Stochastic volatility: Selected readings*. Oxford University Press.

Taylor, S. J., & Xu, X. (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance*, *4*(4), 317–340.

Toulis, P., & Airoldi, E. M. (2017). Asymptotic and finite-sample properties of estimators based on stochastic gradients. *The Annals of Statistics*, *45*(4), 1694 – 1727.

Tsay, R. S. (2005). *Analysis of Financial Time Series*. John Wiley & Sons.

Van Haastrecht, A., & Pelsser, A. (2010). Efficient, almost exact simulation of the Heston stochastic volatility model. *International Journal of Theoretical and Applied Finance*, *13*(1), 1–43.

Wood, R. A., McInish, T. H., & Ord, J. K. (1985). An investigation of transactions data for NYSE stocks. *The Journal of Finance*, *40*(3), 723–739.

7 Appendix

7.1 Data

Table 14: 15 Dow Jones Industrial Average stocks considered in this research

Stock Ticker	**Stock Name**
AA	Alcoa Corporation
AXP	American Express Company
BA	The Boeing Company
CAT	Caterpillar
GE	General Electric Company
HD	The Home Depot
HON	Honeywell
IBM	International Business Machines Corporation
JPM	JPMorgan Chase & Co.
KO	The Coca-Cola Company
MCD	McDonald's Corporation
PFE	Pfizer Inc.
PG	The Procter & Gamble Company
WMT	Walmart Inc.
XOM	Exxon Mobil Corporation

7.2 Mathematical Simplifications

7.2.1 Volatility Formula ProPar

Equation 4 is derived as follows. Given the optimization setup, we have the following first-order condition.

$$\begin{aligned} \nabla(y_t \mid h_{t|t}) - P(h_{t|t} - h_{t|t-1}) &= 0 \\ \implies h_{t|t} = h_{t|t-1} + P^{-1}\nabla(y_t \mid h_{t|t}) &= h_{t|t-1} + \eta\nabla(y_t \mid h_{t|t}). \end{aligned}$$

Using the normal distributed model specification for y_t, we can calculate $\nabla(y_t \mid h_{t|t})$ as

$$
\begin{aligned}
\nabla(y_t \mid h_{t|t}) &= \frac{d}{dh_t} \log p\left(y_t \mid \exp(h_t)\right) \Big|_{h_t = h_{t|t}} \\
&= \frac{d}{dh_t} \log \frac{1}{\exp(h_t)\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y_t-\mu}{\exp(h_t)}\right)^2} \Big|_{h_t = h_{t|t}} \\
&= \frac{d}{dh_t} \left(-h_t - \log\sqrt{2\pi} - \frac{1}{2}\frac{(y_t-\mu)^2}{\exp(2h_t)} \right) \Big|_{h_t = h_{t|t}} \\
&= -1 + \left(\frac{y_t-\mu}{\exp(h_{t|t})}\right)^2 .
\end{aligned}
$$

Hence, $h_{t|t} = h_{t|t-1} + \eta\left(-1 + (\frac{y_t-\mu}{\exp(h_{t|t})})^2\right)$.

7.2.2 Lambert W solution

We start with the volatility formula under the ProPar framework (Equation 4):

$$
\begin{aligned}
h_{t|t} &= h_{t|t-1} + \eta \left[\left(\frac{y_t-\mu}{\exp\left(h_{t|t}\right)} \right)^2 - 1 \right] \\
&= h_{t|t-1} + \eta \left[\frac{(y_t-\mu)^2}{\exp\left(2h_{t|t}\right)} - 1 \right],
\end{aligned}
\tag{43}
$$

multiplying everything with 2, this is equivalent to writing

$$
2h_{t|t} = 2\eta \left(\frac{(y_t-\mu)^2}{\exp\left(2h_{t|t}\right)} \right) + 2(h_{t|t-1} - \eta). \tag{44}
$$

This leads to the final rewritings:

$$
\begin{aligned}
& 2h_{t|t} \exp\left(2h_{t|t}\right) && = 2\eta(y_t-\mu)^2 + 2(h_{t|t-1}-\eta)\exp\left(2h_{t|t}\right) \\
& (2h_{t|t} - 2(h_{t|t-1}-\eta))\exp\left(2h_{t|t}\right) && = 2\eta(y_t-\mu)^2 \\
& (2h_{t|t} - 2(h_{t|t-1}-\eta))\exp\left(2h_{t|t}\right)\exp\left(-2(h_{t|t-1}-\eta)\right) && = 2\eta(y_t-\mu)^2\exp\left(-2(h_{t|t-1}-\eta)\right) \\
& (2h_{t|t} - 2(h_{t|t-1}-\eta))\exp\left(2h_{t|t} - 2(h_{t|t-1}-\eta)\right) && = 2\eta(y_t-\mu)^2\exp\left(-2(h_{t|t-1}-\eta)\right) \\
\implies & (2h_{t|t} - 2(h_{t|t-1}-\eta)) && = W_0\left(2\eta(y_t-\mu)^2\exp\left(-2(h_{t|t-1}-\eta)\right)\right),
\end{aligned}
$$

where we used in the last step that $W_0(y)$ is the solution to $x\exp(x) = y$. Finally, rewriting this solution for $h_{t|t}$ yields $h_{t|t} = h_{t|t-1} - \eta + \frac{1}{2}W_0\left(2\eta(y_t-\mu)^2\exp\left(-2(h_{t|t-1}-\eta)\right)\right)$.

7.2.3 Implicit Hessian under a Student's t distribution

Under our modeling assumptions, the implicit Hessian (i.e. the derivative of the score with respect to the updated volatility) under the IBT-EGARCH model can be formulated as:

$$H(y_t \mid h_{t|t}) = \frac{d}{dh_{t|t}} \nabla(y_t \mid h_{t|t}) = \frac{d}{dh_{t|t}} (\nu+1) \frac{\varepsilon_t^2}{\varepsilon_t^2 + \nu} - 1,$$

where $\varepsilon_t^2 = (y_t - \mu)^2 \exp(-2h_{t|t})$. The derivative of ε_t^2 with respect to $h_{t|t}$ is equal to $-2(y_t - \mu)^2 \exp(-2h_{t|t})$. Hence, we can evaluate the Hessian as:

$$\begin{aligned} H(y_t \mid h_{t|t-1}) &= (\nu+1) \frac{(\varepsilon_t^2 + \nu)(-2(y_t-\mu)^2 \exp(-2h_{t|t})) - \varepsilon_t^2(-2(y_t-\mu)^2 \exp(-2h_{t|t}))}{(\varepsilon_t^2 + \nu)^2} \\ &= (\nu+1) \frac{-2\nu(y_t-\mu)^2 \exp(-2h_{t|t})}{(\varepsilon_t^2 + \nu)^2}. \end{aligned}$$

7.2.4 Long-Term Volatility predictions

From Equation 6, we have $\mathrm{E}_t[h_{t+1}] = \hat{\omega} + \hat{\phi} h_{t|t}$. Under the standard ProPar filter, we can write the d-step ahead volatility prediction by recursion:

$$\begin{aligned} h_{t+d|t} &= \omega + \phi h_{t+d-1|t} \\ &= \omega + \phi(\omega + \phi h_{t+d-2|t}) \\ &= \omega(1+\phi) + \phi^2 h_{t+d-2|t} \\ &= \dots \\ &= \omega(1 + \phi + \ldots + \phi^{d-1}) + \phi^d h_{t|t} \\ &= \omega \frac{1-\phi^d}{1-\phi} + \phi^d h_{t|t}. \end{aligned} \tag{45}$$

Next, we compute the expectation of squared returns. The above result implies

$$\begin{aligned} \mathrm{E}_t[y_{t+d}^2] &= \mathrm{E}_t[(\mu + \exp(h_{t+d}) z_t)^2] \\ &= \mu^2 + \mathrm{E}_t[\exp(h_{t+d})^2 z_t^2] \\ &= \mu^2 + \mathrm{E}_t[\exp(h_{t+d})^2] \\ &= \mu^2 + \mathrm{E}_t[\exp(2\omega \frac{1-\phi^d}{1-\phi} + 2\phi^d h_{t|t})] \\ &= \mu^2 + \exp(2\omega \frac{1-\phi^d}{1-\phi} + 2\phi^d h_{t|t}), \end{aligned}$$

where we used that, following the geometric series formula, $(1 + \phi + \ldots + \phi^{d-1}) = \sum_{i=1}^{d} \phi^{i-1} = \frac{1-\phi^d}{1-\phi}$, whenever $\phi < |1|$. We can impose this restriction in our maximum likelihood optimization

to stimulate optimization speed. This constraint can be considered reasonable as it would indicate that the importance of the updated volatility estimate with respect to the 1-step ahead predicted volatility estimate lies between 0 and ± 100%. The rest of the importance is captured by ω. Besides, the constraint is also reasonable to prevent explosive behavior in $(1+\phi+\ldots+\phi^{d-1})$ and ϕ^d.

To obtain the expected variance over a period of d days, we compute

$$\begin{aligned} \mathrm{E}_t[\sum_{i=1}^{d} y_{t+i}^2] &= \sum_{i=1}^{d} \mathrm{E}_t[y_{t+i}^2] \\ &= \sum_{i=1}^{d} \mu^2 + \exp(2\omega\frac{1-\phi^i}{1-\phi} + 2\phi^i h_{t|t}) \\ &= d\mu^2 + \sum_{i=1}^{d} \exp(2\omega\frac{1-\phi^i}{1-\phi} + 2\phi^i h_{t|t}), \end{aligned} \tag{46}$$

hence, the expected volatility over a period of d days is the square root of this formula.

For the alternative models (GAS and IBT-EGARCH), we apply a similar approach. The GAS model was modeled similarly to the ProPar model, except that the update $h_{t|t}$ uses argument $h_{t|t-1}$. Nevertheless, $h_{t|t}$ is still an argument for the prediction (Equation 6). Therefore, the d-day volatility prediction under the GAS model is similar to the ProPar model. Trivially, Equation 46 is also applicable to the GAS model, where $h_{t|t}$ is computed differently in the sense that we can immediately calculate it, instead of analytically having to solve for it. The same reasoning can be applied to the IBT-EGARCH model. Note, since the variance of the (Student's t-distributed) shock z_t now is equal to $\frac{\nu}{\nu-2}$, each exponent function should be multiplied by this factor. Here, $\nu > 4$ equals the maximum likelihood estimator for the degrees of freedom.

7.3 Other Simulation Results

Table 15: Average parameter estimates and standard errors in parentheses, MSEs of the filtered volatility updates, and their corresponding DM test statistic.

Scenario	1	2	3	4	5	6	7
	Maximum Likelihood Optimization						
$\hat{\eta}$	0.607(0.044)	0.549(0.032)	0.280(0.019)	0.210(0.016)	0.523(0.032)	0.272(0.019)	0.070(0.007)
	0.058(0.003)	0.057(0.002)	0.059(0.003)	0.059(0.003)	0.059(0.002)	0.059(0.003)	0.024(0.003)
$\hat{\omega}$	0.188(0.017)	0.164(0.013)	0.093(0.009)	0.069(0.007)	0.161(0.013)	0.091(0.008)	0.003(0.003)
	-0.034(0.003)	-0.034(0.002)	-0.003(0.002)	0.002(0.002)	-0.026(0.002)	-0.003(0.002)	-0.208(0.017)
$\hat{\phi}$	0.933(0.004)	0.933(0.004)	0.961(0.003)	0.967(0.003)	0.932(0.004)	0.963(0.003)	0.979(0.003)
	0.951(0.002)	0.950(0.002)	0.980(0.002)	0.980(0.002)	0.965(0.002)	0.978(0.002)	0.270(0.047)
$\hat{\mu}$	-0.017(0.009)	-0.023(0.005)	-0.023(0.008)	-0.019(0.011)	-0.023(0.005)	-0.003(0.008)	-0.005(0.009)
	-0.017(0.008)	-0.023(0.007)	-0.024(0.010)	-0.019(0.013)	-0.023(0.006)	-0.003(0.010)	-0.005(0.010)
	Performance Metrics						
MSE_ProPar	0.824	1.102	1.157	1.129	0.878	1.262	0.199
MSE_GAS	0.876	38.852	210.364	1.389	106.446	2.261	0.245
DM	-1.010	-1.762	-1.028	**-9.746**	-1.381	-3.822	**-11.842**

$T = 4036$. For the parameter estimates, the upper and lower element corresponds to those of the ProPar and GAS filter, respectively. The DM test statistics in bold indicate a p-value lower than the significance level of 0.05.

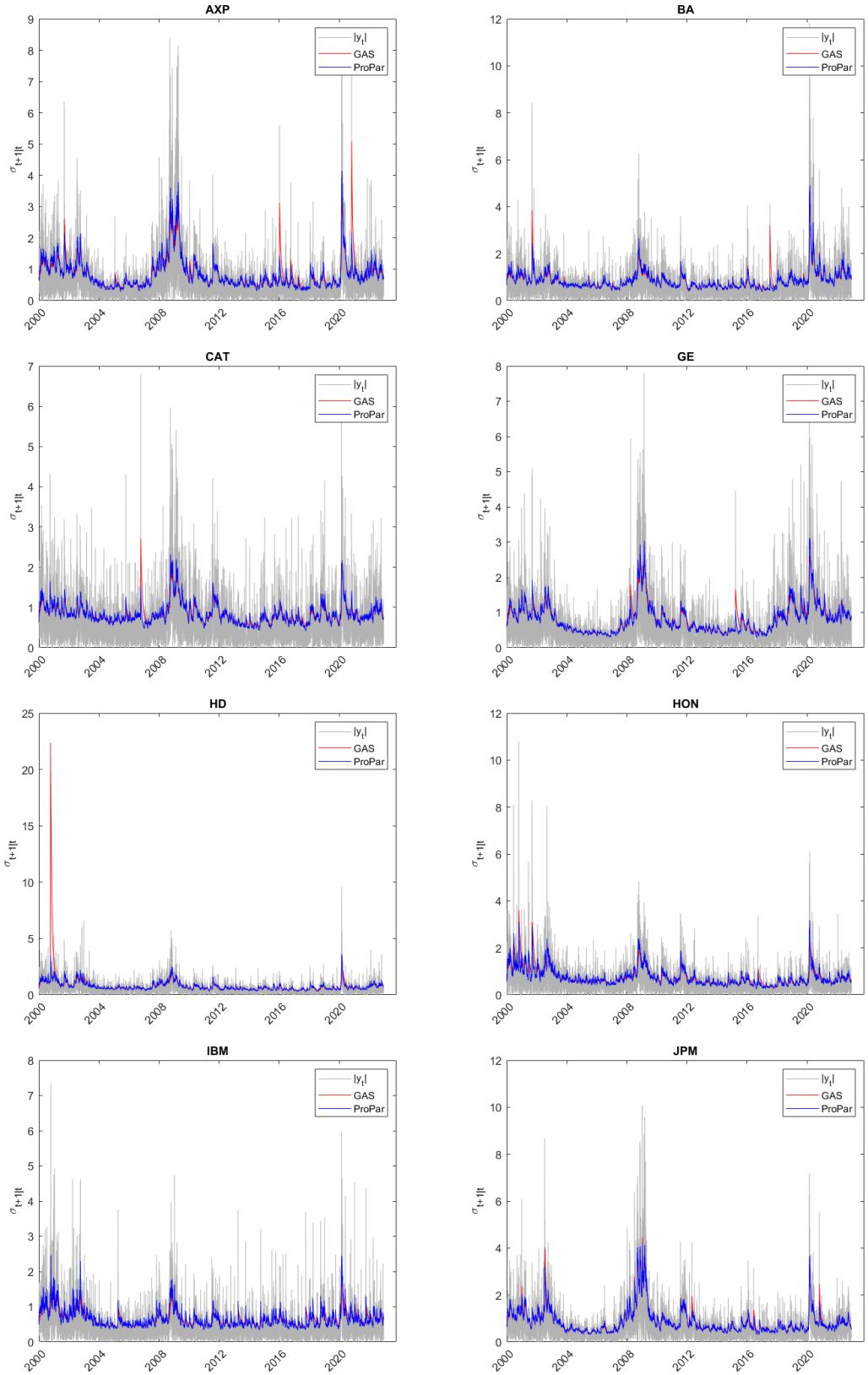
Table 16: Average parameter estimates and standard errors in parentheses, MSEs of the filtered volatility updates, and their corresponding DM test statistic.

Scenario	1	2	3	4	5	6	7
	Maximum Likelihood Optimization						
$\hat{\eta}$	0.668(0.104)	0.639(0.106)	0.317(0.178)	0.239(0.052)	0.601(0.092)	0.329(0.067)	0.088(0.029)
	0.057(0.007)	0.057(0.006)	0.056(0.009)	0.056(0.009)	0.057(0.007)	0.055(0.008)	0.038(0.010)
$\hat{\omega}$	0.163(0.041)	0.135(0.036)	0.106(0.028)	0.076(0.022)	0.121(0.035)	0.086(0.025)	-0.007(0.010)
	-0.094(0.012)	-0.120(0.012)	0.010(0.012)	-0.003(0.006)	-0.093(0.016)	-0.011(0.007)	-0.072(0.013)
$\hat{\phi}$	0.923(0.016)	0.903(0.016)	0.935(0.014)	0.947(0.014)	0.899(0.016)	0.940(0.014)	0.954(0.016)
	0.901(0.008)	0.903(0.008)	0.941(0.010)	0.951(0.009)	0.889(0.014)	0.934(0.010)	0.756(0.020)
$\hat{\mu}$	-0.002(0.022)	-0.003(0.012)	-0.005(0.042)	-0.025(0.043)	-0.004(0.019)	0.008(0.034)	-0.003(0.026)
	-0.003(0.022)	-0.004(0.015)	-0.004(0.043)	-0.025(0.049)	-0.004(0.020)	0.008(0.039)	-0.002(0.028)
	Performance Metrics						
MSE_ProPar	0.572	0.300	1.147	1.129	0.308	1.212	0.188
MSE_GAS	29.949	33.581	2.350	1.389	0.497	1.793	0.209
DM	-1.183	-1.070	**−1.969**	**−11.116**	-1.061	**−2.505**	**5.581**

$T = 504$. For the parameter estimates, the upper and lower element corresponds to those of the ProPar and GAS filter, respectively. The DM test statistics in bold indicate a p-value lower than the significance level of 0.05.

7.4 Empirical Results

Figure 5: Estimated paths for $\sigma_{t+1|t}$ of the ProPar and GAS filter for daily returns, from January 2000 until December 2022



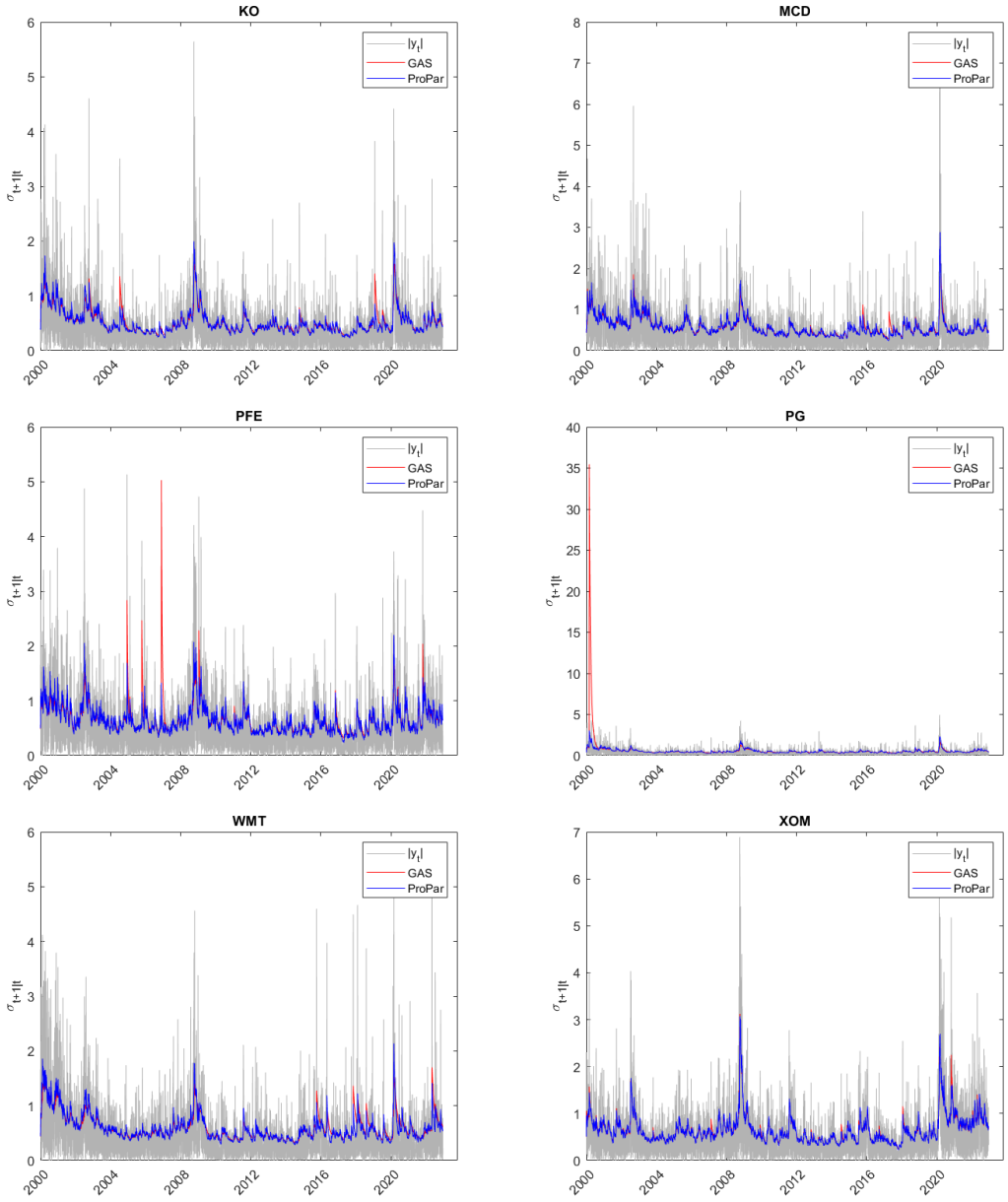
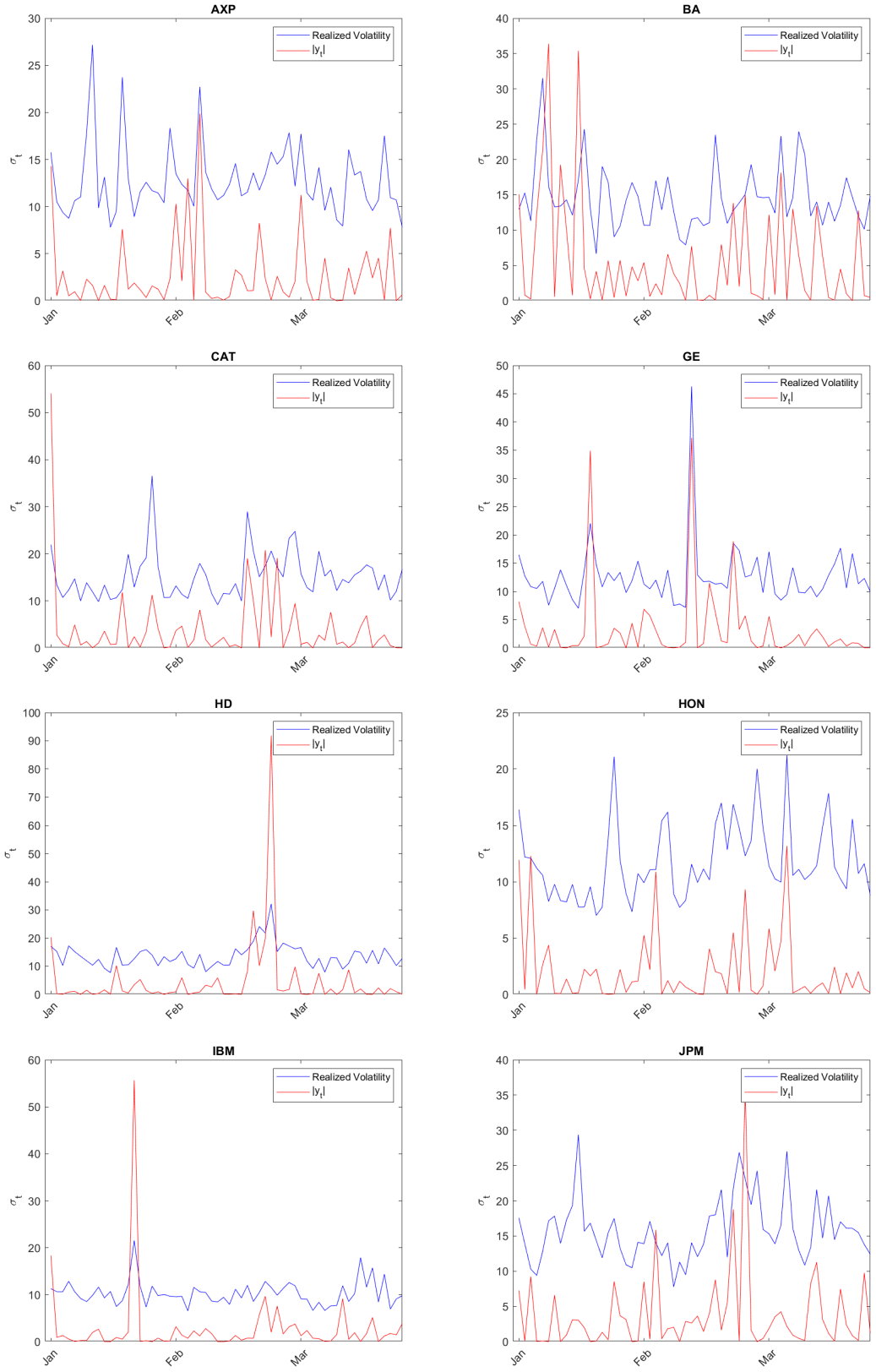
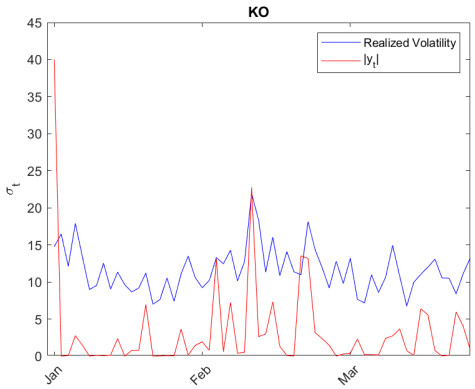
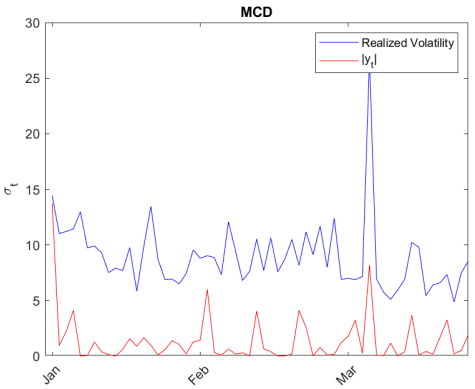
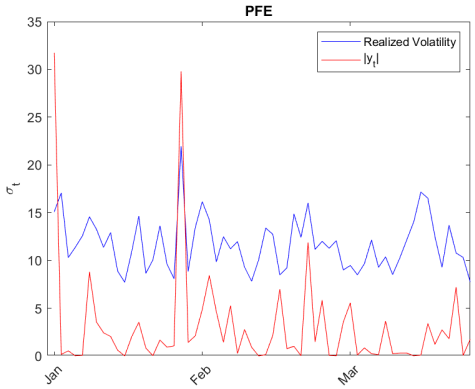

KO
MCD
PFE
PG
WMT
XOM
$|y_t|$
GAS
ProPar
$\sigma_{t+1|t}$
2000
2004
2008
2012
2016
2020

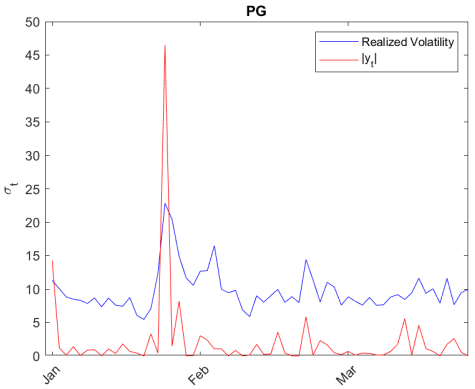
Figure 7: Two daily volatility paths using the realized volatility and absolute daily return.

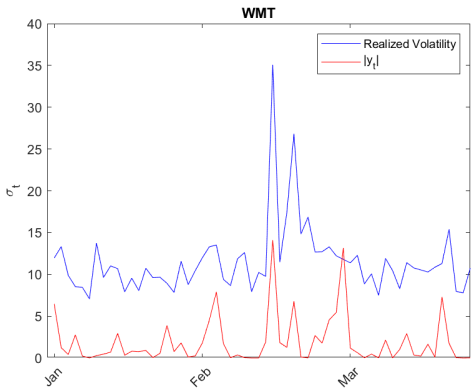


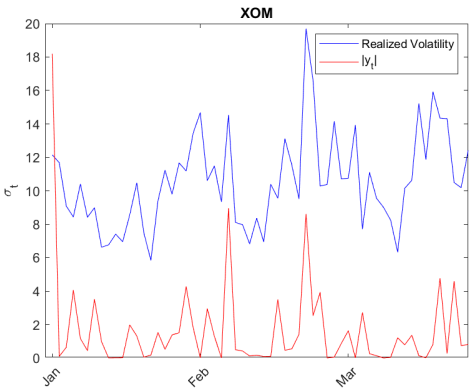

KO
Realized Volatility
|y_t|
σ_t
Jan
Feb
Mar


MCD
Realized Volatility
|y_t|
σ_t
Jan
Feb
Mar


PFE
Realized Volatility
|y_t|
σ_t
Jan
Feb
Mar


PG
Realized Volatility
|y_t|
σ_t
Jan
Feb
Mar


WMT
Realized Volatility
|y_t|
σ_t
Jan
Feb
Mar


XOM
Realized Volatility
|y_t|
σ_t
Jan
Feb
Mar