CERGE - EI

Center for Economic Research and Graduate Education –

Economics Institute

Modeling Conditional Distribution of Intraday Returns

Michal Hakala

Dissertation

Prague 2026

Dissertation Committee:

Stanislav Anatolyev, Ph.D. (CERGE-EI, Chair)

Jozef Baruník, Ph.D. (FSV IES, Charles University)

Veronika Selezneva, Ph.D. (Paris Dauphine-PSL University)

Referees:

Ilze Kalnina, Ph.D. (North Carolina State University)

Robert Luger, Ph.D. (Université Laval)

Contents

Acknowledgement

I would like to express my gratitude to my supervisor, Stanislav Anatolyev, for his guidance throughout my studies. His influence sparked my interest in a deeper understanding of statistics and econometrics. His thorough feedback, inexhaustible patience, and critical insights made this work possible.

I also express my gratitude to Denis Chetverikov for hosting my research stay at UCLA, providing guidance, inspirational knowledge, critical insights, and a positive attitude even in moments of setback. His role was essential for the progress of my work.

I would like to extend my thanks to the CERGE-EI as a community, namely to Veronika Selezneva, for her guidance and support in the early years of my studies. Special thanks belong to the Academic Skill Center – Andrea Downing, Deborah Novakova, and Gray Krueger – for their tireless editing work.

All remaining errors are my own.

Abstract

Time-series modeling of conditional distributions of intraday returns is of great importance to financial professionals and academic researchers. This work contributes to a methodological and empirical body of knowledge on conditional distributions of intraday asset returns. In Chapter 1, we study distributional diurnal patterns. We propose a new semi-parametric modeling framework for capturing distributional diurnal patterns inspired by traditional seasonal adjustment methods compatible with common models in the literature. Capturing distributional diurnal patterns substantially improves forecast precision. In Chapter 2, we study clustered commonality in the stock market to improve intraday volatility forecasts. Using sectors, industries, and data-driven clusters, we extend the heterogeneous autoregressive (HAR) model with relevant groups or clusters of stocks chosen by regularization methods. In Chapter 3, we propose a model with a novel distribution based on Tukey's H transformation to study conditional kurtosis of intraday returns and spillovers of extreme shocks in the stock market. Empirical applications are based on intraday large-cap US-listed stock returns.

Keywords

Conditional distribution, asset returns, intraday returns, time series, diurnal patterns, commonality, realized volatility, kurtosis, extreme risk, spillovers, value-at-risk, expected shortfall

Introduction

Intraday asset returns are high-frequency returns realized within one day. One example is a stock return realized over a period of 5 minutes. The growing importance of high-frequency trading and advancements in computational capabilities have sparked interest in intraday returns among economic and financial researchers. The motivation for research on intraday asset returns is broadly diverse. While practitioners like day traders, market makers, and risk managers are interested in forecasting to gain a competitive advantage, academic financial researchers and policy makers are interested in intraday returns because intraday frequencies may allow the study of phenomena that cannot be studied using daily returns. Intraday frequencies present unique methodological challenges. This work[1] addresses challenges related to modeling conditional distributions of intraday asset returns.

In the first chapter, we study diurnal patterns of return distributions. Diurnal patterns are specific features of intraday returns and represent systematically repeating patterns throughout a day. Diurnal patterns are similar to the seasonality of low-frequency time series. Econometricians have studied diurnal patterns in volatility, trading volume, durations between trades, and cross-sectional correlations extensively, but diurnal patterns in distributions of asset returns remain poorly understood. Because capturing distributional diurnal patterns is essential for precise forecasting of conditional distributions used in risk management and related fields, we aim to fill the gap in the literature. We propose a new semi-parametric modeling framework for capturing distributional diurnal patterns inspired by traditional seasonal adjustment methods compatible with common time series models in the literature, such as E-GARCH. Using a dataset of fifty highly liquid US-listed stocks, we demonstrate superior distributional forecasting performance of models capturing distributional diurnal patterns as opposed to models capturing only the diurnal patterns in volatility.

In the second chapter, we narrow our focus from the entire distribution to the conditional volatility of asset returns. We utilize clustered commonality in the stock market to improve the intraday volatility forecasts for individual stocks. The

[1] All three chapters are the sole work of the author.

commonality is the similarity of a group of assets in terms of price behavior. We examine sectors, industries, and data-driven clusters as potential groups to enhance the heterogeneous autoregressive (HAR) model and apply regularization methods to identify relevant groups for volatility forecasting of individual stocks. Analyzing 363 large-cap US-listed stocks, we demonstrate superior volatility forecasting performance. The forecast improvement is more substantial during trading hours and within sectors of the economy with greater commonality. Our findings are consistent with the liquidity spiral mechanism as an explanation for the existence of a commonality.

The last chapter focuses on the dynamics of intraday returns kurtosis. Asset return kurtosis is a conditional distribution property typically associated with an extreme tail risk. We propose a new model for time-varying kurtosis with a novel distribution based on Tukey's H transformation. Conditional time-varying kurtosis of daily returns is typically higher than three, i.e., the distribution is leptokurtic. However, at intraday frequencies, the stock returns are close to potentially platykurtic shapes, which may complicate the estimation of models established in the literature. Our novel distribution can have leptokurtic and platykurtic shapes. We apply our proposed model to 5-minute frequency returns of 100 highly liquid US-listed stocks to understand the nature of time-varying kurtosis of intraday returns and study spillovers of extreme risk. We show that the main driver behind periods of increased kurtosis of individual stocks is spillovers from other stocks in the market rather than delayed price reaction to new shocking information that is unique to an individual stock.

1 Semi-parametric Framework for Modeling Distributional Diurnal Patterns of Intraday Returns

1.1 Introduction

Intraday asset returns are returns realized within one day, e.g., over a period of 30 minutes. Unlike daily returns, intraday asset returns exhibit strong diurnal patterns similar to the seasonality commonly present in lower-frequency time series. The most studied diurnal pattern of intraday returns is the well-known U-shaped pattern of intraday volatility. The volatility is typically highest immediately after the market opens, lowest around lunchtime, and increases again before the market closes. Research documenting diurnal patterns in other quantities also exists. For example, Stephan and Whaley (1990) document diurnal patterns in trading volumes and price elasticities. Engle and Russell (1998) in their seminal work research diurnal patterns in durations between trades, which sparked a rich stream of subsequent research, e.g., Ghysels et al. (2004) or Anatolyev and Shakin (2007). Another example is the work of Heston et al. (2010), who find diurnal patterns in cross-sectional correlation between stocks. In this study, we examine diurnal patterns in distributions of intraday asset returns beyond volatility. We propose a new semi-parametric modeling framework for capturing distributional diurnal patterns.

Andersen and Bollerslev (1997b) studied diurnal patterns in volatility at various different frequencies and argued that the diurnal volatility pattern can be captured as a simple systematic pattern in marginal volatility throughout the day. "Marginal," in this context, means an unconditional long-term property of the distribution repeating each day at a specific time that can be removed by a time-invariant filter. Based on the idea of a systematic pattern in marginal volatility,

they proposed a simple linear time-invariant filter similar to seasonal adjustment[1]. This simple idea is used frequently, e.g., in the multiplicative decomposition of Feng and McNeil (2008) or the multiplicative GARCH of Engle and Sokalska (2012). The approach is popular because it is simple and computationally inexpensive. Using a more complex approach may provide better forecasts, but the improvement is usually only marginally better (Martens et al., 2002). The class of models based on adjustment for diurnal patterns is estimated in two steps. First, the repeating diurnal patterns are estimated, and the series are adjusted so that the patterns are removed from the marginal distribution. Then, models with no diurnal patterns that are common in daily literature are applied to the adjusted series.

The repeating diurnal patterns may not be constant over a long period of time. Recently, Andersen et al. (2024) examine the hypothesis that marginal diurnal patterns in volatility may be slowly changing over time. Their work is based on asymptotic high-frequency realized measures; thus, it is not directly applicable to the models of conditional distributions. However, it may suggest a potential future direction for research, switching from constant adjustment to slowly drifting adjustment over time.

In contrast to the approach based on the diurnal adjustment of marginal volatility, Bollerslev and Ghysels (1996) proposed the periodic GARCH model, in which diurnal patterns are captured by periodically changing parameters of the GARCH process. The models with periodic parameters are more flexible. For example, periodic GARCH allows conditional variance to be less persistent in the morning and more persistent in the afternoon. The model with periodic marginal volatility is nested within the periodic GARCH since it is equivalent to a process in which only intercepts are systematically changing throughout the day, just like seasonal dummies, and other parameters are fixed. The main limitation of the approach is the very heavy parameterization, as each time of day has its own set of parameters.

In our work, we study diurnal patterns beyond volatility and aim to improve the forecast of conditional distributions of intraday returns. This may include

[1]For examples of time invariant filters for adjusting seasonality, see work Wallis (1974) or Bell and Hillmer (1984).

higher moments, such as skewness or kurtosis, or an entire distribution. Since standardized moments, such as skewness or kurtosis, are invariant to affine transformation, the application of the adjustment-based models is rather rare in the literature on distributional modeling, and periodic-like models are more popular. For example, Coroneo and Veredas (2012) propose periodic quantile regression, in which the conditional distribution is captured by multiple quantile regressions with parameters changing throughout the day. They document diurnal patterns in higher moments – skewness and kurtosis. Similarly, Eckernkemper and Gribisch (2021) propose a periodic extension to the Generalized Autoregressive Score (GAS) model. Long-memory GARCH models with diurnal patterns can also somewhat capture distributional diurnal patterns to a certain degree; e.g., periodic long-memory GARCH can generate diurnal patterns in kurtosis (Bordignon et al., 2007; Rossi & Fantazzini, 2015). However, the flexibility of potential diurnal patterns is limited.

The periodic models are commonly simplified, since full models are heavily parameterized – a separate set of parameters for each intraday period. Therefore, Fourier series or power series are usually used to approximate the relation between the time of the day and the value of the parameter. Less frequently, a non-parametric approach is used. For example, Ergün and Jun (2010), apply the Extreme Value Theory and Autoregressive Conditional Density (ARCD) model of B. E. Hansen (1994) to adjusted series, where the adjustment is achieved by dividing log-returns by diurnal volatility factors estimated by non-parametric smoothing.

Our work contributes to the literature on models based on adjustment. All of the adjustment-related sources cited can be seen as applications of additive and multiplicative decomposition of time series, or a combination of both, for what would be considered a seasonal adjustment in the daily time series literature. In other words, the mean is adjusted by addition, and the volatility by multiplication. This limits the modeling of diurnal patterns only to the first two moments, since higher standardized moments, such as kurtosis and skewness, are invariant to affine transformation. In order to adjust higher moments, a non-linear adjustment has to be used. Our main research question is, can we improve the forecast of the

conditional distribution by adjusting the entire distribution? We clearly need a non-linear adjustment for the goal.

To capture the diurnal patterns in the entire distribution, we propose using the probability integration transformation (PIT) by a marginal distribution at a specific time of the day. For more on the theoretical properties of PIT, see Angus (1994). The PIT can be seen as a non-linear time-invariant filter for any potential distributional diurnal patterns. The adjustment results in a pattern-free series distributed in the $(0, 1)$ interval. To fit the pattern-free series, we develop a new class of models, where processes of conventional time series models, such as GARCH, are transformed by the PIT of their marginal distributions. The resulting models resemble conventional time series models in nature and are consistent with known properties of the distributionally adjusted series.

It is important to note that we are proposing a modeling framework for capturing diurnal patterns that is compatible with typical models used in the intraday literature, e.g. Amado and Teräsvirta (2013), or Amado and Teräsvirta (2017). Our proposed framework is applicable to GARCH-like or ARCD-like models as long as the return process has a stationary marginal distribution, and it is practically feasible to solve for the cumulative distribution function numerically. Our framework excludes, for example, the model of Engle and Sokalska (2012), who use commercially available volatility forecasts. The distribution of the commercially available forecasts is not specified; thus, the marginal distribution is unknown. However, a similar model, RealGARCH (P. R. Hansen et al., 2012) that fully specifies the data-generating process, and hence the marginal distribution, is compatible with the proposed framework. The framework is even compatible with the ARCD models for time-varying skewness and kurtosis pioneered by B. E. Hansen (1994) and further developed over the following three decades, e.g., Jondeau and Rockinger (2003), Brooks et al. (2005), or Anatolyev and Petukhov (2016). For a review, see Soltyk and Chan (2023). Since time-varying higher moments are typically not modeled in the intraday literature, studying the application of our framework with ARCD-like models is outside the scope of this work and is left for future investigation. The idea of distributional adjustment is also compatible with flexible semiparametric models, e.g. Anatolyev and Baruník (2019) or Baur

and Dimpfl (2019).

The main practical application of models for forecasting conditional distributions is risk management or portfolio management. A typical goal is to forecast a risk metric such as Value-at-Risk (VaR), i.e., quantiles of a conditional distribution. Guermat and Harris (2002) use GARCH and EWMA with Student's t distribution for capturing heavy tails and forecasting VaR. Another example is Bali et al. (2008), who study the role of time-varying skewness and kurtosis in the quality of VaR forecasting. In the last decade or two, there has been a slow shift from VaR to Expected Shortfall (ES) as a preferred metric. ES is an expectation of loss conditioned by a loss exceeding a certain quantile. For example, what is the expected loss among the 1% worst losses? Therefore, ES requires a good fit to the tail of the conditional distribution. Examples of ES forecasting include work such as Ergün and Jun (2010), Du and Escanciano (2017), or Meng and Taylor (2020).

Another application of models of conditional distribution, besides forecasting, is statistical inference and testing economic hypotheses. The modeling framework introduced relies on a two-step estimation procedure, with non-parametric diurnal distributional adjustment as the first step. The formal statistical inference requires generalizing the work of Newey (1994) on asymptotic properties of semiparametric estimators from identically independently distributed data to time series data. Developing asymptotic theory for semiparametric estimators applied to time series data is outside the scope of our work; our focus is entirely on improving forecasting performance.

In the applied section of our work, we analyze fifty large-cap US stocks. The preliminary analysis shows diurnal distribution patterns beyond volatility. With the aim of utilizing the patterns, we implement the GARCH and EGARCH models introduced by Bollerslev (1986) and Nelson (1991), respectively, within our framework. To estimate the marginal cumulative distribution functions for the PIT-based adjustment, we use two-dimensional kernel estimation: one dimension representing returns and the other representing intraday time, capturing smooth distributional changes throughout the day. The non-parametric PIT allows high flexibility in potential distributional shapes throughout the day. We compare the distributional fit of EGARCH with the proposed distributional adjustment

against EGARCH with multiplicative volatility adjustment (Engle & Sokalska, 2012) and find significant distributional fit improvement across all fifty analyzed stocks in one-period-ahead out-of-sample distributional forecasts. The models used for the one-period-ahead out-of-sample forecasting are re-estimated annually. In addition to the forecasting assessment with annual re-estimation, we assess practical forecasting performance. We perform backtesting – a moving window exercise with regular re-estimation with every new observation, and predict ES one period ahead. Due to the computational complexity of backtesting, we use the GARCH model instead of the EGARCH, and analyze only four arbitrarily chosen stocks. The results of both exercises show superior out-of-sample forecasting performance and demonstrate the importance of distributional adjustment.

This paper is organized as follows. In section 1.2, we introduce the intraday setup and modeling framework. In section 1.3, we perform a descriptive analysis and preliminary analysis of potential distributional patterns. Then, in section 1.4, we apply the framework to the EGARCH model, estimate the resulting distributionally adjusted EGARCH (DA-EGARCH), and discuss properties of the model. In section 1.5, we benchmark DA-EGARCH against multiplicative EGARCH-t (MEGARCH-t) in terms of VaR and ES forecasting. A conclusion follows in section 1.6.

1.2 Modeling Framework

1.2.1 Intraday Notation

Our empirical work is based on regularly observed thirty-minute returns calculated using open and close prices. For convenience, we introduce two types of indices; a time index representing order of observations in our dataset, and an intra-day double index representing day and time of the day.

Let $t \in \{1, 2, ..., T\}$ be a time index, where T is the total number of observations. We define separate indices for intraday period $\tau(t) \in \{1, 2, ..., \tau_{\max}\}$ and days $d(t) \in \{1, 2, ..., D\}$, where $\tau_{\max}$ is number of intraday periods within each day, and

D is the number of days in the dataset. More precisely,

$$t(d,\tau) = (d-1)\tau_{\max} + \tau, \tag{1.2.1}$$

$$d(t) = \lceil t/\tau_{\max} \rceil\,,\ \tau(t) = t - (d(t)-1)\,\tau_{\max}, \tag{1.2.2}$$

where $\lceil\cdot\rceil$ is the ceiling function. We use both indexings interchangeably. To keep our notation light, we will use t, τ, and d without indicating the input arguments.

For each day and intraday period, we define (log-)returns by

$$r_{d,\tau} = \begin{cases} \log\left(\frac{\text{Close}_{d,1}}{\text{Open}_{d,1}}\right) & \text{for } \tau = 1 \\ \log\left(\frac{\text{Close}_{d,\tau}}{\text{Close}_{d,\tau-1}}\right) & \text{for } \tau > 1 \end{cases} \tag{1.2.3}$$

where $\text{Open}_{d,\tau}$ and $\text{Close}_{d,\tau}$ are open and close prices. We drop the overnight returns as is common practice in the literature. We define, instead, the first return as a return between the moment of the market opening and the end of the first period.

1.2.2 Distributional Adjustment and PIT-based Processes

Our idea is based on the assumption that all diurnal patterns are in marginal distributions and can be filtered by a time-invariant filter. More precisely, we assume there is a function $\mathcal{D}(x,\tau)$, with existing inversion for a given τ, such that

$$r_{d,\tau} = \mathcal{D}(x_t,\tau), \tag{1.2.4}$$

where x_t is latent series without any diurnal patterns. This assumption alone is not sufficient to uniquely identify the function $\mathcal{D}$.

We propose to use the probability integration transform (PIT) for adjustment. Let $F(r,\tau)$ be a marginal cumulative distribution function (cdf) of returns at intraday time τ. Then, any distributional diurnal patterns in the marginal distribution can be fully removed by transforming the returns as

$$p_t = F(r_{d,\tau},\tau). \tag{1.2.5}$$

The next step is to propose a time series model for the conditional distribution of p_t conditioned on $\mathcal{F}_{t-1}$, where $\mathcal{F}_{t-1}$ is all available information at $t-1$. The series p_t is free of any distributional diurnal patterns, but using PIT affects

characteristics that are not relevant to the diurnal patterns. In particular, the marginal distribution of p_t is marginally uniform with a support $\text{Supp}(p_t) = (0, 1)$. Since the marginal distribution of p_t is known, the model should be consistent with it. One such candidate could be a beta distribution with time-varying parameters. However, such a process is not common in the financial literature, and interpreting the model parameters in comparison to traditional GARCH-like models could be challenging.

Therefore, we propose to use a commonly used process in finance for which we can find a marginal cdf and apply PIT to the process to obtain a candidate process for p_t. The idea is identical to PIT adjustment, but with no diurnal patterns. Assume a latent process x_t with known conditional distribution and existing marginal distribution, e.g., EGARCH or GARCHSK (Jondeau & Rockinger, 2003). Denote the marginal cdf of x_t as G, and a marginal probability density function (pdf) as g. From a process x_t with a known form of conditional pdf $g_{x_t|\mathcal{F}_{t-1}}(x)$, we can construct a process p_t as

$$p_t := G(x_t), \tag{1.2.6}$$

with a conditional density

$$k_{p_t|\mathcal{F}_{t-1}}(p) = g_{x_t|\mathcal{F}_{t-1}}\left(G^{-1}(p)\right)\left(\left.\frac{dG^{-1}(s)}{ds}\right|_{s=p}\right). \tag{1.2.7}$$

The conditional density follows directly from properties of PIT, see lemma A.1.1 in the Appendix.

The process p_t, by construction, has a uniform marginal distribution and hence satisfies the requirement imposed by the PIT-based diurnal adjustment. The function $\mathcal{D}(\cdot, \tau) = F^{-1}(G(\cdot), \tau)$, where F^{-1} is the inverse function of F in the first argument, captures the relation between observed returns and the "return-like" latent series

$$x_t = G^{-1}\Big(F(r_{d,\tau}, \tau)\Big) \tag{1.2.8}$$

that is adjusted for distributional diurnal patterns. The relationship (1.2.8) represents the core idea of the proposed adjustment.

1.2.3 Estimation

The modeling framework relies on estimation in two steps. First, we have to estimate the marginal cdfs of returns. The estimated marginal cdfs are used as in (1.2.5) to calculate generated series $\widehat{p_t}$. Using a parametric distribution for F is not a good option, as a potential misspecification can generate x_t with diurnal patterns in mean and variance. Since we do not want to impose distributional assumptions, we estimate the marginal cdf nonparametrically. The second step uses the generated series $\widehat{p_t}$ to estimate the parametric time series model. Technically speaking, the adjustment is model-free, since we estimate the marginal cdf of returns non-parametrically. However, there is an advantage to filtering and visualizing model-based $\widehat{x_t}$ instead of $\widehat{p_t}$. The charts of $\widehat{p_t}$ are hard to understand intuitively. On the other hand, the model-based filtered series $\widehat{x_t}$ is, in nature, closer to what can be intuitively understood as an adjusted series of returns.

We use kernel density-based estimation of the marginal cdf. Let $K(\cdot)$ be the kernel function, and let b_D be bandwidth with the natural assumption that $b_D \xrightarrow{D\to\infty} 0$. Then the estimator of marginal cdf at period τ is

$$\widehat{F}(r,\tau) = \frac{\sum_{d=1}^{D} \int_0^r K\left(\frac{u-r_{d,\tau}}{b_D}\right) du}{D b_D}. \tag{1.2.9}$$

A straightforward usage of a separate kernel density estimator for each intraday period may be unstable, especially in the tails of the distribution, since the number of observations is increasing only with the number of days, and there is no gain from increasing frequency.

Therefore, we propose to use a weighted sum of kernel density estimates of individual intraday periods. Denote the weights by $w(\tau,\tau^*)$, where τ is the period for which we estimate the cdf and τ^* is the period of the univariate kernel cdf estimator. Then the estimator is

$$\widetilde{F(r,\tau)} = w(\tau,\tau)\widehat{F}(r,\tau) + \sum_{\tau^*\neq\tau} w(\tau,\tau^*)\widehat{F}(r,\tau^*). \tag{1.2.10}$$

Naturally, we need constraints $\sum_{\tau^*} w(\tau,\tau^*) = 1$ and $w(\tau,\tau^*) \geq 0$ to ensure that w are legitimate weights. Additionally, we require $w(\tau,\tau^*) \xrightarrow{D\to\infty} 0$, and $w(\tau,\tau) \xrightarrow{D\to\infty} 1$, which gives us consistency as long as the individual cdf estimators are consistent. The individual kernel estimators in the time series setup are

consistent under fairly general assumptions; see Wu et al. (2010). This estimator is generally biased unless there are no diurnal patterns, but it can be argued that returns occurring at similar intraday periods have similar marginal distributions, and hence that the linear combination of kernel estimates has lower variance and relatively negligible bias for reasonably chosen weights.

There are $\tau_{\max}^2$ weights in total. In order to avoid overparametrization for higher frequencies, we capture them non-parametrically by smoothing the densities throughout the day. The smoothing weights are

$$w(\tau, \tau^*) = \frac{L\left(\frac{\frac{\tau}{\tau_{\max}} - \frac{\tau^*}{\tau_{\max}}}{c_D}\right)}{\sum_{\tau^*=1}^{\tau_{\max}} L\left(\frac{\frac{\tau}{\tau_{\max}} - \frac{\tau^*}{\tau_{\max}}}{c_D}\right)}, \tag{1.2.11}$$

where $c_D \xrightarrow{D\to\infty} 0$ is bandwith, and $L(\cdot)$ is the kernel. The intraday periods are scaled by $\tau_{\max}$ to make the estimates standardized across different frequencies.

It would be reasonable to have different bandwidth c_D for different periods of the day, since there are typically larger changes in consequent marginal distributions in the morning than at noon. Therefore, it is desired to have a narrower bandwidth c_D in the morning and a wider bandwidth for the rest of the day, depending on how fast the changes are. However, from an empirical standpoint, it is mainly the marginal volatility that seems to be the source of sharp changes in distribution – volatility drops quickly in the morning and changes slowly for the rest of the day. In the case of other marginal characteristics, as documented in section 1.3.2, we observe a more stable rate of change. Therefore, we standardize the returns and keep the bandwidth constant throughout the day instead. In other words, our estimates of $F(r, \tau)$ first adjust volatility by the multiplicative adjustment, and the kernel density estimate captures any remaining distributional patterns.

In our empirical work, we use the Gaussian kernel, which yields pdf and cdf

estimators

$$\widetilde{f}(r,\tau|b_D,c_D)=\frac{\sum_{\tau^*=1}^{\tau_{\max}}\sum_{d=1}^{D}\phi\left(\frac{\frac{\tau}{\tau_{\max}}-\frac{\tau^*}{\tau_{\max}}}{c_D}\right)\phi\left(\frac{r/\widehat{S}_\tau-r_{d,\tau^*}/\widehat{S}^*_\tau}{b_D}\right)}{\widehat{S}_\tau b_D\sum_{\tau^*=1}^{\tau_{\max}}\phi\left(\frac{\frac{\tau}{\tau_{\max}}-\frac{\tau^*}{\tau_{\max}}}{c_D}\right)}, \tag{1.2.12a}$$

$$\widetilde{F}(r,\tau|b_D,c_D)=\frac{\sum_{\tau^*=1}^{\tau_{\max}}\sum_{d=1}^{D}\phi\left(\frac{\frac{\tau}{\tau_{\max}}-\frac{\tau^*}{\tau_{\max}}}{c_D}\right)\Phi\left(\frac{r/\widehat{S}_\tau-r_{d,\tau^*}/\widehat{S}^*_\tau}{b_D}\right)}{\sum_{\tau^*=1}^{\tau_{\max}}\phi\left(\frac{\frac{\tau}{\tau_{\max}}-\frac{\tau^*}{\tau_{\max}}}{c_D}\right)}, \tag{1.2.12b}$$

$$\widehat{S}_\tau=\sqrt{D^{-1}\sum_{d=1}^{D}(r_{d,\tau}-\overline{r}_\tau)^2}, \tag{1.2.12c}$$

where ϕ is the standard normal pdf, Φ is the standard normal cdf, and $\overline{r}_\tau$ is the average return at intraday time τ. Note that the series of bandwidths for weights may have different rates of convergence. For the series b_D, we require the natural condition $b_D D \xrightarrow{D\to\infty} \infty$, while $c_D D \xrightarrow{D\to\infty} 0$ will not jeopardize the consistency. In fact, setting $c_D = 0$ will set all cross-weights to zero and reduce the weighted estimator to a standard univariate kernel estimator for intraday time τ.

We select the bandwidths by cross-validation. In the empirical part, we maximize cross-validated marginal log-likelihood, i.e.,

$$[b_D,c_D]^T=\underset{[b,c]^T}{\operatorname{argmax}}\sum_{d,\tau}\log\left(\widetilde{f}^{-d,\tau}(r_{d,\tau},\tau|b,c)\right), \tag{1.2.13}$$

where $\widetilde{f}^{-d,\tau}$ is marginal density estimated without return $r_{d,\tau}$.

Using the marginal cdf estimates, we generate a series

$$\widehat{p}_t=\widetilde{F}(r_{d,\tau},\tau|b_D,c_D), \tag{1.2.14}$$

and maximize the log-likelihood

$$\sum_{t=1}^{T} l_t, \text{ where} \tag{1.2.15a}$$

$$l_t=\log\left(g_{x_t|\mathcal{F}_{t-1}}\left(G^{-1}(\widehat{p}_t)\right)\right)+\log\left(\left.\frac{dG^{-1}(s)}{ds}\right|_{s=\widehat{p}_t}\right) \tag{1.2.15b}$$

to estimate the parameters of the time series model.

An asymptotic theory for this class of models, to our best knowledge, is not developed in the literature, and developing such a theory goes far beyond the scope of this paper. Newey (1994), develops an asymptotic theory for semi-parametric

estimators under the i.i.d. assumption. As he notes, his results could be generalized for time series setup under certain regulatory conditions. However, the problem arises from specifying enough general regulatory conditions imposed on time series that would yield a feasible influence function that captures the impact of $\hat{p}_t$ being estimated (generated series) on the second-step of the estimation.

1.2.4 Numerical Algorithm for Marginal Distribution

It is numerically challenging to use a model for p_t based on standard parametric distributional models that are transformed by their marginal cdfs, since models like ARCD (B. E. Hansen, 1994) or even normal GARCH(1,1) typically do not have analytical marginal cdf or quantile function. The marginal cdf or quantile function can be easily obtained by Monte Carlo (MC) simulation. However, for the maximum likelihood estimation (1.2.15b), we also need dG^{-1}/dp, and simulated derivatives are highly inaccurate.

The inverse function theorem allows us to avoid numerical calculation of derivatives dG^{-1}/dp as

$$\frac{dG^{-1}(p)}{dp} = \frac{1}{dG(x)/dx}. \tag{1.2.16}$$

Therefore, we need only marginal density and quantile function. Both can be easily simulated for processes with known distributions conditioned on the past. To be more precise, consider a series of independent shocks $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ that are adapted to $\mathcal{F}_t$ and the process

$$x_t = Y(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, ...), \tag{1.2.17}$$

has known conditional cdf $G_{x_t|\mathcal{F}_{t-1}}$. By the law of iterated expectations, we can find the marginal cdf as

$$\begin{aligned} G(x) &= \mathbb{E}\left[\mathbb{I}\left\{x_t \leq x\right\}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathbb{I}\left\{x_t \leq x\right\}|\mathcal{F}_{t-1}\right]\right] \\ &= \mathbb{E}\left[G_{x_t|\mathcal{F}_{t-1}}(x)\right] \end{aligned} \tag{1.2.18}$$

Conditioning by the past in the context of standard distributional models usually means conditioning by time-varying moments or parameters that are

functions of past shocks. We can simulate the distribution of the time-varying moments or parameters to approximate the resulting expectation (1.2.18). Since Leibniz's conditions for changing the order of differentiation and integration are satisfied for any reasonable model in the context of modeling returns, we can easily obtain the density as well.

To demonstrate the idea, consider, for example, the normal EGARCH(1,1) defined as

$$x_t = \sigma_t \varepsilon_t, \ \epsilon_t \sim N(0,1), \tag{1.2.19a}$$

$$\log(\sigma_t^2) = \beta_1 \log(\sigma_{t-1}^2) + \beta_2 \varepsilon_{t-1} + \beta_3 |\varepsilon_{t-1}|, \tag{1.2.19b}$$

where $x_t = \sigma_t \varepsilon_t$ corresponds to (1.2.17). We would like to solve numerically for marginal cdf and pdf of x_t. The key step is to use the information $\mathcal{F}_{t-1}$ to filter all the necessary conditional parameters for the conditional distribution of x_t. In this case, σ_t is the only time-varying parameter.[2] The conditional distribution is

$$x_t | \sigma_t = \sigma \sim N(0, \sigma^2). \tag{1.2.20}$$

Using (1.2.18), the marginal cdf and pdf can be expressed respectively as

$$G(x) = \mathbb{E}\left[\Phi\left(\frac{x}{\sigma_t}\right)\right], \tag{1.2.21a}$$

$$g(x) = \mathbb{E}\left[\frac{1}{\sigma_t}\phi\left(\frac{x}{\sigma_t}\right)\right]. \tag{1.2.21b}$$

We can easily simulate the marginal distribution of σ_t and approximate the cdf and pdf as

$$G(x) = \mathbb{E}\left[\Phi\left(\frac{x}{\sigma_t}\right)\right] \approx \frac{1}{n_\sigma}\sum_{i=1}^{n_\sigma} \Phi\left(\frac{x}{\sigma_T^i}\right), \tag{1.2.22a}$$

$$g(x) \approx \frac{1}{n_\sigma}\sum_{i=1}^{n_\sigma} \frac{1}{\sigma_T^i}\phi\left(\frac{x}{\sigma_T^i}\right), \tag{1.2.22b}$$

where $\sigma_T^1, \sigma_T^2, ..., \sigma_T^{n_\sigma}$ are simulated volatilities.[3] More precisely, each σ_T is obtained by iterating the law of motion T times with random realizations of ε_t until the initial condition σ_0 becomes negligible. This approximation, in the case of GARCH,

[2] An example of GARCHK with time-varying kurtosis can be found in Appendix A.2.

[3] The upper index denotes trajectory in this context, where T denotes the last simulated value of a particular trajectory.

converges fast, and even a few hundred simulated trajectories provide precise results.

Note that one could potentially simulate directly $x_T = \sigma_T \varepsilon_T$ and use the empirical distribution function to obtain $G(x)$. However, a numerically precise solution of $G(x)$ in the tails of the distribution would require a prohibitive number of simulated observations. The calculation of density $g(x)$ would be even more complicated, as the empirical distribution function is a step function, and we cannot simply calculate numerical derivatives by taking differences over a small interval. One could use a kernel density estimate, but an accurate calculation of the density using the kernel density estimate requires substantially more simulated observations of x_T than using the known parametric conditional distribution for $x_T|\sigma_T = \sigma$. The equality (1.2.18) is absolutely essential as it makes our framework numerically feasible.

The equality (1.2.18) allows fast simulation of cdf and pdf, but the quantile function has to be calculated numerically from the simulated cdf. The distributional adjustment by the marginal cdf estimates yields generated series $\widehat{p}_t$, but the corresponding latent "return-like" process x_t is unknown. Therefore, to compute the log-likelihood, we have to numerically solve for the corresponding values of x_t to be able to evaluate the quantile function and the derivative at $\widehat{p}_t$ in (1.2.15b). A brute force approach, going observation by observation and running some root-finding algorithm like Newton–Raphson, is extremely slow. However, we can use stylized facts to propose a reasonably fast algorithm for solving the quantile function for a fine grid of potential input argument values. It is reasonable[4] to expect that the marginal distribution of x_t will be distributed around 0, i.e., we expect $G^{-1}(0.5) \approx 0$. We also know the derivative of the quantile function at $p = 0.5$; thus, we can use it for linear approximation of points close to $p = 0.5$. Finally, a random variable transformed by PIT is uniformly distributed. Therefore, it is optimal to use an equidistant grid for p. Using these three facts,

[4]We are free to choose marginal median, since marginal properties of x_t do not impact p_t. It may be challenging to choose parameters for some processes such that the median is close to 0. Even then, it is a reasonable guess for the models used to model asset returns. However, if there is a reasonable expectation that the median is significantly different from zero, then the second step of the algorithm can be refined by running a root-finding algorithm to find the median.

we propose the following algorithm. To simplify notation, let $Q(p) := G^{-1}(p)$ be the quantile function, $p \in (0,1)$.

1. Set a fine equidistant grid $p_0, p_1, ..., p_{\lceil \frac{m}{2} \rceil}, ..., p_{m-1}, p_m$ with an odd number of points[5] for $p_i \in (0,1)$, where $0 < p_0 < p_1 < ... < p_{\lceil \frac{m}{2} \rceil} = 0.5 < ... < p_m < 1$.

2. Set $Q^0(p_{\lceil \frac{m}{2} \rceil}) := 0$.

3. Calculate

$$Q^0(p_{\lceil \frac{m}{2} \rceil - i}) = Q^0(p_{\lceil \frac{m}{2} \rceil - i + 1}) - \frac{p_{\lceil \frac{m}{2} \rceil - i + 1} - p_{\lceil \frac{m}{2} \rceil - i}}{g\left(Q^0(p_{\lceil \frac{m}{2} \rceil - i + 1})\right)},$$
$$Q^0(p_{\lceil \frac{m}{2} \rceil + i}) = Q^0(p_{\lceil \frac{m}{2} \rceil + i - 1}) + \frac{p_{\lceil \frac{m}{2} \rceil + i} - p_{\lceil \frac{m}{2} \rceil + i - 1}}{g\left(Q^0(p_{\lceil \frac{m}{2} \rceil + i - 1})\right)},$$

   for $i = 1, 2, ..., \lceil \frac{m}{2} \rceil$.

4. Repeat

$$\begin{pmatrix} Q^j(p_0) \\ Q^j(p_1) \\ \vdots \\ Q^j(p_m) \end{pmatrix} = \begin{pmatrix} Q^{j-1}(p_0) \\ Q^{j-1}(p_1) \\ \vdots \\ Q^{j-1}(p_m) \end{pmatrix} - \begin{pmatrix} \frac{G(Q^{j-1}(p_0)) - p_0}{g(Q^{j-1}(p_0))} \\ \frac{G(Q^{j-1}(p_1)) - p_1}{g(Q^{j-1}(p_1))} \\ \vdots \\ \frac{G(Q^{j-1}(p_m)) - p_m}{g(Q^{j-1}(p_m))} \end{pmatrix}$$

   for $j = 1, 2, ...,$ maxiter, where maxiter is the maximum number of iterations, or until some convergence criteria are met.

A convergence criterion may be, for example,

$$\left\| \left[G\Big(Q^j(p_0)\Big) - p_0, G\Big(Q^j(p_1)\Big) - p_1, ..., G\Big(Q^j(p_m)\Big) - p_m \right]^\intercal \right\| \leq error, \quad (1.2.23)$$

where $\|.\|$ is the euclidean norm and *error* is a reasonable small number. The algorithm has, in a certain sense, two phases. During the first phase, corresponding to points two and three, an initial guess based on a linear approximation of the quantile is computed. The fourth point refines the precision of the guess from the first phase using the Newton–Raphson method. The Newton–Raphson method can be replaced by other root-solving algorithms, but the derivatives of the quantile function are known; thus, it is reasonable to take advantage of it.

[5] We use an odd number of points only to simplify the notation. The idea works just as well with an even number of points.

The euclidean norm can be replaced by more or less strict norms. In our implementation, we use a more strict max-norm that returns the largest difference. Unless stated explicitly otherwise, we run the algorithm with the following settings

- $m = 2T$,
- $error = 10^{-6}$,
- number of conditional variance trajectories $n_\sigma = 1,000$,
- number of simulated periods: 200.

We set $m = 2T$ to a construct grid covering the values $\widehat{p}_t$ that are close to the boundaries 0 or 1. In close proximity to 0 or 1 a small difference in p potentially translates to large differences in $Q(p)$ and potentially a numerical imprecision in likelihood calculation. Since we use a Guassian kernel density estimation for the adjustment, the smallest value of $\widehat{p}_t$ will be larger than $1/2T$ and the largest $\widehat{p}_t$ will be smaller than $1 - 1/2T$. Therefore, all values of $\widehat{p}_t$ should be covered by the range (p_0, p_m). In our experience, the 4th step is very rarely repeated more than once when the number of gridpoints $m \geq 1000$, as the precision of the linear approximation grows with m. The speed of our proposed algorithm is crucial for practical applications. For further discussion of computation speed, see Appendix A.3.

1.2.5 Similarities with Multiplicative Adjustment

Let us look at a trivial toy example, where the proposed distributional adjustment is equivalent to a multiplicative adjustment.

Consider a true data-generating process

$$r_{d,\tau} = S_\tau \varepsilon_t, \quad \varepsilon \sim N(0,1). \tag{1.2.24}$$

The multiplicative adjustment estimates the scale factors

$$\widehat{S}_\tau = \sqrt{\frac{1}{D}\sum_{d=1}^{D}\left(r_{d,\tau} - \overline{r}_\tau\right)^2} \xrightarrow{P} S_\tau, \tag{1.2.25}$$

and adjusts the series as

$$\widehat{r}_t^{adjusted} = \frac{r_{d,\tau}}{\widehat{S}_\tau} \xrightarrow{P} \frac{r_{d,\tau}}{S_\tau} = \varepsilon_t = r_t^{adjusted}. \tag{1.2.26}$$

The adjustment standardizes the series, i.e., $Var\left[r_t^{adjusted}\right] = 1$.

Alternatively, we can make adjustments with respect to some reference intraday period as

$$\widehat{r}_t^{adjusted} = r_{d,\tau}\frac{\widehat{S}_{ref}}{\widehat{S}_\tau} \xrightarrow{P} S_{ref}\varepsilon_t = r_t^{adjusted}, \tag{1.2.27}$$

yielding $Var\left[r_t^{adjusted}\right] = S_{ref}^2$, where S_{ref} is the standard deviation for the reference intraday time. Another potential option is using geometric mean as a reference value, yielding $Var\left[r_t^{adjusted}\right] = (S_1 S_2 ... S_{\tau_{max}})^{2/\tau_{max}}$. In general, we can scale the series to any arbitrary reference value. The chosen reference value affects the interpretation of the further analysis based on the adjusted series.

Let us now compare the multiplicative adjustment to the proposed distributional adjustment. Adjusting the process $r_{d,\tau}$ via the proposed framework uses non-parametric estimation of the marginal cdf $F(r,\tau)$. In this example, the true marginal distribution is $N(0, S_\tau^2)$; thus, using a consistent estimator of $\widehat{F}$ yields

$$\widehat{p}_t = \widehat{F}(r_{d,\tau},\tau) \xrightarrow{P} \Phi\left(\frac{r_{d,\tau}}{S_\tau}\right) = p_t. \tag{1.2.28}$$

For modeling p_t, we chose a latent "adjusted $r_{d,\tau}$-like" process

$$x_t = \sigma^x \varepsilon_t^x,\ \varepsilon_t^x \sim N(0,1). \tag{1.2.29}$$

The model process for p_t is then constructed as

$$p_t = \Phi\left(\frac{x_t}{\sigma^x}\right) = \Phi\left(\frac{\sigma^x\varepsilon_t^x}{\sigma^x}\right) = \Phi(\varepsilon_t^x). \tag{1.2.30}$$

Note that the process p_t is invariant to values of $\sigma^x > 0$. In general, given by PIT, the marginal distribution of x_t does not impact the distribution of p_t. We are free to choose the values of $\sigma^x > 0$ the same way as we can choose the reference value for the multiplicative decomposition. If we set $\sigma^x := \widehat{S}_{ref}$, the resulting series x_t and $r_t^{adjusted}$ will be asymptotically equal.

The additive adjustment gives the freedom of centering the series around an arbitrary value, and the multiplicative adjustment gives the freedom to choose an arbitrary scale. In our framework, following from the properties of PIT, one has the freedom to choose the marginal distribution of x_t. For example, level, scale, and kurtosis. Consider process

$$x_t = \alpha_x + \sigma^x\varepsilon_t^x,\ \varepsilon_t^x \sim t(v_x), \tag{1.2.31}$$

Table 1.1 Traded amount as proxy for liquidity.

	Most liquid	Median liquid	Least liquid
Traded cash	\$9,891,988	\$1,168,095	\$100,009

An average traded amount over 1 minute during the main trading session.

where $\sigma^x > 0$, t is Student's t-distribution and v_x are degrees of freedom. The process (1.2.31) transformed by PIT will yield an identical p_t process as (1.2.29) transformed by PIT.

1.3 Data

1.3.1 The Dataset

The data used in the empirical study comprises fifty time series of highly liquid stocks obtained from the Kibot[6] database for the period from November 29, 2015 to November 8, 2019. The time series includes open and close prices observed over one-minute intervals. We aggregate the one-minute data to thirty-minute data, then the prices are transformed into log returns according to (1.2.3).

The dataset comprises stocks with the following symbols: AAPL, ADBE, AMD, AMZN, AUY, AVGO, BA, BABA, BAC, BHC, BIDU, BP, BRK, C, CAT, CELG, CLF, CRM, CSCO, CTL, DB, DIS, F, FB, FCX, GE, GOLD, GOOG, GOOGL, GS, HD, CHK, IBM, INTC, JD, JNJ, JPM, KMI, MCD, MRO, MSFT, MU, NFLX, NOK, NVDA, PBR, PYPL, QCOM, RDS.A, and RIO. Table 1.1 summarizes liquidity in terms of trade volume during the main trading session. We use all fifty stocks to obtain some of our aggregated results. In addition, we study in more detail four arbitrarily chosen symbols – AAPL, JPM, MCD, and MSFT.

[6] www.kibot.com (accessed 15 November 2019).

1.3.2 Preliminary Exploratory Analysis

Table 1.2 displays several sample statistics of four arbitrarily chosen series. Since the statistics are computed from all observations of the main trading session, it represents average values throughout the day. Data poses, by a rule of thumb, no significant unconditional skewness. In contrast, the returns of all stocks are leptokurtic.

Table 1.2 Descriptive statistics of log-returns for Apple, JPMorgan Chase, McDonald's, and Microsoft.

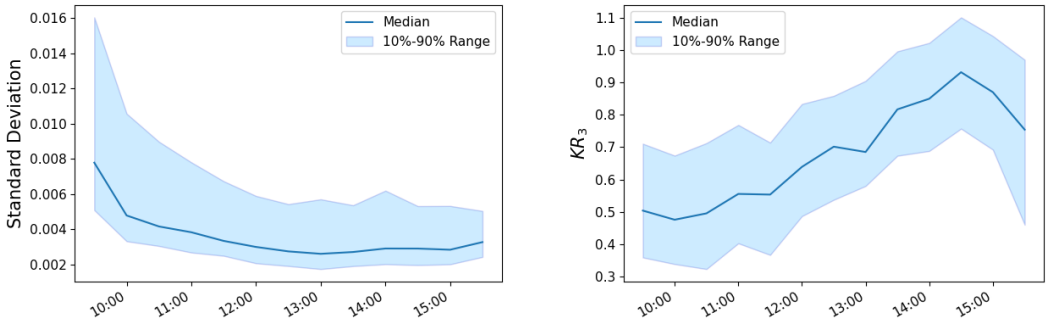
	AAPL	JPM	MCD	MSFT
Number of observations	13,195	13,195	13,195	13,195
Days (D)	1,015	1,015	1,015	1,015
Intraday periods (τ_{max})	13	13	13	13
Mean	3.91×10^{-5}	1.33×10^{-5}	7.79×10^{-6}	8.71×10^{-6}
Standard deviation	3.17×10^{-3}	2.99×10^{-3}	2.43×10^{-3}	3.12×10^{-3}
Skewness	0.287	0.239	0.431	−0.212
Excess kurtosis	12.997	6.931	16.592	6.433
KR_3	1.026	0.783	0.914	0.935
Minimum	−0.026	−0.023	−0.027	−0.026
Maximum	0.054	0.033	0.041	0.027

All series start on November 29, 2015, at 9:30, end on November 8, 2019, at 16:00, and are observed in 30-minutes long intervals. Excess kurtosis is computed using the Fisher formula subtracting 3, i.e., 0 is the excess kurtosis of the normal distribution. KR_3 is a robust alternative measure of kurtosis by Kim and White (2004).

Now, let us focus more on variance with kurtosis and their diurnal patterns. The average sample kurtosis is relatively stable due to the high number of observations. However, when we attempt to estimate marginal kurtosis for each intraday period separately, the number of observations drops to roughly a thousand. Since the moment-based estimation of kurtosis may be heavily biased in an upward direction in the presence of outliers, any plots of diurnal patterns in kurtosis in our dataset are unreadable due to random extreme spikes. There was no clearly visible pattern in kurtosis. Therefore, we use a robust alternative measure of kurtosis by Kim

Figure 1.1 Diurnal patterns of volatility and kurtosis of 50 stocks.

(a) Volatility **(b)** Kurtosis



and White (2004) that is given as

$$KR_3(\tau) = \frac{U_{0.05}(\tau) - L_{0.05}(\tau)}{U_{0.5}(\tau) - L_{0.5}(\tau)} - 2.59, \tag{1.3.1a}$$

$$U_\alpha(\tau) = \frac{1}{\alpha}\int_{1-\alpha}^{1} F^{-1}(y,\tau)\,dy, \quad L_\alpha(\tau) = \frac{1}{\alpha}\int_{0}^{\alpha} F^{-1}(y,\tau)\,dy, \tag{1.3.1b}$$

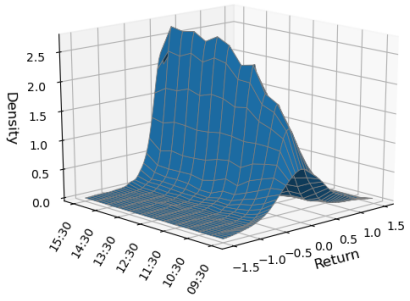
for the intraday period τ and estimated by their sample counterparts. Under normality, it holds that $KR_3(\tau) = 0$.

We estimate the marginal standard deviation and KR_3 for each intraday period of the fifty stocks. The results are summarized in Figure 1.1. The plot demonstrates a clear pattern of increasing kurtosis throughout the day, followed by a drop at the end of the main trading session. The pattern seems stable, since it is clearly visible from the cross-sectional quantiles of KR_3 represented by the light blue area. It is important to note that the metric KR_3 measures "something like kurtosis," not the kurtosis itself. Therefore, the same pattern may not be present in the actual kurtosis.

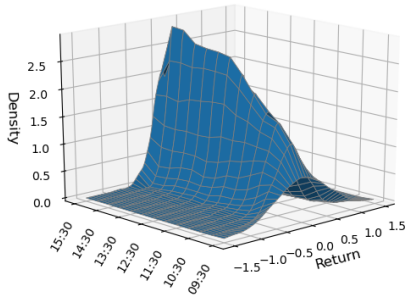
As was described in the previous section, we intend to capture these diurnal patterns and possibly other distributional patterns non-parametrically. We use the proposed first step estimator to visualize estimated marginal distributions throughout the day. Since the diurnal pattern in volatility is much more prominent than other potential diurnal patterns, we also estimate the densities of standardized series, where each return is divided by the corresponding intraday marginal standard deviation. Thus, the standardized density captures the remaining

Figure 1.2 Non-parametric estimates of marginal densities for AAPL and MCD.

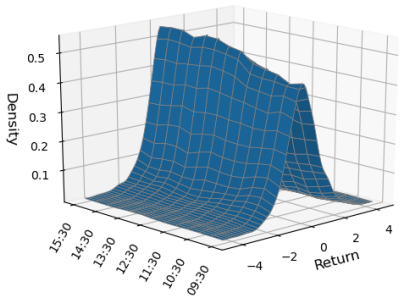
(a) AAPL



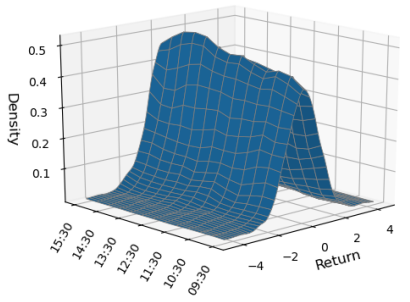
(b) MCD



(c) AAPL – Standardized



(d) MCD – Standardized



Standardized densities are estimated from series, where returns are standardized by corresponding marginal standard deviation.

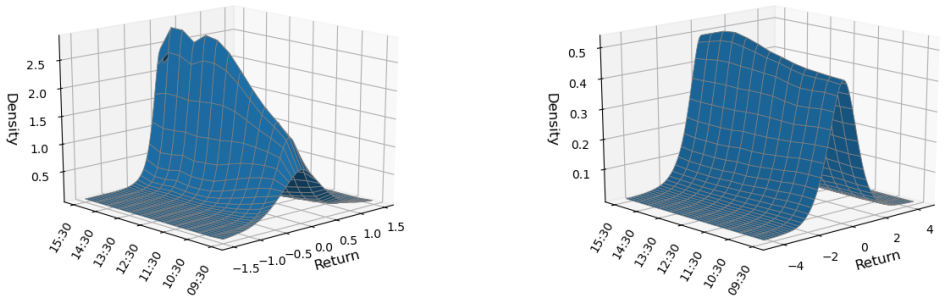
distributional diurnal patterns after the diurnal pattern in volatility is removed by scaling.

The resulting plots for AAPL and MCD are in Figure 1.2. The densities of the returns are changing significantly throughout the day; most of the pattern is given by volatility alone. However, as the estimates for standardized series suggest, there are distributional diurnal patterns beyond volatility. The shapes are in line with the pattern observed for kurtosis. In particular, we can see that the peak of the standardized density is lower in the morning and higher in the afternoon.

A similar pattern as for AAPL and MCD is observed for all stocks in our dataset. In Figure 1.3, we take the average of densities across all considered

Figure 1.3 The average of non-parametric estimates of marginal densities of all 50 stocks.

(a) Average Density **(b)** Average Standardized Density



stocks. The average pattern is the same as the individual patterns, i.e., volatility is the main driver in the diurnal patterns, but there are diurnal patterns beyond volatility.

The distributional diurnal adjustment assumes that diurnal patterns are stable over time. To assess stability, the density in Figure 1.3 is estimated separately for each year in Appendix A.4 and discussed further.

1.4 DA-EGARCH Model

1.4.1 Model Specification

As described above, the modeling framework is applicable to conventional GARCH-like models. One of the most popular choices is the EGARCH model introduced by Nelson (1991). Engle and Sokalska (2012) used the EGARCH in an intraday setup and extended the model with scale factors for multiplicative adjustment of volatility. In this section, we use EGARCH as an example and apply it within our framework. The resulting model is addressed as distributionally adjusted EGARCH, DA-EGARCH for short.

Let us first introduce multiplicative EGARCH-t, hereafter referred to as

MEGARCH-t. The log-returns in MEGARCH-t follow

$$r_{d,\tau} = S_\tau \mu + S_\tau \sigma_t \varepsilon_t, \tag{1.4.1}$$

$$\log(\sigma_t^2) = \beta_0 + \beta_1 \log(\sigma_{t-1}^2) + \beta_2 \varepsilon_{t-1} + \beta_3 |\varepsilon_{t-1}|, \tag{1.4.2}$$

$$\varepsilon_t \sim t(v), \tag{1.4.3}$$

where μ, β_0, β_1, β_2, and β_3 are parameters of the process, the innovations ε_t follow the t distribution with v degrees of freedom, and S_τ are scale factors. Note that we keep the degrees of freedom fixed throughout the day. Preliminary analysis of using different v_τ for each time of the day showed an instability and poor performance, likely caused by a low number of observations within each intraday time period. Therefore, we keep the degrees of freedom fixed. The model is estimated in two steps. First, the diurnal scale factors are estimated by (1.2.12c) and returns adjusted as $r_{d,\tau}/\widehat{S}_\tau$. Then, the EGARCH-t process is estimated via MLE using the adjusted series.

Now, we use the same EGARCH dynamics within our framework to model the latent process x_t. We have to slightly reduce the model. First, we do not need the diurnal scale factors, as the process p_t is already adjusted for the diurnal patterns. Second, we do not need to consider time-invariant properties; μ, β_0, and v are not uniquely identified. Therefore, we are free to choose the parameters. This follows from the fact that the process p_t is constructed from x_t by PIT using the marginal distribution of x_t. We set $\mu = \beta_0 = 0$ and replace the t distribution with the standard normal distribution, i.e., $v \to \infty$. Therefore, the DA-EGARCH is given by

$$r_{d,\tau} = F^{-1}(p_t, \tau), \tag{1.4.4a}$$

$$p_t = G(x_t), \tag{1.4.4b}$$

$$x_t = \sigma_t \varepsilon_t, \tag{1.4.4c}$$

$$\log(\sigma_t^2) = \beta_1 \log(\sigma_{t-1}^2) + \beta_2 \varepsilon_{t-1} + \beta_3 |\varepsilon_{t-1}|, \tag{1.4.4d}$$

$$\varepsilon_t \sim N(0,1), \tag{1.4.4e}$$

where F^{-1} is the marginal quantile function of $r_{d,\tau}$ for intraday time τ, and G is the cdf of the marginal distribution of x_t.

The model is estimated in two steps as described in 1.2.3. First, the returns are adjusted by the marginal cdf as $\widehat{p}_t = \widetilde{F}(r_{d,\tau}, \tau | b_D, c_D)$, where $\widetilde{F}$ is estimated by

Table 1.3 Estimates of DA-EGARCH model.

	Bandwidths		EGARCH		
	b_D	c_D	β_1	β_2	β_3
AAPL	0.208	0.290	0.975	−0.035	0.162
JPM	0.226	0.223	0.976	−0.039	0.119
MCD	0.217	0.397	0.951	−0.024	0.155
MSFT	0.202	0.224	0.970	−0.061	0.165

b_D - return bandwidth, c_D - intraday time bandwidth

$\log(\sigma_t^2) = \beta_1 \log(\sigma_{t-1}^2) + \beta_2 \varepsilon_{t-1} + \beta_3 |\varepsilon_{t-1}|$

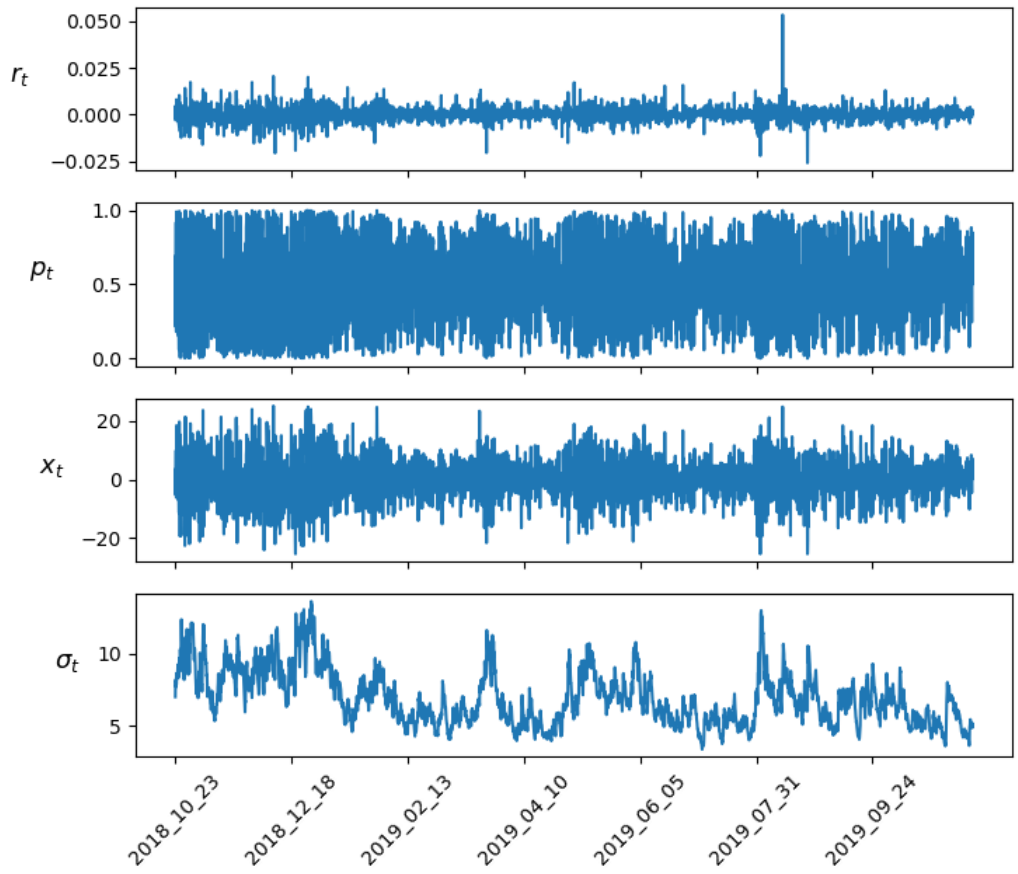
(1.2.12b) with the bandwidths selected by cross-validation. Then, the parameters β_1, β_2, and β_3 are estimated using MLE.

1.4.2 Model Estimates

The DA-EGARCH estimates for the four arbitrarily chosen stocks covering the period from November 29, 2015 to November 8, 2019, are in Table 1.3. The bandwidths for both dimensions are relatively close, with the time bandwith c_D being usually slightly higher. This is a consequence of scaling returns and intraday period indices by marginal standard deviation and number of intraday periods, respectively. Values of c_D are small, i.e., the distribution is changing throughout the day, which suggests diurnal patterns in marginal distributions beyond simple volatility patterns. The EGARCH process parameters are, with high persistence and stronger reaction to negative returns in patterns, not different from typical estimates in the literature.

Figure 1.4 shows a roughly one-year-long series of returns and filtered series. The series x_t can be considered to be a distributionally adjusted series of returns. The scale of the series x_t is completely different from that of r_t. However, we are free to choose the intercept of the log-variance law of motion, which is currently 0; therefore, we may choose an intercept such that the scales of x_t and r_t are the same when such a comparison is desired. The key observation is that the filtered series x_t are in line with stylized facts about the high persistence of volatility, and asymmetric effects of shocks (McAleer, 2014; Nelson, 1991; Rodriguez & Ruiz,

Figure 1.4 AAPL stock, filtered series of the DA-EGARCH model.



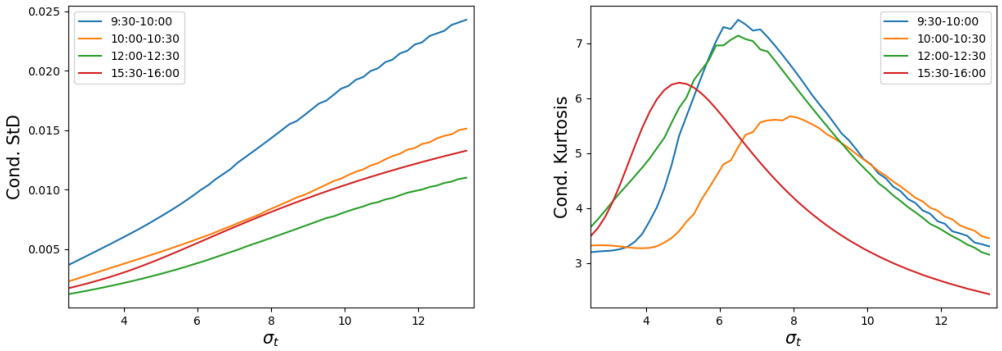
2012).

The transformation $F^{-1}\Big(G(\sigma_t\epsilon_t), \tau\Big)$ is non-linear. Therefore, the standard deviation does not map proportionally to σ_t. Furthermore, the conditional distribution of log-returns also exhibits a time-varying kurtosis that depends on σ_t. Since the transformation is systematically changing throughout the day, the relation between moments and σ_t is also changing throughout the day. We computed conditional moments of log-returns as a function of σ_t and time of the day for the DA-EGARCH model estimated for the AAPL stock presented above. The functions for other stocks may look slightly different, since the estimated marginal distributions are not identical. However, the shapes, in the overall nature, are the same for other stocks as well.

The shapes of resulting functions are presented in Figure 1.5. We have selected only 4 out of 13 intraday periods, but the functions smoothly transition from one to another. It is not surprising that the conditional standard deviation functions

Figure 1.5 AAPL stock, relation between conditional standard deviation and kurtosis of log-returns and standard deviation of the underlying normal EGARCH process in the DA-EGARCH model.

(a) Standard deviation **(b)** Kurtosis



are monotonic in σ_t, all looking very similar. The kurtosis functions have humps with peaks between 5 and 8, which is also the region where σ_t has the highest density. The hump shape is not exclusively a feature of the data but potentially a partial artifact of the model. Different specifications of x_t, e.g., processes with time-varying kurtosis, may yield different shapes. An interesting observation is potentially platykurtic distributions for the end of the day, when σ_t is high.

1.5 Distributional Fit

1.5.1 Distributional Forecasting and Precision Metrics

The main motivation for capturing the diurnal patterns in the distribution is to improve the forecasts of the conditional distribution. We assess out-of-sample performance and benchmark the proposed distributional adjustment to the common multiplicative adjustment. For the benchmarking, we do not use likelihood-based Kullback–Leibler information criteria (Kullback & Leibler, 1951) such as Akaike's information criterion (Akaike, 2003), as the likelihoods and criteria within our semiparametric framework are not comparable to the fully parametric models. Instead, we use conditional quantiles, i.e., Value-at-Risk (VaR), to assess the quality of fit.

The one-period ahead quantile predictions, conditioned on information $\mathcal{F}_{t-1}$, for the DA-EGARCH are obtained in two steps.

1. Filter $\widehat{\sigma}_t$ using the series $\widehat{x}_{t-1}, \widehat{x}_{t-2}, ..., \widehat{x}_1$, where $\widehat{x}_t = \widehat{G}^{-1}(\widehat{p}_t)$ with $\widehat{G}$ being the marginal cdf computed using the MLE estimates $\widehat{\beta_1}, \widehat{\beta_2}$, and $\widehat{\beta_3}$.

2. Calculate desired quantile $p \in (0, 1)$ as

$$Q_{x_t}(p|\widehat{\sigma}_t) = \widehat{\sigma}_t \Phi^{-1}(p), \tag{1.5.1}$$

$$\widehat{Q}_{p_t}(p|\widehat{\sigma}_t) = \widehat{G}\Big(Q_{x_t}(p|\widehat{\sigma}_t)\Big), \tag{1.5.2}$$

$$\widehat{Q}^{DA}_{r_t}(p|\widehat{\sigma}_t) = \widetilde{F}^{-1}\Big(\widehat{Q}_{p_t}(p|\widehat{\sigma}_t), \tau(t)|b_D, c_D\Big), \tag{1.5.3}$$

   where Q_{x_t}, $\widehat{Q}_{p_t}$, and $\widehat{Q}^{DA}_{r_t}$ are, respectively, DA-GARCH predictions of quantiles for x_t, p_t and r_t conditioned on $\sigma_t = \widehat{\sigma}_t$, and $\widetilde{F}$ is estimate of $r_{d,\tau}$ marginal cdf.

To assess the precision of VaR predictions, we follow Du and Escanciano (2017) and define a hit indicator variable as $h_t(p) = \mathbb{I}(r_t \leq Q_{r_t}(p))$, where $Q_{r_t}(p)$ is quantile of r_t conditioned on $\mathcal{F}_{t-1}$. For a correct conditional quantile $Q_{r_t}(p)$, the hit rate and variance are, respectively,

$$\mathbb{E}\left[h_t(p)\right] = p, \ \text{Var}[h_t(p)] = p(1-p). \tag{1.5.4}$$

The sample analog of hit rate is

$$\widehat{h}(p) = \frac{1}{T}\sum_{t=1}^{T} \mathbb{I}(r_t \leq \widehat{Q}_{r_t}(p)), \tag{1.5.5}$$

where $\widehat{Q}_{r_t}$ is the predicted quantile of r_t conditioned on $\mathcal{F}_{t-1}$.

We calculate the hit rates for all fifty stocks for DA-EGARCH and ME-GARCH-t models. Presenting them all in raw form makes interpretation infeasible, so we aggregate the hit rates across the stocks. Since the models can underforecast for some stocks and overforecast for others, we calculate absolute errors in hit rates. Let $\widehat{h^i}(p)$ be the sample hit rate for stock i. Then, the hit rate absolute error is

$$\text{HRAE}_p = \frac{1}{n}\sum_{i=1}^{n} \left|\widehat{h^i}(p) - p\right|. \tag{1.5.6}$$

In recent years, Expected Shortfall (ES) replaced VaR as a preferred risk management metric. We assess the quality of fit also in terms of ES. ES is defined as

$$\mathrm{ES}_p = \frac{1}{p}\int_0^p Q_{r_t}(\gamma)\,d\gamma, \tag{1.5.7}$$

where Q_{r_t} is the conditional quantile for r_t. To assess the quality of ES predictions, Du and Escanciano (2017) introduce cumulative violation

$$\frac{1}{p}\int_0^p h_t(u)\,du = \frac{1}{p}\Big(p - F_{r_t|\mathcal{F}_{t-1}}(r_t)\Big)\mathbb{I}\Big(r_t \le F^{-1}_{r_t|\mathcal{F}_{t-1}}(p)\Big),$$

where $F_{r_t|\mathcal{F}_{t-1}}$ is the cdf of r_t conditioned on $\mathcal{F}_{t-1}$. The expectation and variance of the cumulative violation for a correct conditional distribution are, respectively,

$$\mathbb{E}\left[\frac{1}{p}\int_0^p h_t(u)\,du\right] = \frac{p}{2}, \quad \mathrm{Var}\left[\frac{1}{p}\int_0^p h_t(u)\,du\right] = p\left(\frac{1}{3} - \frac{p}{4}\right).$$

The sample analog of the expected cumulative violation is

$$\widehat{\mathrm{CV}}(p) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{p}\Big(p - \widehat{F}_{r_t|\mathcal{F}_{t-1}}(r_t)\Big)\mathbb{I}\Big(r_t \le \widehat{F}^{-1}_{r_t|\mathcal{F}_{t-1}}(p)\Big), \tag{1.5.8}$$

where $\widehat{F}_{r_t|\mathcal{F}_{t-1}}$ is the model cdf prediction of r_t conditioned on $\mathcal{F}_{t-1}$.

For the same reason as in the case of VaR, we aggregate the results across all fifty stocks. We calculate the average cumulative violations for all fifty stocks and average the absolute errors as follows

$$\mathrm{CVAE}_p = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{\mathrm{CV}^i}(p) - p/2\right|, \tag{1.5.9}$$

where $\widehat{\mathrm{CV}^i}$ is average cumulative violation for stock i.

1.5.2 Quality of Fit

We assess the quality of fit by forecasting one-period-ahead out-of-sample. Using out-of-sample forecasts is crucial, as the non-parametric diurnal adjustment has a larger potential for over-fitting than parametric models. Since the DA-EGARCH is computationally demanding, we perform a moving window exercise with a two-year long estimation window, with annual re-estimation. The first window starts on November 29, 2015 and ends on November 29, 2017. The EGARCH parameters and bandwidths of DA-EGARCH and MEGARCH-t are

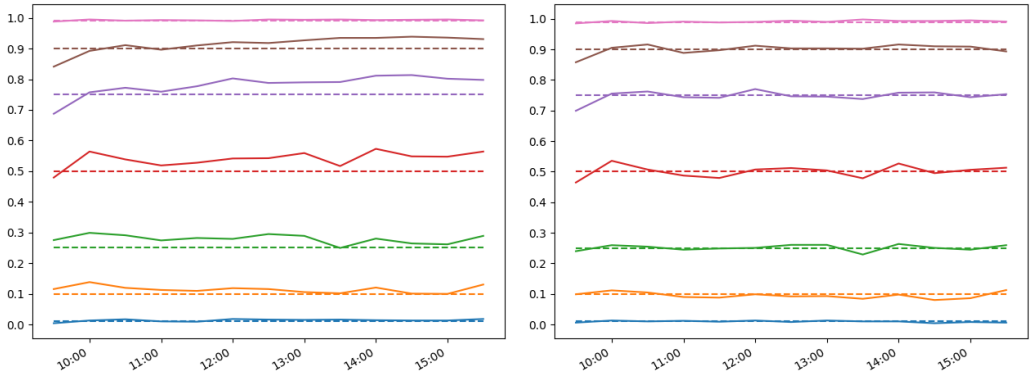
re-estimated annually. The first estimation window ends on November 29, 2017, and out-of-sample one-period ahead forecasts are made for the observations from November 30, 2017, till November 29, 2018. Then, the model parameters and bandwidths are re-estimated using the second two-year window, and out-of-sample one-period ahead forecasts for the rest of the dataset are computed. The out-of-sample one-period ahead forecasts for each time t are calculated with updated diurnal adjustments and filtered conditional variances using the most recent two years of observations available in $\mathcal{F}_{t-1}$. This procedure ensures that no information at t or later is used for distribution forecasting at time t.

We calculate the sample analogues of the hit rates for DA-EGARCH and MEGARCH-t for the Apple Inc. stock for individual intraday periods to assess the general quality of fit and explore potential diurnal patterns in the prediction errors. The results for the two models are shown in Figure 1.6. The prediction errors of MEGARCH-t are visibly larger than the prediction errors of DA-EGARCH. In the case of MEGARCH-t, the hit rates above 50% increase over time, while the hit rates below 50% decrease. This is in line with the diurnal patterns shown in the exploratory analysis in Figure 1.3, where the center of the distribution seems to have lower density in the morning than later in the day. The pattern in prediction errors demonstrates that MEGARCH-t fails to capture the diurnal pattern in the distribution. With the DA-EGARCH, we do not observe such a pattern in predictions, except for the first thirty minutes, where 50% and higher hit rates are missing their expected values. It seems that both MEGARCH-t and DA-EGARCH fail to fit the possible asymmetry of the distribution that is unique to the opening minutes. In the case of DA-EGARCH, the fit could be improved by tweaking the weights of kernel estimation for the opening hours.

The HRAE metrics based on all fifty stocks grouped by intraday period are shown in Table 1.4. With the exception of tails represented by the 1% and 99% percentiles, the DA-EGARCH provides a better distributional fit than MEGARCH-t. We observe the same phenomena as for AAPL: the prediction errors are the largest during the opening periods. The results demonstrate potential limitations of using non-parametric methods for the diurnal distributional adjustment. Generally, non-parametric methods are more demanding in terms of

Figure 1.6 Comparison of distributional fit of MEGARCH-t and DA-EGARCH.

(a) MEGARCH-t **(b)** DA-EGARCH



Results for AAPL stock using 1%, 10%, 25%, 50%, 75%, 90%, and 99% percentiles throughout the day. The dashed lines represent the expected hit rates, and the solid lines are the average hit rates. Prediction period: from November 30, 2017 to November 8, 2019. Models are annually re-estimated with a two-year estimation window.

sample size than parametric methods. In the extreme tails, by definition, only a few observations are available. Therefore, non-parametric adjustment may not be precise enough to achieve significant improvement. However, DA-EGARCH still outperforms MEGARCH-t for most intraday periods.

The results for CVAE_p are in Table 1.5. DA-EGARCH strongly outperforms MEGARCH-t in the central part of the distribution. The performance of DA-EGARCH for 1% or 10% percentiles is better for most of the period.

Note that cumulative violations are focused on the left side of the distribution. The cumulative violation errors for the 99% percentile are impacted not only by the tail, but also by the center of the distribution. The error in the tail is of a smaller scale than in the center of the distribution. Therefore, the fact that DA-EGARCH strongly outperforms MEGARCH-t for the 99% percentile does not mean that it provides strongly better predictions of the right tail.

1.5.3 Practical Backtesting

The main practical application of the proposed model is ES forecasting of the left tail of the distribution. It is common practice to regularly re-estimate models to keep parameters updated with the most recent events. To study the

Table 1.4 Comparison of $\text{HRAE}_p \times 100$ for MEGARCH-t and DA-EGARCH.

EGARCH Model*	M-t	DA	M-t	DA	M-t	DA	M-t	DA	M-t	DA	M-t	DA	M-t	DA
Percentile ($p \times 100$)	1.00%		10.00%		25.00%		50.00%		75.00%		90.00%		99.00%	
09:30	0.390	**0.269**	3.287	**0.980**	3.702	**1.782**	2.768	**1.588**	3.337	**1.831**	3.448	**1.718**	0.471	**0.314**
10:00	0.287	**0.268**	1.942	**0.499**	2.314	**1.006**	2.114	**1.151**	1.703	**1.144**	1.141	**0.645**	**0.255**	0.272
10:30	0.254	**0.164**	1.481	**0.561**	1.877	**0.763**	2.515	**1.220**	2.212	**1.042**	1.065	**0.619**	0.251	**0.212**
11:00	0.349	**0.268**	0.868	**0.794**	1.561	**1.036**	2.659	**1.463**	2.153	**1.026**	1.210	**0.595**	0.304	**0.220**
11:30	0.263	**0.220**	0.934	**0.714**	1.607	**1.019**	1.933	**1.243**	1.532	**0.664**	0.777	**0.530**	0.269	**0.208**
12:00	**0.230**	0.230	1.016	**0.623**	1.571	**0.750**	2.465	**0.989**	1.706	**0.598**	1.036	**0.638**	0.257	**0.222**
12:30	0.326	**0.261**	0.915	**0.815**	2.147	**0.858**	2.442	**1.045**	1.864	**0.986**	0.988	**0.624**	0.290	**0.237**
13:00	0.336	**0.207**	0.647	**0.581**	1.542	**1.285**	2.814	**1.361**	2.360	**0.871**	1.428	**0.650**	**0.272**	0.272
13:30	**0.257**	0.260	1.195	**0.676**	2.679	**1.055**	2.932	**1.315**	2.528	**0.980**	1.575	**0.950**	**0.246**	0.381
14:00	**0.205**	0.237	1.184	**0.623**	1.989	**0.746**	2.567	**1.144**	2.597	**0.671**	1.451	**0.700**	0.293	**0.261**
14:30	**0.239**	0.317	1.051	**0.621**	2.341	**0.891**	2.985	**1.091**	3.195	**1.039**	1.976	**1.028**	**0.255**	0.350
15:00	**0.216**	0.314	1.619	**1.009**	2.554	**0.980**	2.738	**1.358**	1.910	**1.193**	1.358	**0.742**	**0.283**	0.467
15:30	0.302	**0.233**	1.419	**1.173**	2.015	**1.838**	2.886	**1.686**	2.249	**1.423**	1.466	**0.788**	**0.298**	0.332

$\text{HRAE}_p = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{h}^i(p) - p\right|$, Number of Stocks (n): 50

*M-t and DA represent MEGARCH-t and DA-EGARCH, respectively.

Bold numbers signify the model with lower HRAE_p. Prediction period: from November 30, 2017 to November 8, 2019. Models are annually re-estimated with a two-year estimation window.

Table 1.5 Comparison of $\text{CVAE}_p \times 100$ for MEGARCH-t and DA-EGARCH.

EGARCH Model*	M-t	DA	M-t	DA	M-t	DA	M-t	DA	M-t	DA	M-t	DA	M-t	DA
Percentile ($p \times 100$)	1.00%		10.00%		25.00%		50.00%		75.00%		90.00%		99.00%	
Expected CV%	0.5%		5.00%		12.50%		25.00%		37.50%		45.00%		49.50%	
09:30	0.198	**0.140**	1.644	**0.464**	2.881	**0.962**	2.799	**1.171**	2.191	**1.172**	1.991	**1.075**	1.895	**1.005**
10:00	0.185	**0.137**	1.053	**0.267**	1.753	**0.428**	1.930	**0.643**	1.743	**0.774**	1.560	**0.760**	1.436	**0.713**
10:30	0.155	**0.093**	0.903	**0.307**	1.377	**0.428**	1.591	**0.524**	1.714	**0.586**	1.654	**0.575**	1.544	**0.549**
11:00	0.219	**0.151**	**0.467**	0.495	**0.789**	0.833	1.441	**1.060**	1.778	**1.067**	1.691	**0.949**	1.568	**0.846**
11:30	0.127	**0.107**	0.543	**0.418**	0.866	**0.549**	1.189	**0.777**	1.330	**0.773**	1.232	**0.692**	1.140	**0.644**
12:00	0.184	**0.116**	0.530	**0.449**	0.975	**0.676**	1.411	**0.588**	1.569	**0.388**	1.487	**0.361**	1.381	**0.346**
12:30	0.187	**0.164**	**0.442**	0.675	1.032	**0.717**	1.511	**0.675**	1.553	**0.629**	1.413	**0.542**	1.292	**0.490**
13:00	0.211	**0.135**	0.424	**0.327**	0.732	**0.403**	1.287	**0.720**	1.544	**0.844**	1.509	**0.822**	1.415	**0.795**
13:30	**0.164**	0.169	**0.566**	0.607	1.353	**0.785**	2.021	**0.997**	1.902	**0.920**	1.801	**0.835**	1.680	**0.767**
14:00	0.145	**0.128**	0.602	**0.540**	1.250	**0.517**	1.574	**0.494**	1.589	**0.425**	1.529	**0.425**	1.422	**0.406**
14:30	**0.134**	0.173	**0.576**	0.678	1.296	**0.708**	1.769	**0.682**	1.728	**0.679**	1.677	**0.611**	1.567	**0.590**
15:00	**0.155**	0.202	**0.816**	0.823	1.607	**1.030**	2.096	**1.154**	2.038	**1.024**	1.883	**0.884**	1.734	**0.779**
15:30	0.152	**0.126**	0.756	**0.414**	1.345	**0.979**	1.824	**1.221**	2.056	**1.176**	1.992	**1.077**	1.872	**1.014**

$\text{CVAE}_p = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{\text{CV}^i}(p) - p/2\right|$, Number of Stocks (n): 50

*M-t and DA represent MEGARCH-t and DA-EGARCH, respectively.

Bold numbers signify the model with lower CVAE_p. Prediction period: from November 30, 2017 to November 8, 2019. Models are annually re-estimated with a two-year estimation window.

practical performance of the model, we perform a moving window exercise with more frequent re-estimation. We estimate the model using a window of fixed length with 1,950 observations, which corresponds to 150 business days, and make one-period-ahead forecasts. We repeat this for every window from the start to the end of our dataset, i.e., we estimate 10,356 models and forecast the one-period-ahead conditional distribution.

Unfortunately, the DA-EGARCH is computationally too demanding to re-estimate the model 10,356 times. Therefore, we benchmark a distributionally adjusted GARCH(1,1) model against multiplicative GARCH(1,1)-t, respectively hereafter referred to as DA-GARCH and MGARCH-t. The DA-EGARCH model presented in the previous section can be estimated in tens of seconds, even with a good initial guess for parameters, which is too demanding for 10,356 re-estimations. For more details on computational complexity, see Appendix A.3. To process the exercise in a reasonable amount of time, we switched to a symmetric GARCH(1,1) specification that has only two parameters, and, on top of that, the parameters are more restricted in comparison to the EGARCH specification. This allowed us to pre-compute the marginal cdf, quantile function, and derivative of the quantile function of the normal GARCH process on a grid for a reasonable region of parameters. Additionally, we limit ourselves to only four arbitrary stocks from our dataset, AAPL, JPM, MCD, and MSFT, to keep the processing time feasible.

The DA-GARCH model is specified as

$$r_{d,\tau} = F^{-1}(p_t, \tau), \tag{1.5.10}$$

$$p_t = G(x_t), \tag{1.5.11}$$

$$x_t = \sigma_t \varepsilon_t,\ \varepsilon_t \sim \text{i.i.d. } N(0,1), \tag{1.5.12}$$

$$\sigma_t^2 = (1 - \alpha_1 - \alpha_2) + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \varepsilon_{t-1}^2 \sigma_{t-1}^2, \tag{1.5.13}$$

where $F^{-1}(r, \tau)$ is marginal cdf of $r_{d,\tau}$ at intraday period τ, and $G(x)$ is marginal cdf of $x_t = \sigma_t \varepsilon_t$. The marginal distribution of $r_{d,\tau}$ is estimated as for the DA-EGARCH specification, but we update the bandwidth, using the cross-validation maximizing likelihood, only once every two days. Note that the DA-GARCH specification has only two parameters, $\alpha_1, \alpha_2 \geq 0$ with the restriction that $\alpha_1 + \alpha_2 < 1$. Therefore, we can specify a fine grid for α_1 and α_2, and pre-calculate $G^{-1}(p)$ and $dG^{-1}(p)/dp$

for all reasonable values of parameters.

The competing MGARCH-t model is specified as

$$r_{d,\tau} = S_\tau \mu + S_\tau \sigma_t \varepsilon_t, \tag{1.5.14}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \varepsilon_{t-1}^2 \sigma_{t-1}^2, \tag{1.5.15}$$

where $\varepsilon_t \sim t(v)$, with $t(v)$ being Student's t distribution with v degrees of freedom.

In Table 1.6, we report the aggregated performance for all intraday periods and the left tail of the distribution, as practitioners are likely interested in the overall performance in the left tail. Cumulative violations corresponding to quantile p should, for a correct model, on average be equal to $p/2$ and independent. Empirical values above $p/2$ suggest that the model systematically underestimates mass in the tail below p. There is a clear pattern in the table: MGARCH-t with diurnal factors allocates significantly less mass in the tail than the DA-GARCH model. Overall, the DA-GARCH has average cumulative violations substantially closer to the desired values. However, there is still room for improvement, since the observed values are frequently at least two standard errors away from the expected values $p/2$.

The DA-GARCH performs surprisingly well in the moving-window exercise compared to the CVAE calculated with annual model re-estimation. The model assumes fixed parameters over time. However, by running the moving window, we practically allow for slow variation of parameters and marginal distribution in time. Further exploration and formalization of the framework with time-varying marginal distributions is beyond the scope of this work. However, extending our framework with a non-stationary, smoothly trending marginal distribution could potentially achieve further improvement. Feng and McNeil (2008) proposed such an idea for a slowly trending non-stationary scale of asset returns and reported positive empirical results. Our results also align with the work of Andersen et al. (2024), who recently developed a new method for estimating time-varying diurnal patterns in volatility. The authors apply their method to the S&P 500 index and document a slow shift in volatility diurnal patterns over time.

Table 1.6 Out-of-sample average cumulative violations (CV).

Window length			1,800 observations	
No. of windows			10,356 observations	
Percentile ($p \times 100$)	1.00%	3.00%	5.00%	10.0%
Expected $CV\%$	0.50%	1.50%	2.50%	5.00%
Standard Errors $\times 100$	0.06	0.10	0.13	0.17
AAPL				
MGARCH-t	0.68%	1.81%	2.83%	5.28%
DA-GARCH	0.60%	1.57%	2.48%	4.74%
JPM				
MGARCH-t	0.64%	1.76%	2.89%	5.44%
DA-GARCH	0.59%	1.53%	2.56%	4.94%
MCD				
MGARCH-t	0.64%	1.74%	2.80%	5.24%
DA-GARCH	0.59%	1.51%	2.47%	4.77%
MSFT				
MGARCH-t	0.71%	1.87%	2.86 %	5.29%
DA-GARCH	0.63%	1.63%	2.51%	4.75%

One-period-ahead out-of-sample average cumulative violations estimated using 150-day moving windows. The columns represent 1%, 3%, 5%, and 10% quantile-based average cumulative violations, respectively. All results are reported in percentages ($\times 100$) for better readability. A perfect model should have average cumulative violations equal to the expected cumulative violations (CV%).

1.6 Conclusion

We propose a new semi-parametric modeling framework for intraday returns, in which diurnal distributional patterns are captured by PIT using marginal distributions. We estimate the marginal distributions using a kernel density estimator. The intraday returns are then diurnally adjusted in distributions by the estimated cumulative distribution functions, removing any distributional diurnal patterns. The adjusted series are then modeled by a new class of processes that are consistent with the properties of the adjusted series and constructed by PIT from the conventional models for modeling returns, which is important for utilizing stylized facts and comparing our work to others.

The key difference between our approach and conventional methods for diurnal adjustment is the ability to capture diurnal patterns beyond mean and volatility. Our analysis shows a clear diurnal pattern in the entire distribution. We apply the distributional adjustment to GARCH and EGARCH models, and benchmark the resulting models against GARCH and EGARCH with multiplicative adjustment. Out-of-sample forecasting of VaR and ES demonstrates improved precision and the practical importance of capturing distributional diurnal patterns.

Our research can continue in multiple directions. First, in this work, we avoid studying the asymptotic properties of the framework. The knowledge of relevant asymptotic distributions is not important for forecasting, but the framework may allow testing economic hypotheses that would not be possible without distributional adjustment. Therefore, there may be a motivation to extend the asymptotic theory developed by Newey (1994) for semi-parametric methods and independent data to time series data. Second, the proposed framework freely captures any marginal distribution of returns, while conventional models impose implicit assumptions about marginal distribution. Therefore, models in daily setups may potentially benefit from our idea as well, even though there is no need for diurnal adjustment. Finally, the third direction is extending our framework with slowly trending non-stationary distributions, i.e., expanding the idea of Andersen et al. (2024) to the entire distribution. Allowing for slowly trending non-stationary distributions and distributional diurnal patterns could potentially further improve the forecasts.

2 Intraday Volatility Forecasting with Lassoing Clustered Commonality

2.1 Introduction

Intraday volatility forecasting is an important aspect of risk management and financial derivative pricing and, consequently, an essential task for market makers, option traders, and brokers. The importance of forecasting intraday volatility for practitioners is demonstrated by the fact that financial institutions routinely provide intraday forecasting services. For example, in late 2015 Deutsche Börse launched intraday volatility forecasting for the DAX index, the Euro Stoxx 50 index, and the Euro-Bund[1] as part of their Real-Time Analytics services.

Naturally, intraday volatility forecasting has attracted significant attention from researchers over the years. Conditional volatility or variance for each individual time period is not directly observable. The methods for obtaining conditional volatility can be roughly categorized into two groups: model-based and model-free.

Model-based measures use time series models to filter out unobserved series of conditional volatility, pioneered by Engle (1982) in daily setup. The ideas were later extended to an intraday setup by Bollerslev and Ghysels (1996) and extensively empirically studied by Andersen and Bollerslev (1997a). The intraday literature is similar to the daily literature, but differs in a typical intraday periodicity - diurnal U-shaped patterns, where volatility is typically higher in the morning, lower around noon, and again higher before the market closes. For more examples on intraday volatility modeling, see Corsi (2009) or Engle and Sokalska (2012).

The model-free measures, frequently referred to as realized measures, rely on high-frequency data and asymptotic properties of asset returns as sampling frequency grows. Merton (1980) was the first to notice that the integrated variance

[1]Deutsche Börse. "Intraday Volatility Forecast For Dax, Euro Stoxx 50 and Euro-Bund." *Risk.net*, www.risk.net/media/download/934501. Accessed 5 April 2025.

of the time-continuous process can be measured with arbitrary precision as the sampling frequency increases. However, it was only Andersen and Bollerslev (1998), who first studied realized volatility, the approximation for integrated volatility, and advocated it as a good measure. Andersen et al. (2003) formalized links between realized volatility and conditional variance. For a review, see McAleer and Medeiros (2008).

Although the realized measures are model-free, a model for forecasting is needed. This is a main advantage of model-free measures, as there are countless possibilities for forecasting the realized measure by any time series forecasting method or models without the need to filter unobserved volatility. Andersen et al. (2003) examined several autoregressive GARCH-like specifications. Over the last two decades, more models have been developed, for example RealGARCH by P. R. Hansen et al. (2012). Another exceptionally popular model is the heterogeneous auto-regressive (HAR) model by Corsi (2009). Many researchers have also applied various machine learning methods, for example Christensen et al. (2023) compare the HAR model, random forest, lasso, and neural networks. For a review of applications of machine learning methods for predicting realized volatility, see Gunnarsson et al. (2024).

Our work contributes to the literature on forecasting intraday realized volatility and builds on the idea of commonality, which was pioneered by Chordia et al. (2000), and generally describes how groups of assets share common determinants manifesting as comovements in returns, liquidity, or volatility. Karolyi et al. (2012) study how commonality varies across countries and Dang et al. (2015) expanded the research by studying the role of commonality in the impacts of firm-level news across different countries. Bollerslev et al. (2018) significantly improved the out-of-sample predictions of the HAR model by exploiting commonality within groups of assets such as equities, fixed incomes, foreign exchanges, etc., and universal commonality with the market as a whole. In our work, we try to use commonality across groups of assets to improve volatility forecasts.

Our research idea expands on the work of Zhang et al. (2024), who study the performance of various statistical and machine learning models to forecast intraday volatility. They try to exploit the commonality in volatility across equities to

improve the forecasting of individual equities. They construct a simple market-level volatility and use it as a predictor. We take the idea of equity commonality in volatility one step further. We ask the following research question: is there a more granular structure than the entire equity market that can be used for further improvement of intraday volatility forecasts? For example, could we break down the market into sectors and use signals only from relevant sectors for each stock? Also, if there is such a commonality, what is the optimal grouping for intraday volatility forecasting – is it sector, industry, or data-driven clustering of stocks? We break the equity market down into groups and try to utilize potential commonality across the groups.

To answer the questions, we analyze a dataset of 363 equities from S&P 500, covering the period from October 29, 2015, to November 8, 2019. We consider several possible groupings, involving sectors and industries based on the Global Industry Classification Standard (GICS) defined by MSCI and Standard & Poor's.[2] The GICS is based on firm business characteristics, but may not represent an optimal grouping structure from a commonality perspective. Therefore, we also create a data-driven grouping using the k-shape clustering algorithm. Finally, for comparison with the results of Zhang et al. (2024), we also consider the entire market as a possible single group.

We calculate representative volatility for each group and use these volatilities as predictors for a HAR-like model. As some groupings involve a large number of categories, simple OLS easily leads to overfitting and extremely poor out-of-sample performance. In addition, for a specific stock, the signals from some groups may be more important than those from other groups. Therefore, we use the least absolute shrinkage and selection operator (lasso) to let the data decide which groups are important for each stock and which are not. To be more specific, we use the original lasso by Tibshirani (1996), the adaptive lasso by Zou (2006), and the clustering lasso by Bühlmann et al. (2013).

We perform an out-of-sample evaluation of intraday realized volatility forecasts (realized log-variance, to be more precise) and benchmark different groupings and

[2]MSCI. "The Global Industry Classification Standard (GICS)". www.msci.com/our-solutions/indexes/gics. Accessed 1 November 2024.

lasso variations. Our empirical results demonstrate that breaking the commonality of the market level to higher granularity improves the performance of the forecast.

We also study two alternative lag structures. The HAR model uses daily, weekly, and monthly realized volatilities that approximate a long memory of realized volatility. The natural question is whether the commonality in volatility manifests itself only through an immediate intraday reaction or whether there is a possible impact from lower frequencies. We calculate daily, weekly, and monthly volatilities for the group representing realized volatilities. Including lower-frequency commonality does not generally improve forecasting precision, which suggests that there is no or a very weak low-frequency commonality relevant to high-frequency volatility and commonality.

Another research problem we tackle is the lifetime of relevant signals for forecasting. Relevance and commonality with specific clusters or groups may vary over time. If that is indeed the case, a frequently updated model with a short estimation window may potentially capture such a time-varying commonality. Chinco et al. (2019) successfully applied lasso to predict individual stock returns with a lagged cross-section of returns for all available traded stocks as predictors. They use a short moving window for 1-minute forward predictions and lasso to identify potential short-lived signals. A trading strategy based on return predictions of their lasso approach resulted in profit outperforming market return and other trading strategies based on competing forecasting models. Inspired by this work and the idea of detecting short-lived signals, we rerun our exercise with 15 and 60 day-long moving windows. Unfortunately, our results show inferior results compared to annually updated models with a long estimation window. Our findings do not, however, contradict the results of Chinco et al. (2019) as we work with lower frequencies because of a limitation arising from the inability to reliably calculate the realized volatility at higher frequencies due to microstructure noise. The phenomena observed at 1-minute frequency returns may simply not translate to our lower-frequency 30-minute realized volatilities.

Our results can be explained by several theoretical mechanisms described in the literature on commonality in liquidity. To shed more light on the role of commonality in forecast accuracy improvement, we propose a commonality

measure within and between sectors to relate forecast improvements to market commonality. Empirical results show that the commonality within sectors varies between sectors, and sectors with higher intrasector commonality are more predictable in terms of the lower MSE. The empirical results align with the liquidity spirals of Brunnermeier and Pedersen (2009) as a possible mechanism behind the commonality.

This paper is organized into eight sections. The next section introduces realized log-variance. Section 2.3 describes the methods used, including the HAR model, lasso, adaptive lasso, and clustering lasso. Section 2.4 introduces metrics for measuring empirical forecasting performance. Section 2.5 describes data, sectors, and industries. Section 2.6 introduces the model specifications and presents the out-of-sample performance results. Next, section 2.7 relates the forecasting results to the literature on commonality. Finally, section 2.8 concludes.

2.2 Volatility Measures

Let p_t be the log-price, where t is continuous time, measured in minutes. Andersen and Bollerslev (1998), Andersen et al. (2001), and Barndorff-Nielsen and Shephard (2002) pioneered the idea of realized volatility, formalized the theoretical framework, and studied the distributional properties of realized volatility, respectively. The key assumption is that log-price follows a diffusion process

$$p_t = \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dB_s, \tag{2.2.1}$$

where μ_s and σ_s are instantaneous spot drift and spot volatility, respectively, and B_s is a standard Brownian motion. Consider the partition $0 = t_0 \leq t_1 \leq t_2 \leq \ldots \leq t_n = t$ of the time interval $[0, t]$ and define the log-return $r_{t_\iota} = p_{t_\iota} - p_{t_{\iota-1}}$ for $\iota = 1, 2, ..., n$. As the norm of the partition $\max_\iota(t_\iota - t_{\iota-1}) \xrightarrow{n\to\infty} 0$, the sum of squared returns converges in probability to the integrated variance

$$\sum_{\iota=1}^{n} r_{t_\iota}^2 \xrightarrow{\text{in probability as } n\to\infty} \int_0^t \sigma_s^2 \, ds. \tag{2.2.2}$$

In other words, the integrated variance can be measured with an arbitrary precision as the sampling frequency grows. This is a theoretical result with several practical limitations to achieving arbitrary precision.

Real-world asset returns are not observed in continuous time and include microstructure noise. Microstructure noise is a non-systematic random deviation of the observed price from the true asset price. In the presence of microstructure noise, the realized variance is inconsistent for the integrated variance. P. R. Hansen and Lunde (2006), Andersen et al. (2011) and a few others studied the impact of the microstructure noise and proposed estimators robust to the noise. However, the robust estimators typically require choosing some sampling frequency parameters, and selecting them is typically rather an art than exact science. We take a pragmatic approach to the microstructure noise by approximating the integrated volatility with lower-frequency returns, where the impact of microstructure noise is negligible. Zhang et al. (2024) use 1-minute frequency returns, but their dataset includes only the top 100 stocks of the S&P 500 index, while we also include less liquid stocks from the index. We settle on a 3-minute sampling frequency. The potential presence of microstructure noise is discussed and tested in Appendix B.3.

The sum of squared log-returns for the fixed-frequency, finite-sample analog of limit (2.2.2) is realized variance. The square root of the realized variance is realized volatility. Instead of realized variance or volatility, we model realized log-variance[3] defined over period $(t-h, t]$ as

$$RV_t^{(h)} = \log\left(\frac{1}{h}\sum_{s\in(t-h,t]} r_s^2\right). \tag{2.2.3}$$

We use 3-minute returns and predict realized log-variance every 30 minutes; thus, $t \in \{30, 60, 90, ..., T\}$, and $s \in \{3, 6, 9, ..., T\}$ for the rest of our work. Normally, the realized variance is not standardized by the length h of the window $(t-h, t]$. We standardize by the duration h to make realized log-variances comparable across various window lengths, e.g., when comparing realized log-variance calculated over the last 30 minutes and log-variance calculated over the last day. This normalization makes exploratory analysis easier and has no impact upon forecasting performance in our setup.

[3]In the relevant literature, $RV_t^{(h)}$ frequently denotes realized variance or volatility. We use it to denote the realized log-variance to make the notation lighter and aligned with the work of Zhang et al. (2024).

Using realized log-variances instead of realized variances brings several advantages. The logarithm mitigates the impact of large extreme returns. As Andersen et al. (2003) report, the use of realized log-variances may improve forecasting performance of linear models. Using linear models with log-variances is similar to using multiplicative specification for GARCH-like models, e.g., E-GARCH by Nelson (1991).

Another consideration in using realized variance as a proxy for integrated variance is discontinuous jumps. The log-price process may exhibit, in addition to continuous price changes following (2.2.1), discontinuous jumps. In such cases, the sum (2.2.2) converges to the sum of integrated volatility and the sum of squared jumps. Barndorff-Nielsen and Shephard (2004) proposed a decomposition based on realized volatility and the bi-power variation that allows the realized volatility to be decomposed into integrated volatility and the sum of squared jumps. In our work, we focus on volatility without distinguishing whether it comes from a continuous diffusion component or a discontinuous jump component. The decomposition could potentially be useful if applied to decompose past realized volatilities that are used as predictors. For example, Andersen et al. (2007) or Bu et al. (2023) report somewhat improved forecast precision. As the decomposition requires choices of one or more hyper-parameters for which there is no broadly agreed standard, it is beyond the scope of our work.

2.2.1 Time and Cross-sectional Aggregation

We aim to forecast 30-minute realized log-variances. All of the models considered are built on the basic idea of the HAR model (Corsi, 2009), which uses realized volatilities or log-variances of various lengths to account for a potential long-memory of realized log-variance. The HAR model uses daily, weekly and monthly volatilities or log-variances as predictors. We define the realized log-variances with four different window lengths:

- Intraday 30-minute realized log-variance

$$RV_t^{\mathrm{I}} = RV_t^{(30)}. \tag{2.2.4}$$

- Daily realized log-variance

$$RV_t^{\mathrm{D}} = RV_t^{(390)}. \tag{2.2.5}$$

- Weekly realized log-variance

$$RV_t^{\mathrm{W}} = RV_t^{(1950)}. \tag{2.2.6}$$

- Monthly realized log-variance

$$RV_t^{\mathrm{M}} = RV_t^{(8580)}. \tag{2.2.7}$$

We assume that trading is done only during the main trading hours, 9:30–16:00, with a day count convention of 5 and 22 business days within a week and month, respectively. Consequently, the series of daily, weekly, and monthly log-variances are given by series of overlapping moving windows of 390, 1950, and 8580 minutes duration, respectively. We aim to forecast realized log-variance over 30-minute intervals. Therefore, we have chosen $h = 30$ for an intraday log-variance.

Forecasting over intervals different from 30 minutes is possible. However, one has to make sure that realized log-variances are not biased by microstructure noise or poor approximation because of the low number of returns within the interval. Another potential consideration is the number of non-overlapping intervals within one day being a whole number. Zhang et al. (2024) experiment with intervals of 10, 30, and 65 minutes.

The clustering and grouping models applied in our work use cross-sectionally averaged realized log-variances by data-driven clusters or groups based on business classification to sectors, industries, or market as predictors. Henceforth, we will use the terms group and cluster interchangeably.

As we work with multiple stocks we have to add an additional index indicating stock $i \in \{1, 2, ..., m\}$. For each stock i, we calculate the realized log-variances $RV_{i,t}^{\mathrm{I}}$, $RV_{i,t}^{\mathrm{D}}$, $RV_{i,t}^{\mathrm{W}}$, and $RV_{i,t}^{\mathrm{M}}$ defined above.

Consider partitioning the set of stocks into k groups G_1, G_2, ..., G_k, where G_c is a set of stock indices belonging to the cth group. Our goal is to calculate group averages and use them as predictors. However, the past of stock i is potentially more important for predicting realized log-variances of stock i than the

past of other stocks. Therefore, we always include stocks' own past directly. To avoid including the same information twice in the models, we do not include the modeled stock i in the group averaging. In other words, for each stock i, we create a partitioning G_0^i, G_1^i, G_2^i, ..., G_k^i, where $G_0^i = \{i\}$ and $G_c^i = G_c \setminus \{i\}$ for $c \geq 1$.

Inspired by Zhang et al. (2024), who aggregates the entire market by a simple average, we aggregate the individual realized log-variances by a simple average across stocks within a group as

$$\begin{aligned} RV_{i,t}^{\mathrm{I,c}} &= \frac{1}{|G_c^i|} \sum_{j \in G_c^i} RV_{j,t}^{I}, \\ RV_{i,t}^{\mathrm{D,c}} &= \frac{1}{|G_c^i|} \sum_{j \in G_c^i} RV_{j,t}^{D}, \\ RV_{i,t}^{\mathrm{W,c}} &= \frac{1}{|G_c^i|} \sum_{j \in G_c^i} RV_{j,t}^{W}, \\ RV_{i,t}^{\mathrm{M,c}} &= \frac{1}{|G_c^i|} \sum_{j \in G_c^i} RV_{j,t}^{M}, \end{aligned} \tag{2.2.8}$$

where $|\cdot|$ is cardinality.

2.2.2 Diurnal Adjustment

Intraday asset returns exhibit diurnal patterns similar to seasonality in low-frequency data.[4] Throughout the day, variance typically exhibits a U-shaped diurnal pattern with high variance in the morning, low variance around noon, and again increased variance before the market closes. The diurnal patterns of volatility can be part of the model, e.g. periodic GARCH Bollerslev and Ghysels (1996), or the data may be adjusted for diurnal patterns first and then the model can be applied to the adjusted data, e.g., Engle and Sokalska (2012). All candidate models are linear. To keep the modeling of diurnal patterns consistent across models without being affected by the shrinkage of the applied methods, we first adjust all intraday realized log-variances.

It is convenient to introduce a new time indexing for days and intraday time within a day. Consider continuous time $t \in (0, T]$ that covers D whole days of length T/D. The time can be partitioned by days as $(0, t_1]$, $(t_1, t_2]$, ..., $(t_{D-1}, t_D]$. Let d denote the day covering interval $(t_{d-1}, t_d]$ and $\tau \in (0, t_d - t_{d-1}]$ time within

[4]For an empirical study on diurnal patterns in variance, see Andersen and Bollerslev (1997a).

the day. This gives a one-to-one mapping between t and d, τ given by

$$t(d,\tau) = (d-1)\frac{T}{D} + \tau, \tag{2.2.9}$$

$$d(t) = \left\lceil t/\frac{T}{D} \right\rceil, \quad \tau(t) = t - (d(t)-1)\frac{T}{D}, \tag{2.2.10}$$

where $\lceil \cdot \rceil$ is the ceiling function. Thus, $RV^I_{i,t} = RV^I_{i,d(t),\tau(t)}$, or to keep it simple, just $RV^I_{i,d,\tau}$.

We estimate diurnal factors for intraday time τ as

$$S_{i,\tau} = \frac{1}{D}\sum_{d=1}^{D}\left(RV^I_{i,d,\tau}\right) \tag{2.2.11}$$

and adjusted series as

$$RV^{I*}_{i,t} = RV^I_{i,d,\tau} - S_{i,\tau}. \tag{2.2.12}$$

Note that, as RV is log-variance, the additive adjustment is similar to multiplicative adjustment of realized variance. Using an identical procedure, we also adjust the intraday group averages $RV^{\mathrm{I,c}}_t$ from (2.2.8).

2.3 Models

This section describes the models applied in our empirical study, and k-shape clustering algorithm. All of the models considered forecast intraday realized log-variances $RV^I_{i,t}$ calculated over non-overlapping 30-minute windows. To avoid heavy notation for time indices, t will have increments of 30 minutes for the rest of our work.

2.3.1 HAR-D

Corsi (2009) proposed a heterogeneous autoregressive (HAR) model for predicting daily realized variance. The main benefits of the HAR model are its simplicity and ability to replicate long memory observed in the stock return volatility. Its key idea is to use the realized variances over a longer period of time as predictors to capture long memory. Specifically, daily, weekly, and monthly realized variances are used as predictors for predicting daily variance. Formally speaking, the resulting process is an AR process of order 22 (assuming the day count convention

of 22 trading days per month) with constrained estimation and, as such, the process does not exhibit long memory. However, the process empirically captures what could be considered long memory in a looser sense. It has been questioned whether the imposed lag structure is optimal. Audrino and Knaus (2016) used Least Absolute Shrinkage and Selection Operator (lasso) for feature selection to explore whether the HAR lag structure can be easily recreated by lasso applied to realized variances of various meaningful durations other than daily, weekly and monthly. The lasso does not recreate the HAR lag structure, nor does it outperform out-of-sample forecasting by the HAR model.[5]

Recently, Zhang et al. (2024) applied the HAR model in an intraday setup by adding average realized log-variance for each intraday time period as a predictor. As our data are already adjusted for diurnal patterns, we do not add the average for a given time. Instead, the following simple specification is used:

$$RV^{I*}_{i,t+30} = \beta^0_i + \beta^I_i RV^{\mathrm{I}*}_{i,t} + \beta^D_i RV^{\mathrm{D}}_{i,t} + \beta^{\mathrm{W}}_i RV^{\mathrm{W}}_{i,t} + \beta^M_i RV^{\mathrm{M}}_{i,t} + \varepsilon_{i,t+30}. \tag{2.3.1}$$

The model is estimated by ordinary least squares (OLS). Although the model differs slightly from that of Zhang et al. (2024), we will also address the HAR model with diurnal adjustment as HAR-D.

The HAR-D model can use external regressors. In our case, we use group averages as defined by (2.2.8). In general, all considered models are linear in the form

$$RV^{\mathrm{I}*}_{i,t+30} = \sum_{j=0}^{p} \beta_{i,j} X_{i,t,j} + \varepsilon_{i,t+30}, \tag{2.3.2}$$

where $X_{i,t,0}, X_{i,t,1}, X_{i,t,2}, ..., X_{i,t,p}$ are predictors for $RV^{\mathrm{I}*}_{i,t+30}$, with $X_{i,t,0} = 1$ representing the intercepts, and $\beta_{i,0}, \beta_{i,1}, ..., \beta_{i,p}$ being their corresponding coefficients. The predictors include:

- own predictors $RV^{\mathrm{I}*}_{i,t}$, $RV^{\mathrm{D}}_{i,t}$, $RV^{\mathrm{W}}_{i,t}$, and $RV^{\mathrm{M}}_{i,t}$,
- cluster predictors $RV^{\mathrm{I}*,\mathrm{c}}_{i,t}$, $RV^{\mathrm{D},\mathrm{c}}_{i,t}$, $RV^{\mathrm{W},\mathrm{c}}_{i,t}$, and $RV^{\mathrm{M},\mathrm{c}}_{i,t}$ for $c \in (1, k)$.

The full specification then has $p = 4+k\times4$ predictors. Note, the intraday predictors $RV^{\mathrm{I}*,\mathrm{c}}_{i,t}$ are group averages $RV^{\mathrm{I},\mathrm{c}}_{i,t}$ adjusted for diurnal patterns as described in

[5]For more empirical studies on the HAR lag structure, see Audrino et al. (2019) or Qiu et al. (2019).

section 2.2.2. Since OLS is prone to overfitting, anything more granular than market averages leads to very poor out-of-sample performance. However, this model serves as a baseline, and all lasso estimators, with the exception of cluster group lasso (CGL), do variable selection over an identical set of predictors.

2.3.2 Lasso

The Least Absolute Shrinkage and Selection Operator (lasso), developed by Tibshirani (1996), estimates a linear model with concurrent regularization and variable selection. As such, it potentially enhances prediction accuracy and overcomes the overfitting of OLS. The lasso was originally proposed for cross-sectional data. Later, Wang et al. (2007) established theoretical properties of the lasso in a time series setup.

The lasso estimates are obtained as

$$\widehat{\beta}_{\text{Lasso},i} = \arg\min_{b} \sum_{t\in\mathcal{T}} \left[RV^{\text{I*}}_{i,t+30} - \sum_{j=0}^{p} b_j X_{i,t,j} \right]^2 + \lambda_{\text{Lasso},i} \sum_{j=1}^{p} |b_j|, \tag{2.3.3}$$

where $b = (b_0, b_1, ..., b_p)^\intercal$, and $\lambda_{\text{Lasso},i}$ is the nonnegative regularization parameter, and $\mathcal{T} = \{30, 60, ..., T\}$ is the set of time indices in our dataset. The lasso objective function has two terms; the first is the sum of squared residuals identical to OLS, while the second is typically addressed as an l_1 penalty.

The l_1 penalty potentially shrinks the parameters of irrelevant predictors to 0; therefore, when the described predictors are used, the lasso is selecting not only relevant groups or clusters, but also the length of the relevant history.

The intercept b_0 is always selected and excluded from shrinkage. Since the effective penalty associated with predictor $X_{i,t,j}$ depends on the scale of the predictor, the predictors are standardized to apply an even penalty to all variables. This is done for all lasso-like methods in our work.

The hyperparameter $\lambda_{\text{Lasso},i}$ must be selected. We perform the selection by cross-validation and minimize the sum of squared errors for omitted observations. Following Tibshirani (1996), the cross-validation is performed by splitting the estimation sample into 5 non-overlapping time-ordered windows of equal size. More specifically, consider the set of all time indeces t in our data set $\mathcal{T} = \bigcup_{l=1}^{5} \mathcal{T}_l$, where $\mathcal{T}_l$ is the l-th window. $\mathcal{T}_1$ is the first window with the first 20% of observations, $\mathcal{T}_2$

represents the next 20% of observations, and so on. Denote the (2.3.3) estimator for given λ and omitted observations in $\mathcal{T}_l$ as

$$\widehat{\beta}^{-\mathcal{T}_l}_{\text{Lasso},i}(\lambda) = \arg\min_b \sum_{t\in\mathcal{T}\setminus\mathcal{T}_l} \left[RV^{\text{I*}}_{i,t+30} - \sum_{j=0}^{p} b_j X_{i,t,j} \right]^2 + \lambda \sum_{j=1}^{p} |b_j|, \tag{2.3.4}$$

and squared errors for omitted observations as

$$SSE(\mathcal{T}_l, \lambda) = \sum_{t\in\mathcal{T}_1} \left[RV^{\text{I*}}_{i,t+30} - \sum_{j=0}^{p} \widehat{\beta}^{-\mathcal{T}_l}_{\text{Lasso},i}(\lambda) X_{i,t,j} \right]^2. \tag{2.3.5}$$

The hyperparameter is then selected as

$$\widehat{\lambda}_{\text{Lasso},i} = \arg\min_\lambda \sum_{l=1}^{5} SSE(\mathcal{T}_l, \lambda). \tag{2.3.6}$$

2.3.3 Adaptive Lasso

Although the lasso turned out to be highly successful, Zou (2006) pointed out that it is not consistent in variable selection. The consistency, also named the oracle property, means the estimation procedure asymptotically identifies the right submodel with non-zero coefficients. Zou proposed an adaptive modification to achieve the oracle property. The adaptive lasso slightly modifies the penalty:

$$\widehat{\beta}_{\text{AdaLasso, i}} = \arg\min_b \sum_{t\in\mathcal{T}} \left[RV^{\text{I*}}_{i,t+30} - \sum_{j=0}^{p} b_j X_{i,t,j} \right]^2 + \lambda_{\text{AdaLasso,i}} \sum_{j=1}^{p} w_{i,j} |b_j|, \tag{2.3.7}$$

where $w_{i,j}$ are penalty weights of individual variables chosen to achieve the oracle property. Concretely, setting $w_{i,j} = 1/|\widehat{\beta}_{i,j}|^\gamma$, where $\widehat{\beta}_{i,j}$ is a root-T-consistent estimator of $\beta_{i,j}$, will result in the oracle property. Zou (2006) proposes standard OLS or ridge regression. Ridge regression addresses a potential problem with highly correlated predictors and high dimensionality. In our case, some predictors are strongly correlated, and OLS estimates are extremely noisy; thus, we have to use ridge regression.

The ridge regression is similar to the lasso given by (2.3.3), but instead of the l_1 penalty, the ridge regression uses the l_2 penalty:

$$\widehat{\beta}_{\text{Ridge},i} = \arg\min_b \sum_{t\in\mathcal{T}} \left[RV^{\text{I*}}_{i,t+30} - \sum_{j=0}^{p} b_j X_{i,t,j} \right]^2 + \lambda_{\text{Ridge},i} \sum_{j=1}^{p} b_j^2. \tag{2.3.8}$$

The hyperparameters $\lambda_{\text{AdaLasso,i}}$ and $\lambda_{\text{Ridge},i}$ are selected by the same procedure as $\lambda_{\text{Lasso},i}$.

The original adaptive lasso was intended for cross-sectional data. However, the work of Audrino and Camponovo (2015) extends the oracle properties of adaptive lasso to time series and lag selection.

The adaptive lasso requires an additional hyperparameter γ. We do not use cross-validation to select the parameter itself but, following the recommendation of Zou (2006), we simply estimate three different specifications with $\gamma \in \{0.5, 1, 2\}$.

2.3.4 Clustering Lasso

The clustering lasso is not a single method, but rather a variety of lasso methods that utilize potential clusters across individuals (stocks in our case) to improve prediction accuracy. The clustering lasso algorithms work in two steps. In the first step, a clustering algorithm is applied to cluster predictors together. In the second step, the lasso is applied utilizing cluster information. Bühlmann et al. (2013) proposed two variations of the clustering lasso, the cluster group lasso (CGL) and the cluster representative lasso (CRL). The CGL keeps original data and selects either all or no variables from each cluster, while the CRL first averages data by clusters and then selects averages.

Let us first introduce the CGL. Consider again the clustering $G_0^i, G_1^i, G_2^i, ..., G_k^i$ for stock i. Recall that $G_0^i = \{i\}$ and other clusters are created as $G_c^i = G_c \setminus \{i\}$. Let $\boldsymbol{\beta}_{i,j}$ for $j \geq 1$ is a vector of four parameters for realized log-variances $\mathbf{X}_{i,t,j} = (RV_{j,t}^{\mathrm{I}^*}, RV_{j,t}^{\mathrm{D}}, RV_{j,t}^{\mathrm{W}}, RV_{j,t}^{\mathrm{M}})^\intercal$ of stock j. Stacking vectors $\boldsymbol{\beta}_{i,j}$ results in a single vector of parameters $\beta_i = (\beta_{i,0}, \boldsymbol{\beta}_{i,1}^\intercal, \boldsymbol{\beta}_{i,2}^\intercal, ..., \boldsymbol{\beta}_{i,m}^\intercal)^\intercal$, where m is the number of stocks analyzed, and $\beta_{i,0}$ is the intercept (scalar).

As before, we exclude the intercept $\beta_{i,0}$ from variable selection and shrinkage. The CGL estimator is given by

$$\widehat{\beta}_{CGL,i} = \arg\min_b \sum_{t \in \mathcal{T}} \left[RV_{i,t+30}^{I*} - b_0 - \sum_{j=1}^{m} \mathbf{b}_j^\intercal \mathbf{X}_{i,t,j} \right]^2 + \lambda_{\mathrm{CGL},i} \sum_{c=0}^{k} w_c \left| \sum_{j \in G_c} \mathbf{b}_j^\intercal \mathbf{X}_{i,t,j} \right|, \tag{2.3.9}$$

where w_c is a cluster specific penalty weight, typically pre-specified as $w_c = \sqrt{|G_c|}$. The hyperparameter $\lambda_{\mathrm{CGL},i}$ is, as for other methods, selected by cross-validation. The linear combination $\sum_{j \in G_c} \mathbf{b}_j^\intercal \mathbf{X}_{i,t,j}$ represents the contributions to the predicted

values by individual clusters. Therefore, the CGL tends to choose all variables from cluster c or none of its variables.

We were unsuccessful in applying the CGL in our empirical setup. The CGL leads to high prediction errors, significantly underperforming even the baseline HAR-D model. We do not report the high prediction errors, as they have zero information value. However, we believe that the fact that the CGL does not work in our setup may potentially be valuable information. Its poor performance is not surprising, as even Bühlmann et al. (2013) documents an inferior performance of the CGL in some simulation setups. In our case, it is likely given by the poor signal-to-noise ratio combined with relatively large clusters. All variables within a chosen cluster are non-zero. Therefore, choosing any cluster with a stock that experiences an episode of high volatility outside the CGL estimation window may have a substantial negative impact on forecasting performance. The CRL may address this shortcoming by averaging out the noise.

The CRL is identical to a standard lasso (2.3.3), but uses data-driven clustering and the resulting cluster averages as potential predictors instead of market, sector, or industry averages. Our implementation is slightly different from Bühlmann et al. (2013), who apply clustering to all available potential predictors. We, instead, cluster stocks using k-shape clustering (described below in subsection 2.3.5). We do not use all predictors for clustering. Instead, we cluster stocks using only intraday 30-minute realized log-variances. Based on the stock clustering, we then calculate intraday, daily, weekly, and monthly cluster averages.

2.3.5 k-shape Clustering

The k-shape clustering was introduced by Paparrizos and Gravano (2015) as a fast algorithm to cluster time series. There are other popular clustering methods, such as dynamic time-warping clustering, which are numerically more demanding. As our dataset is large, algorithm speed is important.[6]

The k-shape clustering is similar to the k-means clustering applied to cross-sectional data. To better convey the idea of the k-shaped algorithm, let us describe

[6]For a review of clustering time series methods, see Liao (2005).

the k-means algorithm first.[7] We would like to find clusters of similar observations. For consistency with the previous text, denote $X_i = (x_{i,30}, x_{i,60}, ..., x_{i,T})^\intercal$ for $i \in \{1, 2, ..., m\}$. In the time series setup, X_i is a sampled series, but for the context of this paragraph, X_i represents a multivariate observation. Let index $c \in \{1, 2, ..., k\}$ denote cluster, and G_c is a set of indices of observations belonging to cluster c. The goal of k-means clustering is to find a clustering minimizing distance between cluster means $\mu_c^{\text{mean}} = 1/|G_c| \sum_{i \in G_c} X_i$ and individual observations in the cluster. Formally, the goal is to find

$$\{G_1, G_2, ..., G_k\} = \underset{\{G_1^*, G_2^*, ..., G_k^*\}}{\arg\min} \sum_{c=1}^{k} \sum_{i \in G_c^*} ||X_i - \mu_{c^*}^{\text{mean}}||^2, \tag{2.3.10}$$

where $|| \cdot ||^2$ is the squared Euclidean distance and $\mu_{c^*}^{\text{mean}}$ is the mean of cluster G_c^*. The k-means clustering algorithm can be described in four steps.

1. Randomly assign each observation X_i to one of the k clusters.

2. Calculate cluster means μ_c^{mean} for $c \in \{1, 2, ..., k\}$.

3. Each observation X_i reassign to the cluster with the closest cluster mean, i.e., solve

$$\underset{c^*}{\arg\min} ||X_i - \mu_{c^*}^{\text{mean}}||^2. \tag{2.3.11}$$

4. Repeat steps 2 and 3 until the cluster assignment no longer changes.

The algorithm does not guarantee that the final clustering is a global optimum solving (2.3.10). Therefore, the algorithm is run multiple times, and the best solution is chosen.

The k-means algorithm should not be applied to time series data, as the means and the distance in (2.3.10) ignore the distinctive features of the time series data. More specifically, the mean μ_c^{mean} ignores the possibility that one series could be very similar to another series, but lagged by one or more periods. Therefore, μ_c^{mean} should not be a simple element-wise average of $(x_{i,30}, x_{i,60}, ..., x_{i,T})^\intercal$ for $i \in \{1, 2, ..., m\}$, but a statistic that represents the cluster well even when some series are lagging or leading with respect to others. Instead of calculating the means

[7] For a survey on k-means algorithm, see Ahmed et al. (2020).

that represent the clusters, k-shaped clustering calculates representative series. The representative series are addressed as average shapes of clusters, hence the k-shape name. Another consideration is the distance that is used for measuring similarity and reassignment. The squared Euclidean distance $||X_i - \mu_{c^*}^{\text{mean}}||^2$ takes element-wise differences and ignores potential shifts in time. The k-shape clustering is almost identical to the k-means clustering: the only difference being the distance and shape.

To introduce the average shape, we first have to introduce the cross-correlation measure of two series with a shift $\delta \in \mathbb{Z}$. Consider two series centered around their long-run means, $X = (x_{30}, x_{60}, ..., x_T)^\intercal$ and $Y = (y_{30}, y_{60}, ..., y_T)^\intercal$[8]. The cross-covariance of the two series, where X and Y are relatively shifted by δ periods, is defined as

$$CC_\delta(X,Y) = \begin{cases} \sum_{t=30}^{T-30\delta} x_{t+30\delta} y_t, & \delta \geq 0, \\ \sum_{t=30-30\delta}^{T} x_{t+30\delta} y_t, & \delta < 0. \end{cases} \tag{2.3.12}$$

The shifted cross-covariance (2.3.12) can be used to construct an average shape. Consider cluster c with m_c series $X_1, X_2, ..., X_{m_c}$, each with length T. The average shape $\mu_c = (\mu_{30,c}, \mu_{60,c}, ..., \mu_{T,c})^\intercal$ of the cth cluster is given by

$$\mu_c = \arg\max_{\mu} \sum_{i=1}^{m_c} \frac{\max_\delta CC_\delta(X_i, \mu)}{\sqrt{CC_0(X_i, X_i)\, CC_0(\mu, \mu)}}. \tag{2.3.13}$$

The ratio in (2.3.13) is the highest cross-correlation between series X_i and μ that can be achieved by relative shifting of the two series. The purpose of seeking the maximum is to capture potentially lagging or leading cross-dependence between the two series. The average shape then maximizes the average linear cross-dependence between the shape and all the series within cluster c.

One more definition is needed to state the k-shape algorithm: a measure of distance between two series. Again, consider two centered series X and Y. The k-shape algorithm uses the following distance

$$\mathcal{D}(X,Y) = 1 - \max_h \frac{CC_h(X,Y)}{\sqrt{CC_0(X,X)\, CC_0(Y,Y)}}. \tag{2.3.14}$$

[8]We use 30-minute increments to keep index t consistent throughout the paper.

The second term is the highest cross-correlation that can be attained by the relative shift of two series. Thus, the complementary value to one can be interpreted as a distance.

We are now ready to state the k-shape algorithm. Our goal is to cluster m series $X_1, X_2, ..., X_m$ into k clusters by solving

$$\{G_1, G_2, ..., G_k\} = \underset{\{G_1^*, G_2^*, ..., G_k^*\}}{\arg\min} \sum_{c=1}^{k} \sum_{i \in G_c^*} \mathcal{D}(X_i, \mu_{c^*}), \tag{2.3.15}$$

where μ_{c^*} is (2.3.13) calculated for cluster G_c^*.

The algorithm can be described in four steps.

1. Randomly assign each series to one of the k clusters.

2. Find the average shapes μ_c for each cluster $c \in \{1, 2, \ldots, k\}$, according to (2.3.13).

3. For each series X_i, find cluster c that minimizes the distance $\mathcal{D}(X_i, \mu_c)$, and reassign series X_i to cluster c.

4. Repeat steps 2 and 3 until the cluster assignment no longer changes.

The algorithm is run repeatedly with various initial clusterings in step 1 to avoid a suboptimal clustering caused by an unlucky draw of initial cluster assignment.

The 4-step algorithm represents the overall idea of k-shape clustering, but implementing it directly as described is numerically inefficient. For an efficient implementation that yields identical results, see Paparrizos and Gravano (2015).

Specifically to our case, let $\mathbf{RV}_i^{I*} = [RV_{i,30}^{I*}, RV_{i,60}^{I*}, ..., RV_{i,T}^{I*}]^\intercal$ be the series of adjusted realized log-variances of stock i. We apply the k-shape clustering to the series of adjusted intraday realized log-variances $\mathbf{RV}_1^{I*}$, $\mathbf{RV}_2^{I*}$, ..., $\mathbf{RV}_m^{I*}$. Each stock is assigned to one of the k clusters G_1, G_2, ..., G_k. Note, we do not perform clustering based on daily, weekly, or monthly realized log-variances. We aggregate longer realized log-variances according to the clusters obtained from the clustering of intraday realized log-variances.

2.4 Model Evaluation

We evaluate the forecast performance out-of-sample, using a testing dataset for performance evaluation that differs from the training dataset used for model estimation. Unlike in the case of independent cross-sectional data, with time series data, a time dependence has to be considered. Cerqueira et al. (2020) discuss various out-of-sample evaluation schemes used in the literature and compares them. However, there is no universal consensus on a preferred method.

Here, we follow Zhang et al. (2024) and use a simple chronological train–test split, splitting the dataset of ordered observations into two non-overlapping periods. We use an expanding window with annual re-estimation.

The expanding window with annual re-estimation is our baseline evaluation scheme. To address our research question about the potential existence of short-lived signals, inspired by the work of Chinco et al. (2019), we also use moving windows, with lengths of 15 and 60 days, to represent the training dataset, re-estimating the models on daily bases. To make our results based on the short moving window comparable to those based on the expanding window with annual re-estimation, the last days of the first training windows are identical. Therefore, the out-of-sample forecasts are made for identical periods; only the training datasets differ.

The lasso hyperparameters and the parameters for the diurnal adjustment are based on the training dataset. The k-shape clustering is performed only once, due to its computational demands, and only on data until the end of the first training window. This strict separation ensures that there is no leakage of information from the testing data.

2.4.1 Forecasting Precision

We perform a chronological train–test split, but the forecasts are naturally performed as one-period-ahead predictions. We use three forecast precision metrics to evaluate the quality of the one-period-ahead predictions: mean squared error (MSE), quasi-likelihood (QLIKE), and realized utility (RU).

Patton (2011) studies properties of various loss functions for volatility, including

MSE and QLIKE. We chose MSE and QLIKE as other loss functions considered in his work exhibit some degree of bias arising from the fact that realized volatility is only a proxy for an integrated volatility. Let $\widehat{RV^I_{i,t}}$ be the realized log-variance predicted by a model, and $RV^I_{i,t}$ be the corresponding actual realized log-variance. The MSE and QLIKE are defined as:

$$\text{MSE}_i = \frac{1}{|\mathcal{T}_{\text{test}}|} \sum_{t \in \mathcal{T}_{\text{test}}} \left(RV^I_{i,t} - \widehat{RV^I_{i,t}} \right)^2, \tag{2.4.1}$$

$$\text{QLIKE}_i = \frac{1}{|\mathcal{T}_{\text{test}}|} \sum_{t \in \mathcal{T}_{\text{test}}} \left[\exp\left(RV^I_{i,t} - \widehat{RV^I_{i,t}} \right) - 1 - \left(RV^I_{i,t} - \widehat{RV^I_{i,t}} \right) \right], \tag{2.4.2}$$

where $\mathcal{T}_{\text{test}}$ is a set of time indices corresponding to the testing set.

As we perform diurnal adjustments, the models do not predict $\widehat{RV^{\text{I}}_{i,t}}$ directly, but $\widehat{RV^{\text{I*}}_{i,d,\tau}}$. We then predict the realized log-variances as

$$\widehat{RV^{\text{I}}_{i,t}} = \widehat{RV^{\text{I*}}_{i,d,\tau}} + S_{i,\tau}, \tag{2.4.3}$$

where $S_{i,\tau}$ is given by (2.2.11).

Our goal is to find a model that performs the best for the majority of stocks. In the empirical application, we analyze too many stocks to effectively report the MSE and QLIKE for all of them. In the literature analyzing many assets, it is common to aggregate loss functions like MSE and QLIKE by a simple average (Bollerslev et al., 2018; Gu et al., 2020; Zhang et al., 2024). Averaging loss functions may lead to a loss of information. The model with the best average MSE is not guaranteed to be the model with the best MSE for the majority of the stocks. Therefore, we also report the number of stocks for which one model outperforms the other model in a pairwise comparison.

Additionally, we accompany the reported MSE and QLIKE with Model Confidence Sets (MCS) as developed by P. R. Hansen et al. (2011). The MCS is constructed separately for each metric. We iteratively eliminate the model with the worst MSE or QLIKE and perform hypothesis tests of

$$\mathbb{E}\left[L_{\zeta,t} - L_{\eta,t}\right] = 0 \text{ for } \zeta, \eta \in \mathcal{M}, \tag{2.4.4}$$

where $L_{\zeta,t}$ is a loss function, averaged across stocks at time t for model ζ, and $\mathcal{M}$ is the set of models remaining in the current iteration. More specifically, for MSE and QLIKE, the loss functions for model ζ predicting realized log-variances

$\widehat{RV^I_{\zeta,i,t}}$ for stocks $i \in \{1, 2, ..., m\}$ at time t are

$$L^{MSE}_{\zeta,t} = \frac{1}{m} \sum_{i=1}^{m} \left(RV^I_{i,t} - \widehat{RV^I_{\zeta,i,t}} \right)^2, \tag{2.4.5}$$

$$L^{QLIKE}_{\zeta,t} = \frac{1}{m} \sum_{i=1}^{m} \left[\exp\left(RV^I_{i,t} - \widehat{RV^I_{\zeta,i,t}} \right) - 1 - \left(RV^I_{i,t} - \widehat{RV^I_{\zeta,i,t}} \right) \right]. \tag{2.4.6}$$

Elimination is stopped when no test of (2.4.4) can be rejected. The remaining set obtained by the procedure is called a MCS with the significance level of the tests for (2.4.4) used in the elimination procedure.

There are three important considerations for the correct testing of (2.4.4). Assets are cross-sectionally correlated, predictions by two models for an identical asset and period are potentially correlated, and loss functions (2.4.5) and (2.4.6) may be autocorrelated. Gu et al. (2020) adapted the Diebold-Mariano test for such a setup. The loss functions (2.4.5) and (2.4.6) are averaged across assets, which addresses the cross-sectional correlation of assets. The Diebold-Mariano test is paired; therefore, it takes into account the correlation of $L_{\zeta,t}$ and $L_{\eta,t}$. And standard errors are robust to potential autocorrelation.

The third metric used for assessment of forecast quality of realized utility was introduced by Bollerslev et al. (2018). The motivation behind the realized utility is to establish the utility impact of improving the forecast of realized volatilities within a simple mean–variance utility framework. Bollerslev, Hood, Huss, and Pedersen assume that the expected utility of the investor at time t can be approximated by

$$\mathbb{E}_t[u(W_{t+30})] = \mathbb{E}_t[W_{t+30}] - \frac{1}{2}\gamma^A \operatorname{var}_t(W_{t+30}), \tag{2.4.7}$$

where u is utility function, W_t is wealth, and $\gamma^A \equiv -u''/u'$ denotes the absolute risk aversion of the investor. The investor can allocate fraction x_t of current wealth to a risky asset with return r_{t+30} and the rest to a risk-free asset with return r^f_t. The wealth at time $t+30$ is then given as $W_{t+30} = W_t \left(1 + (1 - x_t) r^f_t + x_t r_{t+30}\right)$. Using $\operatorname{var}_t(r_{t+30} - r^f_t) = \mathbb{E}_t[\exp(RV_{t+30})]$ and assuming a constant Sharpe ratio $SR \equiv \mathbb{E}_t[r_{t+30} - r^f_t]/\sqrt{\mathbb{E}_t[\exp(RV_{t+30})]}$, the expected utility (2.4.7) can be expressed as a function of the ratio x_t as

$$U_t(x_t) = W_t \left(x_t \, SR \sqrt{\mathbb{E}[\exp(RV_{t+30})]} - \frac{\gamma}{2} x_t^2 \mathbb{E}_t[\exp(RV_{t+30})] \right), \tag{2.4.8}$$

where $U_t()$ is the expected utility, and $\gamma = \gamma^A W_t$. An investor maximizing expected utility will allocate wealth with ratio

$$x_t^* = \frac{SR/\gamma}{\sqrt{\mathbb{E}_t[\exp(RV_{t+30})]}}. \tag{2.4.9}$$

We denote an expectation based on model θ as $\mathbb{E}_t^\theta(.)$. An investor using the model instead of an actual expectation allocates wealth as

$$x_t^\theta = \frac{SR/\gamma}{\sqrt{\mathbb{E}_t^\theta[\exp(RV_{t+30})]}}, \tag{2.4.10}$$

believing that it is an optimal allocation. Substituting x_t^θ into (2.4.8) yields utility

$$U_t(x_t^\theta) = W_t \left(\frac{SR^2}{\gamma} \frac{\sqrt{\mathbb{E}_t[\exp(RV_{t+30})]}}{\sqrt{\mathbb{E}_t^\theta[\exp(RV_{t+30})]}} - \frac{SR^2}{2\gamma} \frac{\mathbb{E}_t[\exp(RV_{t+30})]}{\mathbb{E}_t^\theta[\exp(RV_{t+30})]} \right). \tag{2.4.11}$$

Bollerslev et al. (2018) evaluate the expected utility by simply replacing $\mathbb{E}_t[\exp(RV_{t+30})]$ with an actual realized variance $\exp(RV_{t+30})$, averaging over time and dropping the wealth W_t so that the resulting utility is relative to the unit of wealth. The resulting metric

$$RU = \frac{1}{|\mathcal{T}_{\text{test}}|} \sum_{t \in \mathcal{T}_{\text{test}}} \left(\frac{SR^2}{\gamma} \frac{\sqrt{\exp(RV_{t+30})}}{\sqrt{\mathbb{E}_t^\theta[\exp(RV_{t+30})]}} - \frac{SR^2}{2\gamma} \frac{\exp(RV_{t+30})}{\mathbb{E}_t^\theta[\exp(RV_{t+30})]} \right) \tag{2.4.12}$$

is called the realized utility.

We have to choose the Sharpe ratio SR and the risk aversion parameter γ to evaluate the expected utility. Motivated by the empirical literature, Bollerslev et al. (2018) and Zhang et al. (2024) use $SR = 0.4$ and $\gamma = 2$. We set the same values to make our results comparable.

2.5 Data

Our empirical application of the models is based on intraday returns of 363 large stocks obtained from the Kibot database,[9] covering the period from October 29, 2015, to November 8, 2019. The 363 stocks belong to the S&P 500 index; a full list can be found in the appendix.[10] Since trading outside the main trading

[9] www.kibot.com (accessed 15 November 2019).

[10] We included only stocks that have available sector and industry CIGS classifications (see subsection 2.5.1), were included in the S&P 500 index during the analyzed period, and are

session is rather limited and, for some analyzed stocks, practically nonexistent, we limit our work to only the main trading session.

Using the day d and intraday time τ indexing, we calculate 3-minute log-returns from open and close prices observed for non-overlapping 3-minute intervals as

$$r_t = r_{d,\tau} = \begin{cases} \log\left(\frac{\text{Close}_{d,\tau}}{\text{Open}_{d,\tau}}\right) & \text{for } \tau = 3, \\ \log\left(\frac{\text{Close}_{d,\tau}}{\text{Close}_{d,\tau-3}}\right) & \text{for } \tau \in \{6, 9, ..., 390\}. \end{cases} \tag{2.5.1}$$

We model only the main trading hours. Therefore, the return over the first 3 minutes of a trading day is calculated using the open price of the day. Then, we process the returns into 30-minute realized log-variances according to (2.2.3).

The realized log-variances of individual stocks are winsorized to limit the impact of extreme values. Following Zhang et al. (2024), we winsorize the values by the 0.5% and 99.5% percentiles.[11]

2.5.1 Market Sectors and Industries

Models using firm characteristics for grouping are based on the stock classification by MSCI and Standard & Poor's. We use three grouping schemes: the whole market that includes all 363 analyzed stocks, individual sectors, and more granular (sub-)industries,[12] representing various degrees of granularity of the market. There are 11 sectors and 78 industries in our dataset.

The number of stocks and industries is too numerous to present here; for a full list, see Appendix B.1. Figure 2.1 shows the frequency histogram of industries by

available in the Kibot database. The S&P 500 index has several additions and removals every year. There was only one instance of manual removal: PG&E Corp was removed, since the company experienced an extreme price drop in late 2017 and throughout 2018. The volatility was substantially higher than for any other considered stock. Consequently, the reported forecast accuracy metrics, such as mean square error, averaged across the analyzed stocks were exclusively driven by a single company.

[11]Winsorization sets values smaller than the 0.5% percentile to the 0.5% percentile, and the values larger than the 99.5% percentile to 99.5% percentile.

[12]Definition effective from September 28, 2018 till March 17, 2023. We use the definition of sub-industry, as it is the most granular classification. For simplicity, we refer to it as simply 'industry' hereafter.

the number of stocks involved; most industries involve fewer than 5 stocks out of our S&P 363. The histogram demonstrates how finely the market is partitioned by industry but does not say anything about the industries' general sizes.

Table 2.1 lists all sectors and the number of stocks per sector among the 363 analyzed stocks. Industries can be seen to represent a much finer partition of our stocks than sectors, with some of the industries represented only by a single stock.

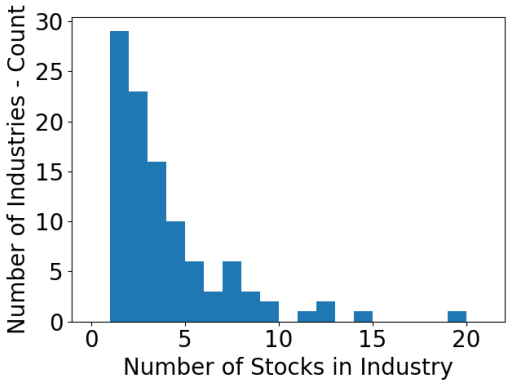


Figure 2.1 S&P industry sizes in our dataset of 363 analyzed stocks.

Sector	Number of Stocks
Communication Services	12
Consumer Discretionary	43
Consumer Staples	28
Energy	14
Financials	52
Healthcare	43
Industrials	48
Materials	16
Real Estate	22
Technology	61
Utilities	24

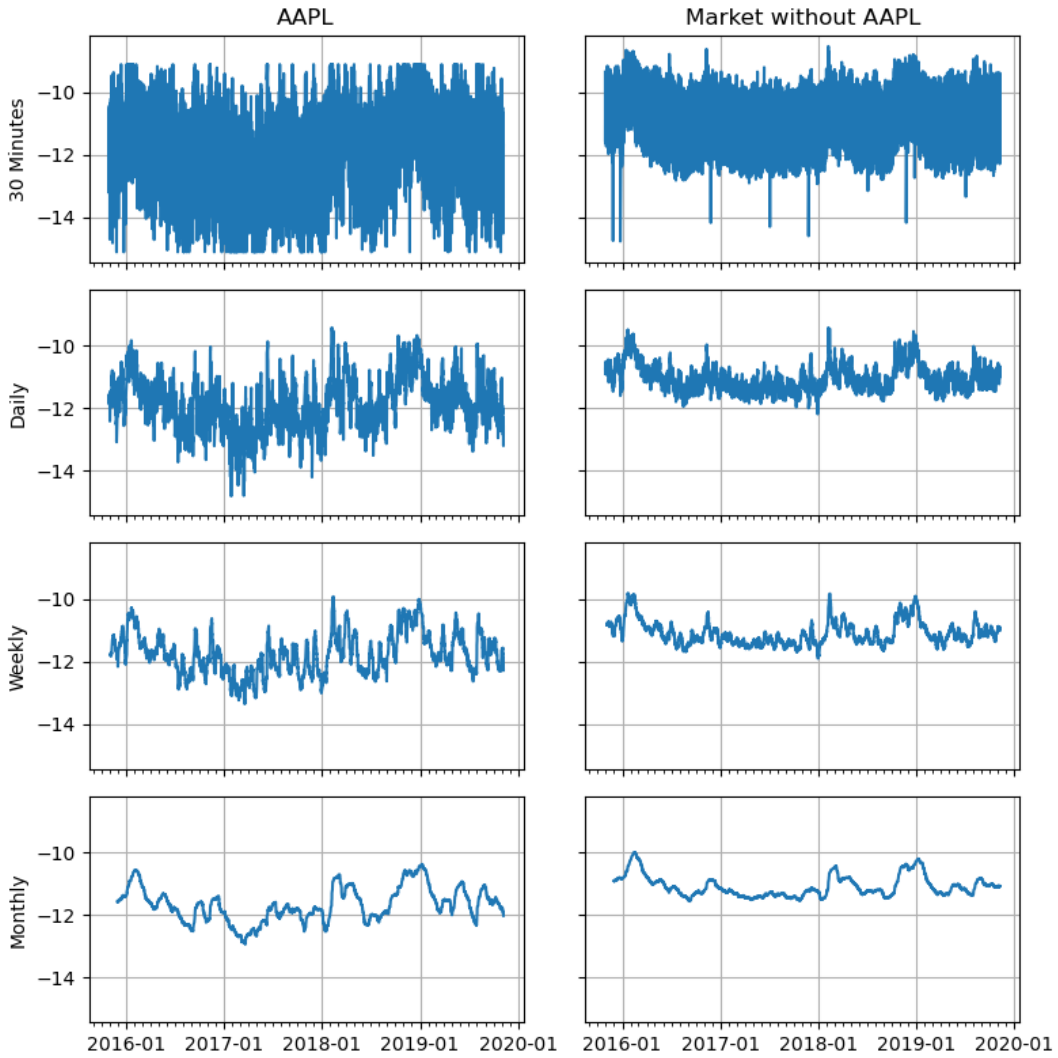
Table 2.1 S&P sector sizes in our dataset of 363 analyzed stocks.

Let us use Apple Inc. stock (symbol AAPL) as an example. Consider a single group representing the market, covering all 363 analyzed stocks. As all the stocks are from the S&P 500 index, the market group well represents stocks with a large capitalization. The averaged realized log-variances representing the market for AAPL alone are calculated from the 362 stocks without AAPL. The plot of all four realized log-variances for AAPL and the market is shown in Figure 2.2. The market average aggregates many stocks; consequently, the market series are smoother, since idiosyncratic risks of individual stocks are canceled out.

2.6 Empirical Performance

The baseline model is the HAR-D model as defined by (2.3.1) with intraday 30 minutes, daily, weekly, and monthly realized log-variances as predictors. Also, the

Figure 2.2 30-minutes, daily, weekly, and monthly realized log-variances for Apple Inc. and averages for the remaining 362 stocks representing the large-cap US market.



HAR-D with external predictors as defined by (2.3.2), which includes predictors based on other stocks averaged according to groups or clusters.

We consider four different grouping schemes. For each group of each grouping scheme and each stock, we calculate group averages as described in Section 2.2.1. The grouping schemes are as follows:

- **Market** – one group, averages of all stocks, with modeled stock being excluded.
- **Sectors** – 11 sectors, averages calculated by sectors.
- **Industries** – 78 industries, averages calculated by industries.
- **Clustering** – we cluster stocks in three different specifications with 5, 10, and 20 clusters by applying the k-shape clustering to intraday realized log-variances.

We tried three clustering specifications to study the potential impact of grouping granularity on the accuracy of the forecast.

By calculating intraday, daily, weekly, and monthly realized log-variances and passing them to the lasso, we indirectly impose lag structure in comparison to including all intraday 30-minute realized log-variances lagged up to a month. We experimented with a completely free lag structure specification, but preliminary results have shown poor performance, which is consistent with the findings of Audrino and Knaus (2016). Therefore, we use a pre-defined HAR structure and let the lasso choose only out of intraday, daily, weekly, and monthly realized log-variances. We consider specifications with two alternative lag structures:

- **Intra** – only intraday predictors based on groups are included among potential predictors. In other words, $RV_{i,t}^{\mathrm{D,c}}$, $RV_{i,t}^{\mathrm{W,c}}$, and $RV_{i,t}^{\mathrm{M,c}}$ for $c \in (1, k)$ are excluded from the estimation. The models still include the long realized log-variances for the own history of the modeled stock.
- **All** – includes intraday, daily, weekly, and monthly realized log-variances for own past and past of groups.

We hypothesize that long realized log-variances of clusters may not be particularly informative for predicting intraday realized log-variances of individual stocks. There are two potential explanations. First, in the long run, a stock price

movements are more affected by the changing intrinsic value of the company than in the short run. Therefore, the stock's own long-run history is more important than the long-term history of other stocks in the market. Second, if there are any volatility spillovers in the long run, they are likely already included in the modeled stock's own history, so including a long history of other stocks would then be redundant.

Table 2.2 shows the number of potential predictors for the groups and lag structures considered. In addition, an intercept is always included.

Table 2.2 Number of potential predictors.

	Lag Structure	
Grouping	Intra	All
Plain HAR-D	4	4
Market	5	8
Sectors	15	48
Industries	82	316
5 Clusters	9	24
10 Clusters	14	44
20 Clusters	24	84

From our specifications, we have excluded the lasso applied directly with ungrouped individual stocks, representing the most "granular" specification. The preliminary results showed poor performance for some stocks with potentially very high MSE. Similar behavior was observed for CGL, which also directly uses all stocks but sets all parameters within clusters to zero or non-zero values. This may be a consequence of the lasso's enforced sparsity, which may simply not match reality as there may be factors shared behind realized log-variance comovements. It is important to note that, for some stocks, applying the lasso with individual stocks provided further improvement over the clustered lasso. However, our goal is to come to fairly general conclusions that are valid for all or most of the analyzed stocks.

We also experimented with the HAR-D model with a more granular grouping

than the market, but OLS has a strong tendency to overfit the data even with as few as 5 clusters. Consequently, the out-of-sample forecast performance was very poor and excluded from the presented results.

2.6.1 Annual Moving Window

The results for annually re-estimated models are shown in Table 2.3. Each model name indicates the method and the grouping used. Forecasting is done for each stock separately; we average the performance across stocks. The results show a clear pattern of models exploiting commonality at a more granular level than the market outperforms HAR-D with market predictors.

Judging from the MSE, the best performing model is estimated with lasso utilizing only intraday average realized log-variances calculated for industries with HAR predictors for own past. It is also the only model left in the 95% model confidence set. Measuring by QLIKE, the adaptive lasso and the clustering representative lasso with 20 clusters perform the best. However, the differences are relatively small, and the most significant improvement comes from breaking down market aggregation into more granular information, rather than various degrees of granularity. The realized utility reveals a similar story, with just a slightly different model ranking order.

The *intra* lag structure specification generally works better, with the market specifications being the only exception. This aligns with our expectation that the longer-term realized log-variances do not contribute much additional information, as potential volatility spillovers are already incorporated into the realized log-variances of own long-term past.

Adaptive lasso, in comparison to lasso, does not bring improvement. As Zou (2006) shows, adaptive lasso has an oracle property but that does not necessarily translate to improved forecasting in a finite sample. The improvement or worsening over the plain lasso depends on the signal-to-noise ratio. Financial data like stock returns or realized log-variances are generally noisy and, as such, likely do not offer a favorable setup for adaptive lasso.

We report MSE, QLIKE and RU averaged across stocks as individual results for each of the 363 stocks would be hard to interpret, and the results presented in

Table 2.3 Comparison of average performance for 363 stocks.

Model	MSE		QLIKE		RU $\times 100$	
	Intra	All	Intra	All	Intra	All
HAR-D	0.500	0.500	0.294	0.294	3.343	3.343
HAR-D, Market	0.499	0.497	0.294	0.294	3.341	3.340
Lasso, Market	0.499	0.497	0.294	0.295	3.340	3.337
Adaptive Lasso, Sectors, $\gamma = 0.5$	0.489	0.494	0.288*	0.293	3.353	3.341
Adaptive Lasso, Sectors, $\gamma = 1$	0.489	0.495	0.288*	0.295	3.354	3.337
Adaptive Lasso, Sectors, $\gamma = 2$	0.490	0.495	0.288*	0.295	3.355	3.336
Lasso, Sectors	0.489	0.491	0.289	0.292	3.352	3.345
Lasso, Industries	0.486*	0.492	0.290	0.294	3.347	3.338
CRL, 5 Clusters	0.489	0.492	0.290	0.295	3.348	3.335
CRL, 10 Clusters	0.489	0.492	0.290	0.295	3.347	3.336
CRL, 20 Clusters	0.488*	0.491	0.288*	0.290	3.353	3.350

* 99% model confidence set, applied only to MSE and QLIKE.

The out-of-sample evaluation period spans August 30, 2017–November 8, 2019. The model is re-estimated annually with expanding window. The Intra and All columns represent two sets of predictors. The setup Intra uses HAR predictors, including intraday, daily, weekly and monthly realized log-variances, for the stock being predicted and only intraday realized log-variances for other clusters or groups. The setup All uses HAR predictors for both the stock and groups or clusters.

this way are comparable to the work of Bollerslev et al. (2018), Zhang et al. (2024), and other similar studies that model many stocks. However, a disadvantage of using statistics averaged across stocks is the loss of information. Some stocks have generally higher volatility than others and consequently a higher level of MSE. Although the differences in MSE across stocks are not substantial, they are still important. Stocks with generally higher MSE contribute more to the average and drive the comparison across methods. A practitioner would likely prefer a model that can consistently deliver superior forecasts for the majority of the stocks rather than a forecast that works better for the more volatile assets driving the average MSE. Therefore, we compare how often one model beats another model when applied to various stocks.

Table 2.4 reports the frequency of one model outperforming another in terms of MSE for the *intra* specification. A further table covering models in *all* specification can be found in the appendix, table C5. Each cell in the matrix presents the case count for how many times (stocks) the model in the row outperformed the model in the column. The best model should, ideally, outperform all other models in at least 50% of cases out of the 363 stocks. The clear winner is the lasso with clustering by industry. The CRL based on 20 clusters comes second, demonstrating that the reported averages across stocks indeed represent somewhat uniform improvement.

The lasso with clustering by industry provides superior predictions. It is important to point out that the period studied does not cover any major stock market crash, such as the Financial Crisis in 2008 or the COVID crash in 2020. The performance of the models during a crisis period is an open question. Zhang et al. (2024) reports improved forecasting precision for autoregressive models utilizing market volatility as a predictor compared to models without market volatility only during March 2020. Otherwise, the post-crash recovery period in terms of forecasting was not visibly different from the pre-crash period. This suggests that utilizing commonality may be beneficial during periods of extreme volatility. However, the positive impact of breaking the market into individual sectors or clusters is unclear. Some clusters may be more relevant during a crisis than others. Moreover, each cluster may have a potentially unique role during various crisis periods. For example, the financial sector during the Financial Crisis

of 2008 is potentially more important for forecasting than utilities. The potential forecasting improvement during historical major stock market crashes remains an open question left for future research.

Table 2.4 Pair-wise counts of MSE model superiority out of 363 stocks - setup Intra.

	HAR-D	HAR-D, Market	Lasso, Market	Adaptive Lasso $\gamma = 0.5$	Adaptive Lasso $\gamma = 1$	Adaptive Lasso $\gamma = 2$	Lasso, Sectors	Lasso, Industries	CRL, 5 Clusters	CRL, 10 Clusters	CRL, 20 Clusters
HAR-D	0 (0.00%)	59 (16.25%)	66 (18.18%)	25 (6.89%)	26 (7.16%)	30 (8.26%)	24 (6.61%)	28 (7.71%)	29 (7.99%)	31 (8.54%)	26 (7.16%)
HAR-D Market	304 (83.75%)	0 (0.00%)	289 (79.61%)	28 (7.71%)	28 (7.71%)	32 (8.82%)	27 (7.44%)	31 (8.54%)	32 (8.82%)	35 (9.64%)	29 (7.99%)
Lasso Market	297 (81.82%)	74 (20.39%)	0 (0.00%)	27 (7.44%)	28 (7.71%)	32 (8.82%)	27 (7.44%)	30 (8.26%)	32 (8.82%)	35 (9.64%)	28 (7.71%)
Adaptive Lasso Sectors, $\gamma = 0.5$	338 (93.11%)	335 (92.29%)	336 (92.56%)	0 (0.00%)	245 (67.49%)	291 (80.17%)	119 (32.78%)	67 (18.46%)	212 (58.40%)	216 (59.50%)	99 (27.27%)
Adaptive Lasso Sectors, $\gamma = 1$	337 (92.84%)	335 (92.29%)	335 (92.29%)	118 (32.51%)	0 (0.00%)	291 (80.17%)	112 (30.85%)	66 (18.18%)	194 (53.44%)	193 (53.17%)	87 (23.97%)
Adaptive Lasso Sectors, $\gamma = 2$	333 (91.74%)	331 (91.18%)	331 (91.18%)	72 (19.83%)	72 (19.83%)	0 (0.00%)	76 (20.94%)	54 (14.88%)	125 (34.44%)	135 (37.19%)	57 (15.70%)
Lasso Sectors	339 (93.39%)	336 (92.56%)	336 (92.56%)	244 (67.22%)	251 (69.15%)	287 (79.06%)	0 (0.00%)	71 (19.56%)	241 (66.39%)	244 (67.22%)	103 (28.37%)
Lasso Industries	335 (92.29%)	332 (91.46%)	333 (91.74%)	296 (81.54%)	297 (81.82%)	309 (85.12%)	292 (80.44%)	0 (0.00%)	317 (87.33%)	304 (83.75%)	259 (71.35%)
CRL, 5 Clusters	334 (92.01%)	331 (91.18%)	331 (91.18%)	151 (41.60%)	169 (46.56%)	238 (65.56%)	122 (33.61%)	46 (12.67%)	0 (0.00%)	190 (52.34%)	72 (19.83%)
CRL, 10 Clusters	332 (91.46%)	328 (90.36%)	328 (90.36%)	147 (40.50%)	170 (46.83%)	228 (62.81%)	119 (32.78%)	59 (16.25%)	173 (47.66%)	0 (0.00%)	80 (22.04%)
CRL, 20 Clusters	337 (92.84%)	334 (92.01%)	335 (92.29%)	264 (72.73%)	276 (76.03%)	306 (84.30%)	260 (71.63%)	104 (28.65%)	291 (80.17%)	283 (77.96%)	0 (0.00%)

The numbers represent absolute numbers of stocks for which the model in the row outperforms the model in the column in terms of out-of-sample MSE.

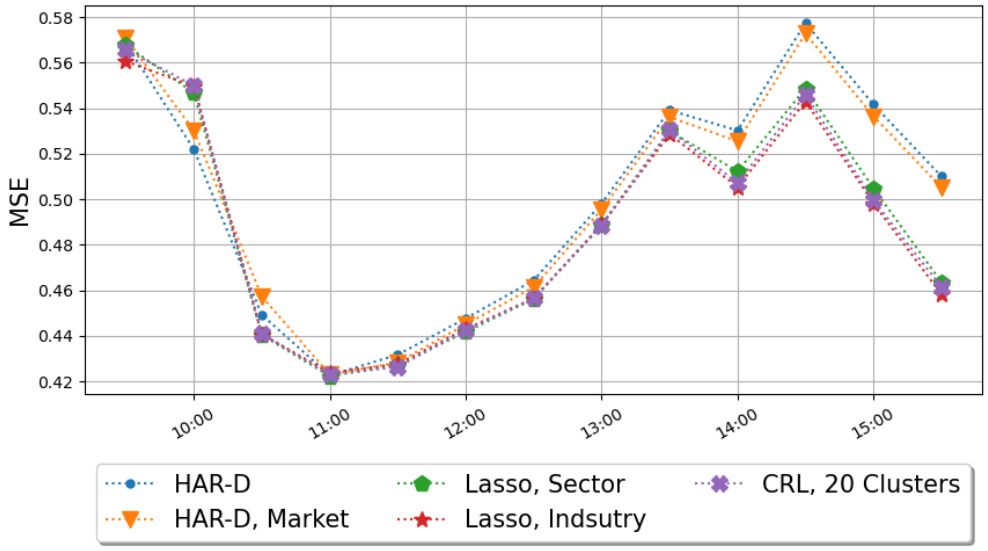
The out-of-sample evaluation period spans August 30, 2017–November 8, 2019.

2.6.2 Diurnal Pattern of Predictability

Another, frequently overlooked, aspect of intraday forecasting is performance across the day. The intraday volatility exhibits the well-documented U-shape pattern. It is not clear, however, how or even whether this pattern somehow translates into forecast precision. Ideally, a better forecasting method should be uniformly superior regardless of the time of day.

Figure 2.3 shows the average MSE for the chosen methods broken by intraday times with the *intra* specification. The improvement achieved by breaking the market into sectors, industries, or data-driven cluster is clearly realized during all periods, with an exception of the second period spanning 10:00 – 10:30. The second period of the day uses the first 30 minutes of the day as predictors. The trading activity during the opening period may be of a different nature than trading later in the day. Lasso selects variables using observations covering all intraday periods, which may be a sub-optimal variable selection when the lagged realized log-variances are of the first period of the day. In addition, incremental improvement seems to be gained throughout the day. These patterns are probably a consequence of the commonality mechanisms described and further explored in the following section 2.7.

Figure 2.3 Average MSE for 363 stocks broken by trading time – *intra* specification.



2.6.3 Short-lived Signals

The commonality and relevance of clusters or groups may vary over time: clusters or groups may be more relevant for some periods and be irrelevant during others. As Chinco et al. (2019) demonstrate, short-lived signals in stock returns can potentially be captured by regularly re-running lasso with a short training window. To test whether this applies in our setup, we update the moving window daily with 15- and 60-day training windows. We estimate our model with the *intra* specification. To make our results comparable to the models re-estimated annually, we limit the evaluation period to the same timeframe, viz. August 30, 2017–November 8, 2019.

Table 2.5 shows results averaged across stocks, with values directly comparable to the annually re-estimated model results in table 2.3. Breaking the market into sectors, industries, or clusters still yields improvement. However, all of the models perform even worse than the annually re-estimated HAR-D model. Therefore, if there are short-lived signals, the signal-to-noise ratio is not sufficiently high for the signals to be captured in our setup.

This is not a definitive proof that the short-lived signals cannot be captured, as it may work only in higher frequencies. Chinco et al. (2019) use 1-minute returns, while we are using 30-minute realized log-variances to avoid microstructure noise. Limit order book data on the best available bid and ask prices, combined with some properly tuned measure of volatility robust to microstructure noise, could allow one to reliably calculate the realized log-variances at higher frequencies than every 30 minutes and, potentially, to detect short-lived signals.[13] However, tuning robust measure parameters is more an art than an exact science; such an investigation is beyond the scope of this paper.

2.7 Market and Sector Commonality

The commonality in liquidity (the common evolution of liquidity of a class of assets) has been extensively and widely studied. Empirical studies are supported

[13]For more on microstructure noise and robust measures, see, for example, Boudt and Zhang (2015), Jacod et al. (2017), or Li and Linton (2022).

Table 2.5 Comparison of average performance for 363 stocks, daily re-estimation with moving window of 15 and 60 day lengths.

Model	MSE		QLIKE		RU $\times 100$	
	15d	60d	15d	60d	15d	60d
HAR-D	0.579	0.511	0.374	0.300	3.112	3.326
HAR-D, Market	0.690	0.513	0.499	0.301	2.742	3.326
Lasso, Market	0.587	0.509	0.364	0.299	3.159	3.331
Adaptive Lasso, Sectors, $\gamma = 0.5$	0.565	0.502	0.351	0.295	3.186	3.340
Adaptive Lasso, Sectors, $\gamma = 1$	0.569	0.503	0.357	0.295	3.169	3.338
Adaptive Lasso, Sectors, $\gamma = 2$	0.575	0.504	0.364	0.296	3.149	3.337
Lasso, Sectors	0.563	0.501*	0.348	0.294*	3.196	3.341
Lasso, Industries	0.569	0.502*	0.355	0.299	3.178	3.326
CRL, 5 Clusters	0.562	0.501*	0.347	0.293*	3.201	3.344
CRL, 10 Clusters	0.565	0.502*	0.360	0.294*	3.156	3.341
CRL, 20 Clusters	0.567	0.501*	0.353	0.296	3.183	3.334

* 99% model confidence set, applied to MSE and QLIKE.

The out-of-sample evaluation period spans August 30, 2017–November 8, 2019. The model is re-estimated daily with a moving window of 15 and 60 business day lengths. The results are for the *intra* specification, which uses intraday, daily, weekly, and monthly realized log-variances of modeled stock and only intraday predictors for groups or clusters.

by various theoretical mechanisms that may explain a commonality in liquidity. Although the focus is liquidity, the mechanisms are interesting: they typically involve a potential commonality in prices, trading volumes, and volatilities, and are, therefore, also relevant to our work as well. The goal of this section is to relate our forecasting results to the existing body of knowledge on commonality in the markets. In the literature, there are two branches of explanations behind commonality: supply-side and demand-side.

Let us first focus on supply-side mechanisms. Brunnermeier and Pedersen (2009) formulate a theoretical model with episodes of sudden decrease of liquidity that creates liquidity spirals (mechanisms in which an externally driven decline in trades lowers market liquidity). Lower market liquidity leads to higher margins for traders and losses on existing positions, as existing positions become more expensive to close due to the lack of liquidity. The losses and higher margins lead to funding problems, which translate into less trading, triggering the next cycle of the spiral. With a positive autocorrelation in volatility and uninformed traders in the market, the liquidity spiral mechanism predicts a common increase in market volatility. As margins typically do not change significantly within a day, spirals based on higher margins seem less relevant to our results. However, the channel based on losses on existing positions is highly relevant for intraday trading. Trading volumes throughout the day exhibit strong diurnal patterns; see, for example, Gourieroux et al. (1999). There is typically less trading later in the morning and in the afternoon compared to market openings; therefore, the commonality could be stronger later in the trading day.

Anagnostidis and Fontaine (2020) document the role of high-frequency trading as a potential spiral amplifier in intraday setup and propose the introduction of maximum bid-ask spread limit rules to contracts with market makers to mitigate spirals. Liquidity spirals can be amplified by high-frequency trading but still directly involve human decision making; we hypothesize that spirals do not start and end instantaneously. Consequently, the commonality generated by liquidity spirals can lead to a predictable increase in volatility. Therefore, volatility forecasting for assets with high commonality may benefit more from including commonality in predictive models than for assets with low commonality.

There are other similar supply-side mechanisms, such as "liquidity black holes" (Bernardo & Welch, 2004; Morris & Shin, 2004), where traders have private trading price limits that generate mutually reinforcing liquidation loops.[14] To the best of our knowledge, there are no theoretical models focusing on commonality within the context of various groups of stocks within a single economy that would perfectly fit our setup. However, the empirical work of Coughenour and Saad (2004) documents commonality between stocks handled by the same specialist firm. Firms or funds that focus on a specific part of the economy could give rise to clustered commonality given by the liquidity black holes that could explain our improvement in volatility forecasts.

The demand-side mechanisms are based on the correlated behavior of institutional investors. A potential reason for correlated behavior may be basket trading, where traders are not interested in individual assets but in baskets of assets. For example, Koch et al. (2016) document high commonality among stocks with high mutual fund ownership arising from funds' correlated liquidity shocks. An increase in the popularity of exchange traded funds (ETFs) and the constantly growing range of ETFs specialized on themed groups of assets and sectors of the economy could lead to clustered commonality due to potential disbalance in inflow or outflow to the funds forming a cluster.

Another demand-side mechanism is the sentiment of noise traders that may induce comovement. Baker and Wurgler (2006) shows that waves of investor sentiment can affect groups of stocks at the same time. As sentiment and fund ownership can vary between clusters of assets, the commonality could vary between clusters as well.

Since the main goal of our work is to utilize commonality for forecasting, we examine the role of commonality in forecasting. We are not trying to explain the driving forces behind the existence of commonality. We limit our analysis only to sectors. We do not perform analysis at the industry or cluster level, as doing so would make the results difficult to present, interpret, and may invoke difficulties due to high dimensionality. Although the sector-based models were not the best performing, their performance was reasonably close to the best performing models

[14]For a review and discussion of other mechanisms, see Karolyi et al. (2012).

with industries and 20 clusters. Based on the mechanisms discussed, we do expect 1) commonality varying across sectors and 2) a positive relationship between commonality and forecast accuracy.

Commonality is understood, in the traditional sense, as comovement during the same period and is typically measured by adjusted R^2. Bollerslev et al. (2018) measures commonality of realized variance by estimating the following regression for each stock and intraday time separately:

$$RV^I_{i,t} = \alpha_i + \beta_i RV^{I,Market}_t + \varepsilon_{i,t}, \tag{2.7.1}$$

where $RV^{I,Market}_t$ is market volatility represented by averaged realized log-variances across stocks and $\varepsilon_{i,t}$ is a zero-mean regression error. For $RV^{I,Market}_t$, we use the definition described in the section on cross-sectional aggregation, with the predicted stock left out to maintain consistency with the rest of our work. The average across adjusted R^2 values of individual stocks is the measure of commonality. Following Zhang et al. (2024), we estimate the regression for each intraday period separately; therefore, we do not apply the diurnal adjustment.

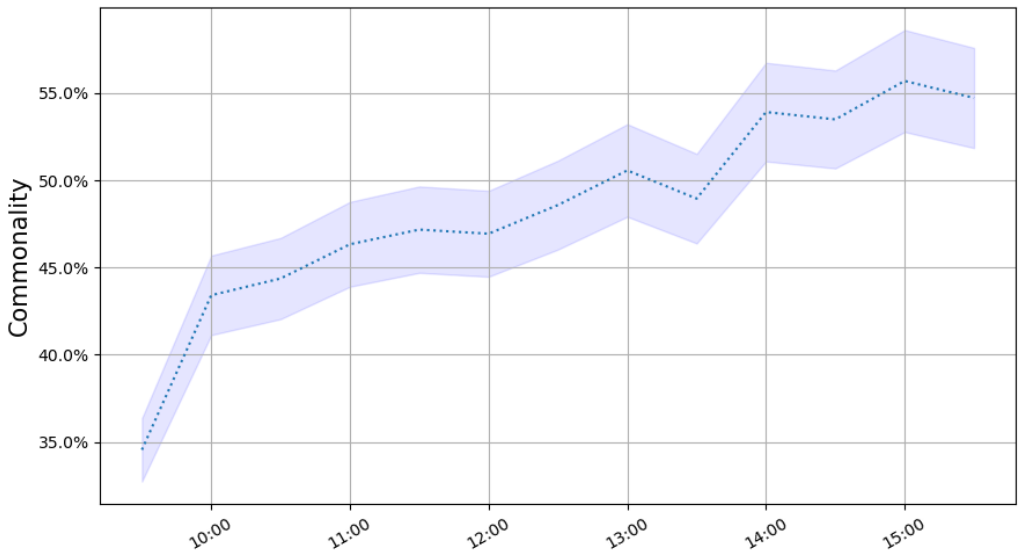
Figure 2.4 shows the commonality of the market throughout the day measured by average R^2 with a standard deviation range. Although we analyzed different periods and have a substantially larger number of stocks than Zhang et al. (2024), the pattern is very similar to theirs. The low commonality and small improvement in the MSE in the morning and the higher commonality with a stronger improvement in the MSE during the later trading hours suggest that volatility in the morning is relatively less driven by common factors and more driven by unique factors of the stock. This pattern is aligned with expectations based on the liquidity spiral.

For studying sector-level commonality, we use a straightforward extension with partial R^2. Our goal is to measure unique and relevant information originating from a specific sector. First, we estimate regression,

$$RV^I_{i,t} = \alpha_i + \sum_{c=1}^{11} \beta_{i,c} RV^{I,c}_t + \varepsilon_{i,t},$$

where $RV^{I,c}_t$ is the average realized log-variances of sector c as defined by (2.2.8). For each sector, we calculate the partial R^2, which we then average across stocks grouped by sector. This results in 11 averages for each of the 11 sectors.

Figure 2.4 Market commonality in realized log-variance – average adjusted R^2.



The resulting average partial R^2s form an 11×11 matrix for each period of the day. Table 2.6 shows the averages for the first 30 minutes of the day. The rows represent sectors for which R^2 is averaged, while the columns represent which sector was left out to calculate partial R^2. The diagonal metrics are substantially higher than all off-diagonal elements, but generally lower than the observed 35% commonality for the entire market. This means that the majority of the comovements are intra-sector comovements or movements of the entire market, while cross-sector comovements unique to a single pair of sectors are weak or rare. The same conclusions hold for all time periods within a day.

Figure 2.5 shows an evolution of the diagonal elements throughout the day. Unlike in the case of market commonality, the diurnal pattern in sector commonalities is not strongly visible. On the other hand, there is a clearly visible non-homogeneous commonality across sectors. If improvement in forecast accuracy is indeed driven by commonality, this may suggest that some sectors may benefit more from breaking market averages into more granular averages than others. This variability across sectors could be explained by basket trading, investor sentiments, and possibly uneven fund ownership across sectors.

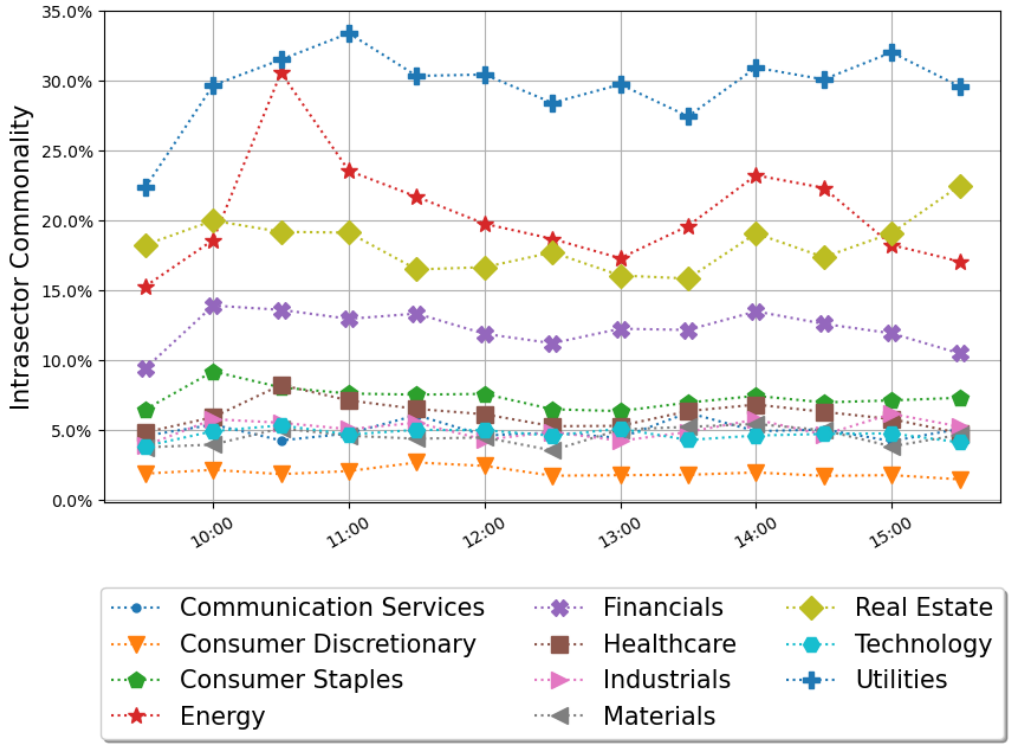
Zhang et al. (2024) use commonality to rank and group stocks into bins separated by quintiles, starting with 20% of their analyzed stocks with the lowest commonality up to group with 20% of stocks with the highest commonality to

Table 2.6 Partial sector commonality matrix in realized log-variance 9:30–10:00.

Averaged \ Reduced	Communication Services	Consumer Discretionary	Consumer Staples	Energy	Financials	Healthcare	Industrials	Materials	Real Estate	Technology	Utilities
Communication Services	4.6%	0.5%	0.5%	0.3%	0.6%	0.3%	0.7%	0.7%	0.8%	1.4%	0.4%
Consumer Discretionary	0.5%	1.9%	0.4%	0.5%	0.3%	0.5%	0.7%	0.4%	0.3%	0.5%	0.5%
Consumer Staples	0.4%	0.6%	6.5%	0.3%	0.2%	0.3%	0.3%	0.5%	0.5%	0.6%	0.5%
Energy	0.2%	0.6%	0.6%	15.3%	0.3%	1.0%	0.1%	0.5%	0.6%	0.1%	0.4%
Financials	0.4%	0.2%	0.3%	0.3%	9.4%	0.5%	0.7%	0.3%	0.4%	0.4%	0.4%
Healthcare	0.4%	0.5%	0.5%	0.4%	0.2%	4.8%	0.3%	0.4%	0.6%	0.6%	0.2%
Industrials	0.2%	0.6%	0.4%	0.3%	0.4%	0.4%	3.9%	0.5%	0.3%	0.4%	0.3%
Materials	0.3%	0.4%	0.5%	1.0%	0.4%	0.6%	0.6%	3.8%	0.6%	0.3%	0.4%
Real Estate	0.3%	0.3%	0.5%	0.3%	0.2%	0.3%	0.2%	0.6%	18.3%	0.4%	0.5%
Technology	0.5%	0.4%	0.4%	0.6%	0.2%	0.4%	0.6%	0.3%	0.6%	3.8%	0.3%
Utilities	0.3%	0.3%	0.3%	0.7%	0.3%	0.6%	0.2%	0.3%	0.9%	0.2%	22.4%

Partial R^2 averaged across stocks within sector by row; column heads indicate which sector was removed.

Figure 2.5 Sector commonality in realized log-variance – average partial R^2.

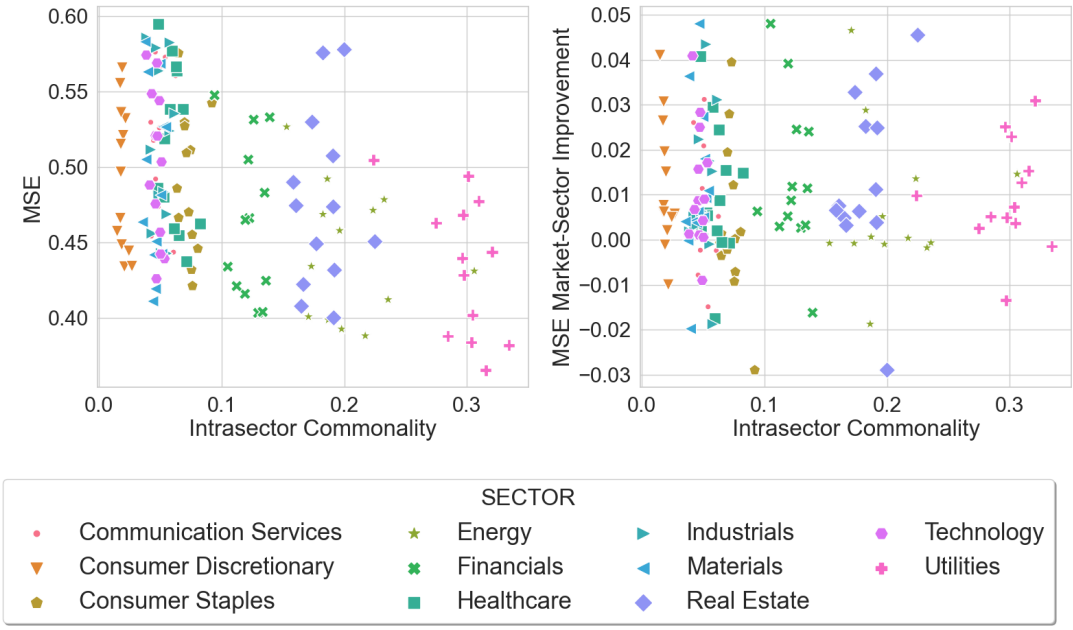


demonstrate the relationship between realized log-variance forecast accuracy and commonality. They find that the quintile of stocks with the lowest commonality have a significantly worse forecast accuracy improvement when including market commonality. We perform a similar exercise to answer whether sectors with strong intrasector commonality benefit more from including sectors among predictors in comparison to the specification with just market-based predictors.

Figure 2.6 shows a relationship between the MSE of the model lassoing sectors and commonality, and the improvement of MSE over the HAR-D model with the market intraday realized log-variance predictor. There is a clear negative relationship between intrasector commonality and MSE of the sector lassoing model, with a correlation coefficient of 44%. Sectors with higher commonality are more predictable.

In contrast, the improvement gained from breaking the market into sectors does not grow with intrasector commonality. The commonality measures comovements within the same period, whereas forecasting relies on lagged values. Since there is no clear trend, the potential explanation for the forecast improvement is liquidity

Figure 2.6 Relationship between intrasector commonality and MSE of lassoing sectors, and comparison against HAR-D model with market intraday realized log-variance predictor.



spirals that take some time to manifest. The forecast improvement likely does not emerge from basket trading, as basket trades are expected to be realized at the same time. The potential disconnection of intrasector commonality and improvement of forecasts could be explained by heterogeneity in presence of various mechanisms generating intra-sector commonality of each sector.

2.8 Conclusion

In this study, we utilize clustered commonality in intraday volatility to improve the forecast accuracy of the intraday realized log-variances of stocks. We break the market into sectors, industries, and data-driven clusters, and lasso their average realized log-variance. Breaking the market into more granular partitions gives superior forecasting performance. Data-driven clusterings of various granularities and GICS classifications provide improvements of similar scale. The improvement is observed in a clear majority of the stocks analyzed. The results suggest that the relevance of clusters and groups is long-term, as shorter moving windows potentially capturing short-lived signals provide inferior performance. The forecast improvements exhibit diurnal patterns, with a stronger improvement in the afternoon trading hours.

We study sector commonality using cross-sectional regressions to better understand the mechanisms behind clustered commonality. We show that intrasector commonality is stronger than cross-sector commonality, and forecasting improvement in terms of MSE is stronger in sectors with stronger intrasector commonality, which could be explained by liquidity spirals.

There are several possible directions for future research. One is to develop a cluster-based panel model. Bollerslev et al. (2018) shows that a panel version of the HAR model, where parameters for the dynamic components are identical for all stocks, outperforms the HAR model estimated separately for individual stocks. Our accuracy improvements and intrasector commonality measures suggest that panel lassoing within clusters or groups could potentially further improve the forecast accuracy in comparison to the market-wide panel estimation. Another research direction could be the use of level-1 limit order book data together with microstructure robust volatility measures to further explore the possibility to capture potential short-lived signals such as Chinco et al. (2019) achieved with return forecasts.

3 Intraday Kurtosis in the Stock Market and Extreme Risk Spillovers

3.1 Introduction

Distributions of asset returns exhibit heavy tails. This is a well-studied phenomenon in the financial and time-series econometrics literature, with heavy tails typically represented by kurtosis. Time series models for asset returns frequently only capture the heavy tails with unconditional kurtosis, which is constant in time. The goal of these models is to improve the quality of distributional forecasting, and kurtosis is not an object of interest per se (Franses & Ghijsels, 1999; He & Teräsvirta, 1999). In contrast, studies on time-varying kurtosis are often interested in testing economic hypotheses, in addition to improving forecasts. Given that high conditional kurtosis is associated with a risk of extreme events, it can be used to study various effects related to extreme risk, such as spillovers of extreme risk within or between markets.

We study the conditional kurtosis of intraday asset returns; returns that are realized within one day, e.g., at a five-minute frequency. Although conditional kurtosis of daily returns has been extensively studied in the literature, conditional kurtosis of intraday returns is poorly understood, leaving a gap that this work aims to fill. The intraday setup poses a unique challenge, because models for daily returns may simply not work in this context. In this work, we aim to better understand conditional kurtosis of intraday asset returns, including its existence and properties, as well as intraday spillovers of extreme risk between stocks. One advantage of the intraday setup is its higher granularity. For example, with daily conditional kurtosis, we cannot answer the question of whether shocking new information is immediately incorporated into the stock price or if the price adjustment is gradual, when the gradual adjustment takes less than one day to fully incorporate the shocking new information. To study conditional kurtosis

and extreme risk spillovers between individual stocks, we develop a model with a novel distribution for time-varying kurtosis in the intraday setup.

The idea of time-varying higher moments—including kurtosis—was pioneered by B. E. Hansen (1994) inventing an autoregressive conditional density (ARCD) model, and was further developed over the following decades by, e.g., Jondeau and Rockinger (2003), Brooks et al. (2005), León et al. (2005), Creal et al. (2013), and Anatolyev and Petukhov (2016), among many others. Research into conditional time-varying kurtosis of asset returns is typically motivated by improved distributional forecasting or economic research questions related to extreme risk.

The goal of modeling kurtosis in the context of forecasting literature is to improve distribution-based risk metrics such as value-at-risk (VaR) or expected shortfall (ES), which are important for risk and portfolio management. VaR is a quantile, and ES is an expectation conditioned on return exceeding a certain quantile. Bali et al. (2008) and Guermat and Harris (2002) exemplify applications of VaR forecasting. More recently, literature and financial practitioners have shifted from VaR to ES, which provides a better summary of tail risk than a simple quantile. Because ES depends on the entire tail of the distribution, the role of conditional kurtosis is even more important (Du & Escanciano, 2017; Ergün & Jun, 2010). Kurtosis also plays a role in portfolio optimization (Zhao et al., 2015).

Extreme risk research focused on kurtosis in the context of economic theory typically studies its role in asset pricing and risk spillovers. Examples of structural asset pricing with conditional kurtosis can be found in C. R. Harvey and Siddique (2000) and Smith (2007). Other researchers focus on understanding the role of kurtosis in pricing of financial derivatives (Hafner & Herwartz, 2001), or the role of kurtosis in the dynamics of volatility indices, such as Lalancette and Simonato (2017). Another research stream relates to extreme risk spillovers between markets in an intraday setup, typically employing a threshold-based definition of extreme returns and studying Granger's causality of extreme events between markets (Hong et al., 2009; Mazzarisi et al., 2020). Another relevant example is Nekhili and Bouri (2023), who study spillovers and kurtosis co-movements using realized measures and conclude that higher moments beyond variance matter for spillovers

between markets.

Bringing the idea of conditional kurtosis to intraday returns proves more challenging than anticipated. We examine a dataset of intraday returns on 100 highly liquid US-listed stocks, covering the period from October 29, 2015, to November 8, 2019. As our first attempt to model intraday kurtosis, we apply the GARCHK-t model of Brooks et al. (2005), originally used in the daily setups. However, we experience difficulties with model estimation, as the conditional kurtosis process degenerates into nested time-constant kurtosis. As kurtosis in an intraday setup is generally lower than in a daily setup, we hypothesize that it could be caused by the imposed assumption of intraday returns always being leptokurtic, i.e., always with conditional kurtosis higher than three.

To address the hypothesis of a potentially platykurtic distribution, we propose to use a distribution of innovations that allows conditional leptokurtic and platykurtic shapes [1]. A generalized skewed t-distribution could be a potential candidate. However, kurtosis of a generalized skewed t-distribution is given by two parameters and there is no one-to-one mapping between kurtosis and the parameters. Therefore, it is not possible to simply specify the dynamics of conditional kurtosis and then map the filtered values of kurtosis to the values of parameters. One-to-one mapping is an essential property for studying extreme risk spillovers, as kurtosis can be interpreted, while multiple latent time-varying parameters would be hard to interpret. Nelson's (1991) generalized error distribution (GED) is an alternative candidate, being a special case of generalized t-distribution that allows leptokurtic and platykurtic shapes to be driven by a single parameter. Preliminary results have shown that the estimates obtained using GED do not degenerate to a time-constant kurtosis. However, the distributional fit in terms of likelihood and information criteria is considerably inferior compared to even the time-constant Student's t-distribution. Additionally, as A. Harvey and Lange (2017) note, the GED shape parameter changes both the heaviness of tails and the peak of the distribution. As the authors discuss, the GED—unlike the t-

[1]Throughout this chapter, a distribution with a platykurtic shape refers to a symmetric distribution with kurtosis lower than that of a normal distribution, and a leptokurtic shape refers to a symmetric distribution with kurtosis higher than that of a normal distribution. Parameters that determine kurtosis will be referred to hereafter as shape parameters.

distribution—is limited in the heaviness of the tails and does not allow fat tails, reflecting undesirable properties for studying extreme risk spillovers. Therefore, we propose a new distribution based on Tukey's H-transformation, in which kurtosis is given by a single parameter and allows leptokurtic and platykurtic shapes at the same time. This proposed distribution will hereafter be referred to as the H-distribution.

Besides the form of the conditional distribution, we also adapt the dynamics of conditional volatility and kurtosis of the GARCHK model to the intraday setup. Applying the GARCHK model in an intraday setup poses a higher risk of mistakenly capturing variance through kurtosis. As Engle and Sokalska (2012) and Bekierman and Gribisch (2021) document, intraday volatility is highly persistent and mixed-frequency models are a potential solution for capturing this high persistence. Based on our empirical observations, persistence typically grows with increasing frequency. Inspired by these two studies on intraday volatility, we use mixed-frequency volatility: one updated daily capturing long-term dynamics, and one updated with every new intraday observation capturing short-term dynamics. We update the daily volatility with the realized volatility of the previous day, and our results confirm the importance of mixed-frequency volatility.

We use the H-distribution, Student's t-distribution, and GED to capture kurtosis. All three distributions have variance influenced by the shape parameter, e.g., degrees of freedom of the t-distribution, which are supposed to represent time-varying conditional kurtosis. With high-frequency stock returns, models can potentially capture high volatility persistence through the process of time-varying parameters intended for kurtosis. Therefore, in addition to the mixed-frequency dynamics for volatility, inspired by Brooks et al. (2005), we allow the free movement of kurtosis with respect to volatility using standardized distributions with unit variances.

We apply the proposed models to 100 highly liquid US-listed stocks, and ask four main research questions: Does conditional kurtosis of intraday stock returns exist? What are the dynamics of conditional kurtosis? Does conditional kurtosis improve the in- and out-of-sample forecast? Do extreme risk spillovers exist between the stocks?

Using the new proposed model based on Tukey's H-transformation, we find a relatively weak yet significant presence of time-varying kurtosis in individual stock returns. None of the reported estimates leads to filtering conditional kurtosis smaller than three, suggesting that the potential platykurtic shapes are not necessary to identify conditional kurtosis. However, the proposed H-distribution still seems better suited for identifying conditional kurtosis of intraday returns than the Student's t-distribution. Furthermore, it makes numerical optimization of maximum likelihood estimation (MLE) more robust, as gradient-based methods such as the Broyden–Fletcher–Goldfarb–Shanno algorithm hits platykurtic shapes on their iterative path to the optimum. We experiment with various specifications of the kurtosis dynamics and find that the quartic specification of Brooks et al. (2005) used for daily returns is also a sensible choice for intraday returns. Capturing conditional kurtosis improves in- and out-of-sample distributional forecasts measured by likelihood forecasts. The improvement in distribution forecasting is rather small, as potential improvements in VaR or ES forecasts was beyond measurable, and as such is not reported.

To study the extreme risk spillovers, we construct a proxy measure for extreme shocks in the market based on filtered standardized returns of 100 highly liquid stocks. This measure does not aim to capture extreme shocks to the market as a whole, but rather to detect one or more stocks in the market experiencing extreme shock, which could then spill over to the kurtosis of other stocks.

By adding the lagged proxy measure for extreme shocks in the market to the kurtosis dynamics, we can test whether new shocking information in the market is immediately incorporated into prices. Empirical results show that conditional kurtoses of individual stocks are causally influenced by extreme shocks of other stocks in the sense of Granger's causality, and are mainly driven by the spillovers from the market rather than by shocks from their own past. These findings suggest that extreme shocks are not immediately incorporated into stock prices, and the main contributors to the potential of extreme risk are not directly linked to individual stocks but represent spillovers from other stocks.

Our empirical findings on extreme risk spillovers are relevant to the literature on the market efficiency and speed of incorporating new information. In general,

prices do not always reflect shocking new information immediately, but rather the information is gradually incorporated, at speeds that are influenced by investor attention (Hirshleifer et al., 2009), the complexity of the shocking news (Loughran & McDonald, 2014), and potentially other factors. The work of Chen (2024) is particularly relevant, because it studies the clustering of discontinuous jumps, which are likely the main driver of conditional kurtosis. The author asks two questions: 1. whether new shocking information is immediately incorporated through one discontinuous jump or if the market incorporates new information as a series of jumps; and 2. if the new information is incorporated as a series of jumps, whether the series of jumps is self-excitation or a spillover from other stocks. Our conclusions are well-aligned with Chen, as we detect periods of increased kurtosis and spillovers as the main drivers. However, Chen (2024) uses daily setup and jump filtration, whereas we model high-frequency intraday kurtosis, and our conclusions are within the scope of minutes. Therefore, we can detect the gradual incorporation of new shocking information that seems immediate from a daily perspective.

In section 3.2, we introduce the notation for our intraday setup and the notion of realized volatility. In section 3.3, we propose our novel distribution based on Tukey's H-transformation and models used for conditional kurtosis. In section 3.4, we describe the dataset of 100 highly liquid US-listed stocks. We apply the proposed models to intraday stock returns in section 3.5, and study the dynamics of conditional kurtosis. Section 3.6 introduces a proxy measure for extreme shocks in the market and studies extreme risk spillovers. Finally, section 3.7 concludes.

3.2 Intraday Setup

In this work, we model intraday log returns. It is convenient to define double indexing for the day and intraday period. $t \in \{1, 2, ..., T\}$ is the time index of ordered observations in our dataset. We define indices for the intraday period $\tau(t) \in \{1, 2, ..., \tau_{max}\}$ and days $d(t) \in \{1, 2, ..., D\}$, where τ_{max} is the number of

intraday periods within each day, and D is the number of days in the dataset, as

$$t(d,\tau) = (d-1)\frac{T}{D} + \tau, \tag{3.2.1}$$

$$d(t) = \left\lceil t/\frac{T}{D} \right\rceil, \quad \tau(t) = t - (d(t)-1)\,\frac{T}{D}, \tag{3.2.2}$$

where $\lceil \cdot \rceil$ is the ceiling function. To keep our notation light, we use τ and d without an input argument, and interchangeably with t. Additionally, we use only d to indicate variables that change only daily or τ to indicate variables that change throughout the day with intraday time τ, but identical values repeating each day.

For each day d and intraday time τ, we define (log) returns as follows:

$$r_{d,\tau} = \begin{cases} \log\left(\frac{\text{Close}_{d,1}}{\text{Open}_{d,1}}\right) & \text{for } \tau = 1 \\ \log\left(\frac{\text{Close}_{d,\tau}}{\text{Close}_{d,\tau-1}}\right) & \text{for } \tau > 1 \end{cases}. \tag{3.2.3}$$

Because returns outside the main trading session are inherently different, we drop overnight returns, which is a common practical solution in the literature (Engle & Sokalska, 2012).

3.2.1 Realized Volatility

We use daily realized volatility to update daily volatility in our models. Merton (1980) notes that volatility can be measured with arbitrary precision as the sampling frequency grows in a time-continuous framework and under fairly general assumptions. Measuring volatility is a challenging task, since real-world returns are burdened with micro-structure noise and discontinuous jumps. For example, Ait-Sahalia et al. (2011), Mykland et al. (2019), and Ait-Sahalia, Jacod and Xiu (2020) address the challenges.

Because we do not aim to decompose volatility into discontinuous jumps and continuous diffusion components, and choosing the appropriate parameters of micro-structure noise robust estimators is not an exact science, we use realized volatility of day d defined as

$$RV_d = \sqrt{\sum_{\tau} r_{d,\tau}^2}$$

based on returns observed at a five-minute frequency. This frequency is commonly used in the literature (Andersen & Bollerslev, 1998) as it provides sufficient precision, and the micro-structure noise is negligible in the case of highly liquid stocks. For a detailed review of realized volatility, see Andersen and Teräsvirta (2009).

3.3 Multiplicative EGARCHK

3.3.1 H-Distribution

This section aims to introduce a novel distribution that can be leptokurtic and platykurtic, in which a single parameter drives the shape. These properties can be achieved using Tukey's H-transformation (Tukey, 1977), defined as:

$$H_h(z) \equiv z \exp\left[\frac{hz^2}{2}\right], \quad z, h \in \mathbb{R}.$$

Let X be the standard normal distribution. Then the random variable

$$Y = H_h(X)$$

for $h \in (0, \frac{1}{4})$ is symmetric around zero and leptokurtic. It follows from symmetricity, convexity, and concavity of the transform. We construct the distribution by transforming the standard normal distribution for convenience, since standard normal distribution is symmetric around zero. For $h = 0$, the random variable Y reduces to standard normal. The transformation can then be seen as increasing kurtosis with respect to the normal distribution.

The H-transformation is not a good candidate for achieving platykurtic shapes. Negative values of h give platykurtic shapes but impose an undesired bound on the support; see Headrick et al. (2008). The H-transformation for negative h is also not monotonic. Consequently, the inversion does not exist, and distribution is less analytically tractable. Therefore, we propose another transformation based on the H-transformation for achieving platykurtic shapes, which is the inverse of the H-transformation achieving leptokurtic shapes. We define this new transformation

as

$$K_h(z) = \begin{cases} z \exp\left(\frac{hz^2}{2}\right) & \text{for } h > 0 \\ z & \text{for } h = 0 \,, \\ \text{sign}(z)\sqrt{\frac{W(z^2|h|)}{|h|}} & \text{for } h < 0 \end{cases} \tag{3.3.1}$$

where $W()$ is the Lambert W function and $\text{sign}(z)$ is a sign of z and $K_h(z)$.

Proposition 3.3.0.1. The inversion of (3.3.1) is given by

$$K_h^{-1}(z) = \begin{cases} \text{sign}(z)\sqrt{\frac{W(z^2|h|)}{|h|}} & \text{for } h > 0 \\ z & \text{for } h = 0 \,. \\ z \exp\left(\frac{|h|z^2}{2}\right) & \text{for } h < 0 \end{cases} \tag{3.3.2}$$

Proof. In Appendix C.2. □

In order to use MLE, we need to derive the conditional density of returns, which requires the density of the standard normal variable transformed by the K transformation. The following proposition 3.3.0.2 describes the distribution of a transformed variable.

Proposition 3.3.0.2. Let X be continuous and symmetrically distributed around zero with the density $f(x)$ and the cumulative distribution function $F(x)$, and $g_{K_h}(y)$, $G_{K_h}(y)$ and y_p being the respective density, cumulative distribution func-

tion, and $p \in (0,1)$ quantile of $Y = K_h(X)$. Then:

$$g_{K_h}(y) = \begin{cases} f\left(\sqrt{\frac{W(y^2h)}{h}}\right)\sqrt{\frac{1}{W(y^2h)}}\frac{|y|\sqrt{h}}{y^2h+\exp(W(y^2h))} & \text{if } h > 0 \\ f(y) & \text{if } h = 0 \\ f\left(y\exp\left(\frac{y^2|h|}{2}\right)\right)\left[\exp\left(\frac{y^2|h|}{2}\right) + y^2\exp\left(\frac{y^2|h|}{2}\right)|h|\right] & \text{if } h < 0 \end{cases}, \tag{3.3.3}$$

$$G_{K_h}(y) = \begin{cases} F\left(\text{sign}(y)\sqrt{\frac{W(y^2h)}{h}}\right) & \text{if } h > 0 \\ F(y) & \text{if } h = 0 \\ F\left(y\exp\left(\frac{y^2|h|}{2}\right)\right) & \text{if } h < 0 \end{cases}, \tag{3.3.4}$$

$$y_p = \begin{cases} F^{-1}(p)\exp\left(\frac{1}{2}h\left(F^{-1}(p)\right)^2\right) & \text{if } h \geq 0,\ p \geq 0.5 \\ \sqrt{\frac{1}{|h|}W\left(\left(F^{-1}(p)\right)^2|h|\right)} & \text{if } h < 0,\ p \geq 0.5 \\ -F^{-1}(p)\exp\left(\frac{1}{2}h\left(F^{-1}(p)\right)^2\right) & \text{if } h \geq 0,\ p < 0.5 \\ -\sqrt{\frac{1}{|h|}W\left(\left(F^{-1}(p)\right)^2|h|\right)} & \text{if } h < 0,\ p < 0.5 \end{cases}. \tag{3.3.5}$$

Proof. In Appendix C.2. □

Moreover, assuming a standard normal distribution of X and $h \geq 0$, the moments coincide with moments derived by Headrick et al. (2008). In particular,

$$\mathbb{E}[K_h(X)^2] = \frac{1}{(1-2h)^{\frac{3}{2}}}, \quad \text{if} \quad h \in \left(0, \frac{1}{2}\right),$$
$$\mathbb{E}[K_h(X)^4] = \frac{3}{(1-4h)^{\frac{5}{2}}}, \quad \text{if} \quad h \in \left(0, \frac{1}{4}\right),$$

which yields kurtosis

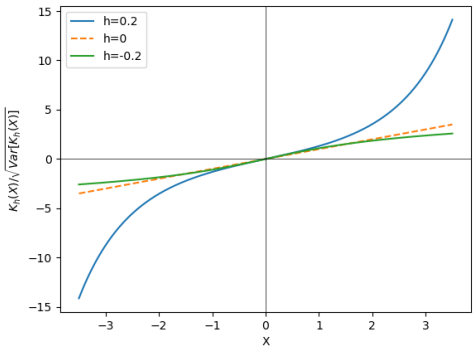
$$\kappa = \mathbb{E}\left[\frac{K_h(X)^4}{\mathbb{E}[K_h(X)^2]^2}\right] = \frac{3\left(1-2h\right)^{\frac{3}{2}}}{(1-4h)^{\frac{5}{2}}} \quad \text{if} \quad h \in \left(0, \frac{1}{4}\right). \tag{3.3.6}$$

Unfortunately, the even moments for negative h do not have closed-form solutions, and must be computed numerically. We pre-compute variances and kurtosis for a fine h grid $[-0.249, -0.248, ..., -0.001, 0.0]$ by Monte Carlo simulation with a sample size of 100,000. Fortunately, the moments have a small variance for $h < 0$, and the simulation error is negligible even for smaller sample sizes. The distribution is well defined even for $h < -0.249$, although it is extremely unlikely to obtain such values of h.

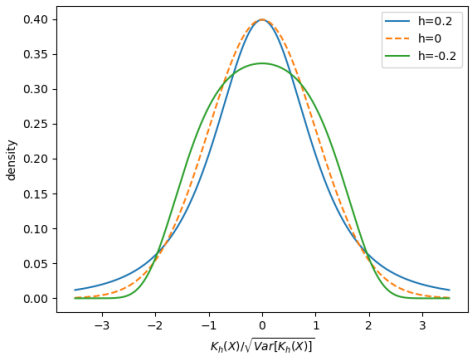
Hereafter, we address a distribution constructed by transforming standard normal distribution by (3.3.1) as the H-distribution[2]. To gain a better idea of possible shapes produced by the proposed transformation, a standardized transformation, i.e., preserving unit variance, of standard normal distribution and its density is shown in Figure 3.1. The dashed line represents the original standard normal distribution, while green is platykurtic, and blue is leptokurtic. An additional comparison with the t-distribution and GED can be found in Appendix C.1. An important property of the proposed distribution is the fact that the transformation has an almost negligible impact on values in the interval $(-1, 1)$, i.e., the center of the distribution, while values at the tails are inflated or shrunk by the h parameter. Therefore, the h parameter represents a tail risk, unlike the degrees of freedom of GED, which also change the center of the distribution.

Figure 3.1 Figures of a standardized K_h transformation of a standard normal random variable and associated probability density function.

(a) Standardized transformation



(b) Density

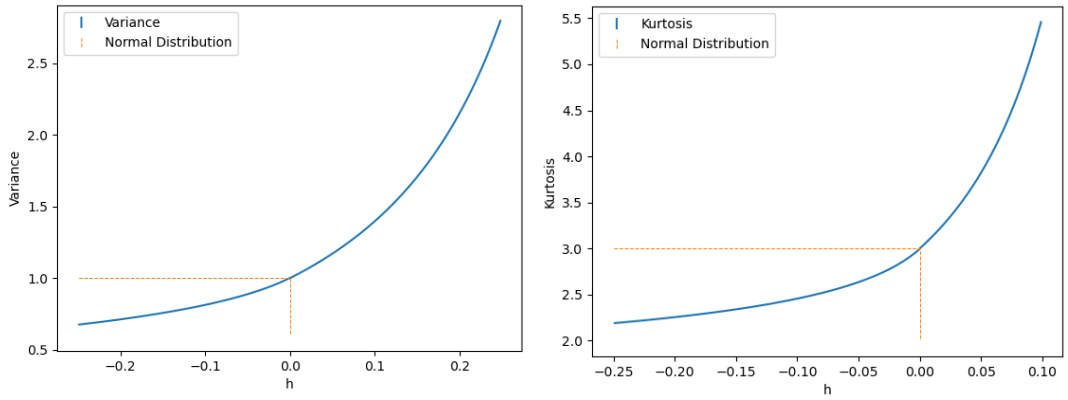


The first four moments of the Tukey transformed standard normal variable exist and are finite if $h < \frac{1}{4}$. The plot of moments for various h is shown in Figure 3.2.

[2]The standard normal random variable transformed by the H-transformation is typically called h-distribution or the more general g-and-h distribution. Because the K-distribution is already a reserved term, we use a capital H to distinguish the original Tukey distribution from our proposed distribution.

Figure 3.2 Figures of H-distribution moments.

(a) Variance **(b)** Kurtosis



Given that we directly model the dynamics of kurtosis, we need to express the shape parameter h as a function of kurtosis. An analytical solution does not exist; therefore, for a given kurtosis κ, we solve numerically for the corresponding h using the Newton-Raphson algorithm.

3.3.2 Student's T-Distribution and Generalized Error Distribution

To assess performance of the proposed H-distribution, we compare models based on the H-distribution against the Student's t-distribution and GED, which are commonly used in modeling conditional distributions of daily returns.

The random variable X following the Student's t-distribution with v degrees of freedom has the probability density function

$$g_{t_v}(x,v) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{\pi v}\Gamma(\frac{v}{2})}\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}}. \tag{3.3.7}$$

The variance and kurtosis are given as

$$Var[X] = \frac{v}{v-2}, \tag{3.3.8}$$

$$\mathbb{E}\left[\frac{X^4}{Var[X]^2}\right] = \frac{6}{v-4}+3. \tag{3.3.9}$$

The random variable X following the GED has the probability density function

$$g_{GED}(x,\nu) = \frac{\nu}{2\Gamma(1/\nu)}e^{-|x|^\nu}, \tag{3.3.10}$$

with variance and kurtosis given as

$$Var[X] = \frac{\Gamma(3/\nu)}{\Gamma(1/\nu)}, \tag{3.3.11}$$

$$\mathbb{E}\left[\frac{X^4}{Var[X]^2}\right] = \frac{\Gamma(5/\nu)\Gamma(1/\nu)}{\Gamma(1/\nu)^2}. \tag{3.3.12}$$

As with the H-distribution, we need to express shape parameters as a function of kurtosis. In the case of the t-distribution, the degrees of freedom v can be analytically expressed from (3.3.9). In the case of GED, an analytical formula does not exist. For given kurtosis, we numerically solve (3.3.12) for shape parameter ν with the Newton-Raphson algorithm.

3.3.3 Model Specification

Inspired by Engle and Sokalska (2012) and Bekierman and Gribisch (2021), who document the potentially long memory of intraday volatility, we specify a model with two frequencies of updating volatility, daily and intraday. Let $r_{d,\tau}$ be intraday returns, then

$$r_{d,\tau} = \mu S_\tau + \sigma_d \sigma_{d,\tau} S_\tau z_{d,\tau}, \tag{3.3.13}$$

where σ_d, $\sigma_{d,\tau}$, S_τ, and $z_{d,\tau}$ are daily volatility, intraday volatility, diurnal factor, and standardized innovations, respectively.

The intraday asset returns exhibit diurnal patterns in volatility with a well-known U-shaped pattern, in which volatility is typically high in the morning, low at noon, and again high near market closing. We capture the diurnal patterns in volatility by factors S_1, S_2, ..., $S_{\tau_{max}}$. Using diurnal factors allows us to estimate the model in two steps, achieving lower numerical complexity of the overall estimation. The first step estimates the factors and adjusts returns. In the second step, we apply MLE with the adjusted returns. It is common practice in modeling intraday returns to use diurnal factors and a two-step estimation (Engle & Sokalska, 2012; Feng & McNeil, 2008).

An alternative approach involves modeling periodic behavior using Fourier or other approximation series (Andersen & Bollerslev, 1997b; Bollerslev & Ghysels, 1996; Bordignon et al., 2007). One advantage of this approach over diurnal factors

is its potentially higher efficiency, though there is a risk of approximation error. Because our final analysis is performed with 100 stocks, we use the more robust and less efficient diurnal factors.

Engle and Sokalska (2012) do not specify dynamics of daily updated volatility. Instead, they used commercially available external daily volatility forecasts. Given that we do not have available daily commercial volatility forecasts for individual stocks, for daily volatility σ_d, we use an EGARCH-like specification based on realized volatility as

$$\log \sigma_d = \beta_1 \log \sigma_{d-1} + \beta_2 \left[\log RV_{d-1} - \mathbb{E}[\log RV_d]\right], \tag{3.3.14}$$

where $\log RV_d$ is the log of realized volatility and $\mathbb{E}[\log RV_d]$ is the long-run mean of the log of realized volatility. The intraday volatility follows the asymmetric EGARCH process:

$$\log \sigma_{d,\tau} = \alpha_0 + \alpha_1 \log \sigma_{d,\tau-1} + \alpha_2 z_{d,\tau-1} + \alpha_3 |z_{d,\tau-1}|. \tag{3.3.15}$$

For details of EGARCH, see Nelson (1991). Note that 3.3.15 is defined only for $\tau > 1$. To filter the first value of the day, we use the last observation of the previous day, that is, $d-1, \tau_{max}$, when $\tau = 1$.

The daily volatility (3.3.14) does not include an intercept, since the intraday volatility (3.3.15) already includes α_0 and the intercept would not be identified. Moreover, demeaning of realized log volatility is not necessary, as $-\beta_2 \mathbb{E}[\log RV_d]$ would be captured by α_0 for a model specification without the demeaning. We perform demeaning to improve interpretability, as σ_d then better represents the relative deviation from long-run volatility.

Conditional kurtosis is captured by the time-varying shape parameters of the H-distribution, Student's t-distribution, and GED. The variance of all three distributions depends on the shape parameter. Following Brooks et al. (2005), we allow for independent movement of kurtosis and variance by enforcing the unit variance of $z_{d,\tau}$ as

$$z_{d,\tau} = \frac{\epsilon_{d,\tau}}{\sqrt{\mathbb{E}[\epsilon_{d,\tau}^2 | I_{d,\tau-1}]}}, \tag{3.3.16}$$

where $\epsilon_{d,\tau}$ follows the H-distribution, Student's t-distribution, or GED, and $I_{d,\tau-1}$ is all the information available at time $d, \tau - 1$. Intraday returns have high

persistence of conditional variance and potentially long memory compared to daily returns, as documented by Engle and Sokalska (2012). Therefore, the kurtosis process in a model without standardization could potentially capture variance rather than kurtosis.

Following Brooks et al. (2005), we model the process of conditional kurtosis and then map it to the corresponding shape parameters. The H-distribution and GED do not have an analytical solution for mapping kurtosis to the shape parameter, meaning that the corresponding shape parameters have to be found numerically. The Student's t-distribution has an analytical formula. Let $\kappa_{d,\tau}$ be conditional kurtosis on day d and time τ, then conditional degrees of freedom of the Student's distribution are:

$$v_{d,\tau} = \frac{2(2\kappa_{d,\tau} - 3)}{\kappa_{d,\tau} - 3}. \tag{3.3.17}$$

Finally, we need to specify the dynamics of conditional kurtosis, and experiment with various specifications and news impact curves. The idea of the news impact curve (NIC) was introduced by Engle and Ng (1993) as a function describing the relationship between past (potentially standardized) returns and conditional volatility. Jondeau and Rockinger (2006) extend the idea to conditional skewness and kurtosis. Anatolyev and Petukhov (2016) study the NIC of conditional skewness with flexible parametric and non-parametric specifications. Later, Anatolyev (2021) researched the NIC for the conditional probability of returns exceeding a certain threshold. All research on the NIC has been conducted in the context of daily returns, and it is unclear whether the results translate to intraday returns. Therefore, we explore the NIC in the empirical section of our work.

Preliminary analysis has shown that specifications inspired by literature using daily returns easily leads to numerically explosive kurtosis when used for intraday conditional distributions. This is caused by relatively extreme sudden changes in volatility and the fact that GARCH-like models assume predetermined volatility based on past observations. For example, there can be a period of several days with very little volatility, during which the filtered intraday volatility slowly drops to values close to 0. Under such conditions, if new shocking information appears, even though the absolute price movement is not extreme relative to the long-term volatility, it holds extreme value relative to the conditional short-term volatility.

This is not a phenomenon unique to conditional kurtosis of intraday returns. In the literature on modelling intraday volatility, it is common to winsorize volatility by replacing extreme values with unconditional quantiles. Typically, any values outside the 0.5% - 99.5% percentile range are set to the 0.5% and 99.5% percentiles (Orlowski et al., 2024; Zhang et al., 2024). In our case, we should not perform winsorization, as the goal of our work is to capture rare events, and setting a cut-off for a quantile could lose more information on the magnitude of the extreme shock than desired. Instead, we limit the values of $z_{d,\tau}$ to the interval $[-6, 6]$ as:

$$B(z) = \min(\max(z, -6), 6). \tag{3.3.18}$$

This eliminates numerically extreme values, while retaining more unaltered observations compared to winsorization. For context, if the standardized shocks were normally distributed with unit variance, the probability of observing a value outside the interval $[-6, 6]$ would be 2×10^{-9}.

In order to capture conditional kurtosis, we consider the following specifications:

1. Constant kurtosis:

$$\kappa_{d,\tau} = \omega_0. \tag{3.3.19}$$

2. Quartic specification:

$$\kappa_{d,\tau} = \omega_0 + \omega_1 \kappa_{d,\tau-1} + \omega_2 B(z_{d,\tau-1})^4. \tag{3.3.20}$$

3. Quadratic-quartic specification:

$$\kappa_{d,\tau} = \omega_0 + \omega_1 \kappa_{d,\tau-1} + \omega_2 B(z_{d,\tau-1})^4 + \omega_3 B(z_{d,\tau-1})^2. \tag{3.3.21}$$

4. Linear log-kurtosis:

$$\log \kappa_{d,\tau} = \omega_0^{\log} + \omega_1^{\log} \log \kappa_{d,\tau-1} + \omega_4^{\log} \Big| B(z_{d,\tau-1}) \Big|. \tag{3.3.22}$$

5. Piece-wise linear log-kurtosis:

$$\log \kappa_{d,\tau} = \omega_0^{\log} + \omega_1^{\log} \log \kappa_{d,\tau-1} + \omega_4^{\log} \Big| B(z_{d,\tau-1}) \Big| + \omega_5^{\log} \left(\Big| B(z_{d,\tau-1}) \Big| - b \right)^+, \tag{3.3.23}$$

where $x^+ = \max(0, x)$, and $b \geq 0$.

The quartic specification is common in the literature (Brooks et al., 2005; León et al., 2005). In the quadratic-quartic specification, we add the quadratic term to allow non-monotonic NIC and the different impact of less shocking—$z \in (-1, 1)$ in this case—and more shocking returns. The advantage of log-kurtosis is that, similarly to E-GARCH, it represents a multiplicative specification. Unlike kurtosis specifications, log-kurtosis specifications do not require powers to be chosen, and the curvature of the NIC is represented by parameters. The piece-wise linear log-kurtosis specification has the same motivation as the quadratic-quartic specification; to allow potentially different reactions to more and less shocking returns. In this case, the interval $(-b, b)$—at which standardized returns are considered less shocking—is the parameter to be estimated.

In addition to the five specifications stated above, we experiment with direct specifications for parameters as used by Jondeau and Rockinger (2003), albeit unsuccessfully. The estimation procedure either fails to identify meaningful parameters or the time-varying kurtosis degenerates to a constant kurtosis. We also experiment with a piece-wise linear log-kurtosis with more than two intervals. However, preliminary analysis shows poor out-of-sample performance, suggesting an overfitting and identification problem. These poorly performing specifications are not presented in this paper, although we consider it important to mention that we were unsuccessful in applying them to intraday returns.

3.3.4 Estimation

The model is estimated in two steps. The first step estimates diurnal factors $\mathbf{S} = [S_1, S_2, ..., S_{\tau_{max}}]^T$ and the long-run mean of realized log-volatility $\mu_{RV} = \mathbb{E}[\log RV_d]$ by a moment estimator. The moment vector for day d is:

$$m_d^1(\mu_{RV}, \mathbf{S}) = \begin{pmatrix} \mu_{RV} - \log RV_d \\ (S_1^2 - r_{d,1}^2) \\ (S_2^2 - r_{d,2}^2) \\ \vdots \\ (S_{\tau_{max}}^2 - r_{d,\tau_{max}}^2) \end{pmatrix}, \tag{3.3.24}$$

resulting in moment conditions $\mathbb{E}[m_d^1(\mu_{RV}, S)] = 0$.

Using the estimates of diurnal factors $\widehat{S_\tau}$, we calculate the series adjusted for a diurnal pattern in scale as:

$$s_{d,\tau} = \frac{r_{d,\tau}}{\widehat{S_\tau}} \xrightarrow[D\to\infty]{P} \mu + \sigma_d \sigma_{d,\tau} z_{d,\tau}. \tag{3.3.25}$$

The log-likelihood of $s_{d,\tau} = r_{d,\tau}/S_\tau$ conditioned on the past for shocks $z_{d,\tau}$ with H-distribution, Student's t-distribution, or GED is, respectively:

$$l_{d,\tau}(\theta|\mu_{RV}, \mathbf{S}) = \log\left(g_{K_h}\left(\frac{r_{d,\tau}/S_\tau - \mu}{\sigma_d\sigma_{d,\tau}}, h_{d,\tau}\right)\frac{1}{\sigma_d\sigma_{d,\tau}}\right), \tag{3.3.26}$$

$$l_{d,\tau}(\theta|\mu_{RV}, \mathbf{S}) = \log\left(g_{t_v}\left(\frac{r_{d,\tau}/S_\tau - \mu}{\sigma_d\sigma_{d,\tau}}, v_{d,\tau}\right)\frac{1}{\sigma_d\sigma_{d,\tau}}\right), \tag{3.3.27}$$

$$l_{d,\tau}(\theta|\mu_{RV}, \mathbf{S}) = \log\left(g_{GED}\left(\frac{r_{d,\tau}/S_\tau - \mu}{\sigma_d\sigma_{d,\tau}}, \nu_{d,\tau}\right)\frac{1}{\sigma_d\sigma_{d,\tau}}\right), \tag{3.3.28}$$

where θ is a vector of parameters that were not estimated in the first step. The second step estimation moments for the remaining parameters are then:

$$m_d^2(\theta, \widehat{\mu_{RV}}, \widehat{\mathbf{S}}) = \frac{1}{\tau_{max}}\sum_\tau \nabla_\theta l_{d,\tau}(\theta|\widehat{\mu_{RV}}, \widehat{\mathbf{S}}). \tag{3.3.29}$$

Assuming that the vector process of intraday returns and realized log-variance is ergodic and stationary at a daily frequency, we can use the generalized method of moments (GMM) framework of L. P. Hansen (1982). By "stacking" the moment vectors of both steps, following Newey and McFadden (1994), as

$$m_d(\mu_{RV}, \mathbf{S}, \theta) = \begin{pmatrix} m_d^1(\mu_{RV}, \mathbf{S}) \\ m_d^2(\theta, \mu_{RV}, \mathbf{S}) \end{pmatrix}, \tag{3.3.30}$$

we can easily find the asymptotic distribution of the estimator given by moment conditions $\mathbb{E}[m_d(\mu_{RV,0}, \mathbf{S}_0, \theta_0)] = 0$, where $\mu_{RV,0}$, $\mathbf{S}_0$, and θ_0 are the true values of the parameters. Engle and Sokalska (2012) provide a brief discussion. The asymptotic distribution of parameter estimates is

$$\sqrt{D}\begin{pmatrix} \widehat{\mu}_{RV} - \mu_{RV,0} \\ \widehat{\mathbf{S}} - \mathbf{S}_0 \\ \widehat{\theta} - \theta_0 \end{pmatrix} \xrightarrow{d} N\left(0, \left(M^{-1}\right)^T V M^{-1}\right), \tag{3.3.31}$$

where

$$\begin{aligned} V =& \mathbb{E}[m_d(\mu_{RV,0}, \mathbf{S}_0, \theta_0) m_d(\mu_{RV,0}, \mathbf{S}_0, \theta_0)^T] \\ &+ 2\sum_{n=1}^{\infty} \mathbb{E}[m_d(\mu_{RV,0}, \mathbf{S}_0, \theta_0) m_{d-n}(\mu_{RV,0}, \mathbf{S}_0, \theta_0)^T], \end{aligned} \tag{3.3.32}$$

and

$$M = \mathbb{E}\left[\frac{\partial m_d(\mu_{RV,0}, \mathbf{S}_0, \theta_0)}{\partial [\mu_{RV,0}, \mathbf{S}_0, \theta_0]^T,}\right]. \tag{3.3.33}$$

The consistency and normality of the two-step estimator is a corollary to Theorem 6.1 in Newey and McFadden (1994), relying, in addition to ergodicity and stationarity stated above, on assumptions (i) $\mu_{RV,0}, \mathbf{S}_0, \theta_0$ being from the interior of parametric space, (ii) $m_d(\mu_{RV}, S, \theta)$ is continuously differentiable in a neighborhood $\mathcal{N}$ of $[\mu_{RV,0}, \mathbf{S}_0, \theta_0]^T$, (iii) $\mathbb{E}[m_d(\mu_{RV,0}, \mathbf{S}_0, \theta_0)] = 0$ and $\mathbb{E}\left[||m_d(\mu_{RV,0}, \mathbf{S}_0, \theta_0)||\right] < \infty$, (iv) $\mathbb{E}[\sup_{[\mu_{RV}, \mathbf{S}, \theta]^T \in \mathcal{N}} ||\nabla_{[\mu_{RV,0}, \mathbf{S}_0, \theta_0]^T} m_d(\mu_{RV}, \mathbf{S}, \theta)||] < \infty$, and (v) non-singular hessian M.

Treating $\log RV_d$ as an external predictor, similarly to Engle and Sokalska (2012), and assuming common parametric restrictions for volatility and kurtosis processes guaranteeing stationarity, non-negativity of volatilities and kurtosis from the range attainable by the used distribution, the conditions (i) - (v) are expected to hold.

A potential violation of the assumption may arise from model distribution misspecification. To be more specific, if the conditional distribution of $z_{d,\tau}$ is not leptokurtic, the condition (i) will not be satisfied for t-distribution. This concern is addressed by using H-distribution and GED. Another concern is the autoregressive coefficients of the volatility processes. As we do not explicitly describe the full data-generating process, the relationship between intraday returns and daily realized log-variances remains unclear, and the conditions cannot be formulated explicitly in terms of model parameters. The processes $\sigma_{d,\tau}$ and σ_d could potentially be non-stationary for a combination of high but not necessarily unit values of α_1 and β_1. Specifying a full description of the data-generating process would require a substantially more complex model and is out of the scope of our work. However, if this was indeed a problem, it would lead to numerical instability of the filtered volatility series. Empirically, we did not register any instability in volatility filtering.

The variance-covariance matrix V can be estimated using the heteroskedasticity- and autocorrelation-consistent (HAC) estimator (L. P. Hansen & Hodrick, 1980; Newey & West, 1986). However, it is important to highlight that Engle and Sokalska (2012) ignore the potential serial correlation of

the moment estimator of the diurnal factors. By construction, the population conditional score functions are independent, and the sources of serial correlation are moment conditions of diurnal factors and—in our case—the long-run mean of realized log-variance estimators. The autocorrelations of squared returns of the same intraday period on different days are generally weak. Therefore, ignoring serial correlation in moment conditions for diurnal factors may be justified. The same argument cannot be used for daily realized log-variances that are highly correlated.

3.3.5 Conditional Density Forecasting

We assess the importance of time-varying kurtosis based on the quality of distributional fit in- and out-of-sample. For the in-sample assessment, we use the likelihood ratio test, comparing each specification with time-varying kurtosis to the specification with constant kurtosis.

For out-of-sample evaluation, we construct a model confidence set (MCS) developed by P. R. Hansen et al. (2011). MCS is an analogous idea to a confidence interval. It is a set of models constructed such that it will contain the best model with a given level of confidence. It is an iterative procedure, where in each step all models within a set of potential models are tested for equivalent forecasting performance. If the hypothesis of equivalent forecasting performance is rejected, the worst-performing model is eliminated, and the procedure continues until it is no longer possible to reject the equivalence.

To apply MCS, we need a loss function that measures forecasting performance. We take inspiration from Anatolyev and Petukhov (2016), who apply a test developed by Amisano and Giacomini (2007) for comparing the forecasted log densities of two models. We use one-period ahead conditional log-likelihoods as a loss function for MCS. Let $l^{\xi}_{d,\tau}$ be the conditional log-likelihood of model ξ, the equivalence test hypothesis is

$$\mathbb{E}[l^{\xi}_{d,\tau} - l^{\eta}_{d,\tau}] = 0 \text{ for all } \xi, \eta \in \mathcal{M}, \tag{3.3.34}$$

where $\mathcal{M}$ is a set of models. Following P. R. Hansen et al. (2011), we use a t-ratio

for a pair-wise comparison of ξ and η:

$$dm_{\eta,\xi} = \frac{\sqrt{D\tau_{max}}\sum_{d,\tau}(l^{\xi}_{d,\tau} - l^{\eta}_{d,\tau})}{\sqrt{\widehat{Var[l^{\xi}_{d,\tau} - l^{\eta}_{d,\tau}]}}}, \tag{3.3.35}$$

where $\widehat{Var[l^{\xi}_{d,\tau} - l^{\eta}_{d,\tau}]}$ is the HAC estimator of the long-run variance of log-likelihood differences. Amisano and Giacomini (2007) show that the test statistic $dm_{\xi,\eta}$ has an asymptotic $N(0,1)$ distribution under a null hypothesis that the forecasts of the two models are equivalent[3]. The null hypothesis (3.3.34) of all models in $\mathcal{M}$ being equivalent naturally maps to a test statistic:

$$DM = \max_{\xi,\eta\in\mathcal{M}} |dm_{\eta,\xi}|. \tag{3.3.36}$$

The asymptotic distribution of DM is nonstandard with nuisance parameters for the dependence of the t-ratios. We use the bootstrap procedure described by P. R. Hansen et al. (2011). As a model elimination rule, we remove the model with the lowest out-of-sample log-likelihood $\sum_{d,\tau} l^{\xi}_{d,\tau}$.

For a given significance α, the equivalence test (3.3.34) and the model elimination are iteratively repeated until no model can be removed. The resulting set is $1-\alpha$ MCS. For convenience, we report MCS p-values. If the MCS p-value of a model is smaller than α, the model does not belong to the $1-\alpha$ MCS.

Practitioners may potentially be more interested in the forecasting performance of conditional VaR and ES. Du and Escanciano (2017) propose tests to assess the performance of forecasts in terms of VaR and ES. We do not report VaR or ES forecast precision, as the improvement in distributional forecast precision arising from conditional kurtosis is too small to be measured by improvement in VaR or ES forecasts.

[3]The test is suitable for comparing nested and non-nested models, when based on out-of-sample forecasts with a fixed estimation sample size. The asymptotic distributions of the in-sample test statistics are different for nested and non-nested models. Additionally, the rate of convergence for comparison of nested and non-nested models is different. This makes in-sample MCS applications with a log-likelihood as a loss function substantially more challenging. Therefore, we do not apply MCS to assess in-sample performance. For further discussion, see P. R. Hansen et al. (2011).

3.4 Data

Our empirical analysis of conditional kurtosis and the study of extreme shock spillovers is based on 100 highly liquid US traded stocks. The entire data set we use covers the period from October 29, 2015, to November 8, 2019. The models described are numerically challenging to estimate at a five-minute frequency. Therefore, we divide the dataset in two and mostly limit our empirical analysis to the period from November 1, 2017, to November 8, 2019. The first period from October 29, 2015, to October 31, 2017, is only used for estimation when we assess out-of-sample performance during the period from November 1, 2017, to November 8, 2019. The analyzed dataset includes 100 highly liquid US-listed stocks with the following symbols: AA, AAPL, ABEV, ADBE, AEO, AGN, AGNC, AMAT, AMD, AMGN, AMZN, AU, AUY, AVGO, AZN, BA, BABA, BAC, BHC, BHP, BIDU, BIIB, BP, BRK.B, BUD, C, CAT, CCL, CELG, CHK, CLF, COST, CRM, CSCO, CTL, CVX, CY, DB, DIS, ECA, ET, EXPE, F, FB, FCAU, FCX, FDX, FEYE, GE, GFI, GILD, GM, GOLD, GOOG, GOOGL, GS, GSK, HBAN, HD, HL, HON, HPE, HPQ, HST, IBM, INTC, ITUB, JD, JNJ, JPM, KEY, KMI, LRCX, M, MA, MCD, MMM, MRO, MSFT, MT, MU, NBR, NFLX, NKE, NOK, NVDA, NYCB, ODP, PANW, PBCT, PBR, PEP, PFE, PYPL, QCOM, RDS.A, RDS.B, RF, RIG, and RIO. The frequency of the original dataset is one minute and includes open, high, low, and close prices together with traded volume. We transform the data to the five/minute log returns, defined by (3.2.3). We use only data from normal trading hours from 9:30 a.m. to 4:00 p.m.

To better understand the liquidity of the stocks considered, we calculate the average traded amount for each stock over one minute during normal trading hours. Table 3.1 shows the percentiles, minimum, and maximum of averages calculated from the 100 stocks. The stock with the highest average traded amount in one minute is Apple with $11,367,352. The least traded stock in our dataset is ODP Corporation, with an average traded amount in one minute of $31,704.

In our analysis, we focus extensively on ODP Corporation (ODP), FedEx (FDX), Apple (AAPL), and JPMorgan Chase & Co (JPM). ODP, FDX, and AAPL represent stocks with the lowest, median, and highest average traded amount, respectively, within our dataset. JPM is added to represent the financial sector,

Table 3.1 Traded amounts.

Minimum	25% percentile	50% percentile	75% percentile	Maximum
\$31,704	\$309,658	\$730,180	\$1,389,365	\$11,367,352

Notes: Summary statistics of one-minute traded amount averages of individual stocks during normal trading hours. Averages are calculated from November 1, 2017, to November 8, 2019 for the 100 stocks listed above.

and with its average one-minute traded value of \$2,884,768 it is among the stocks with a higher traded amount in our dataset.

3.5 Conditional Kurtosis of Intraday Returns

3.5.1 H-distribution Results

The first question to address concerns the existence of time-varying kurtosis. Table 3.2 provides a summary of the five specifications for the dynamics of kurtosis under consideration. Table 3.3 shows estimates of all five specifications with the H-distribution.

Table 3.2 Specifications of kurtosis dynamics.

1	$\kappa_{d,\tau} = \omega_0$
2	$\kappa_{d,\tau} = \omega_0 + \omega_1 \kappa_{d,\tau-1} + \omega_2 B(z_{d,\tau-1})^4$
3	$\kappa_{d,\tau} = \omega_0 + \omega_1 \kappa_{d,\tau-1} + \omega_2 B(z_{d,\tau-1})^4 + \omega_3 B(z_{d,\tau-1})^2$
4	$\log \kappa_{d,\tau} = \omega_0^{\log} + \omega_1^{\log} \log \kappa_{d,\tau-1} + \omega_4^{\log} \left\lvert B(z_{d,\tau-1}) \right\rvert$
5	$\log \kappa_{d,\tau} = \omega_0^{\log} + \omega_1^{\log} \log \kappa_{d,\tau-1} + \omega_4^{\log} \left\lvert B(z_{d,\tau-1}) \right\rvert + \omega_5^{\log} \left(\left\lvert B(z_{d,\tau-1}) \right\rvert - b \right)^+$

Models 2, 3, and 5 significantly outperform the model with constant kurtosis in terms of conditional forecasting, both in- and out-of-sample. This confirms that conditional kurtosis exists and is relevant for forecasting. Although the forecast improvement is statistically significant, the economic significance is rather limited. The two-year period with 78 observations per day yields 39,234 observations. Therefore, much higher values of statistics would be expected if the forecasts were improved for most of the observations. The time-varying kurtosis only improves

the forecast for very few observations, where there is a series of two or more extreme shocks within a short period. For most of the time, the conditional kurtosis is close to the long-term kurtosis.

The minor improvement can be explained by findings in the literature on realized measures and asset returns in a continuous time framework. It is typically assumed that price in continuous time follows a jump-diffusion process, with continuous trends and discontinuous jumps (Ait-Sahalia & Jacod, 2012; Ait-Sahalia et al., 2012). With increasing sampling frequency, the log-return process taken to the fourth power is dominated by discontinuous jumps (Amaya et al., 2015). While kurtosis in a daily setup is driven by strong trends within a day, kurtosis in high-frequency intraday returns is more likely to be driven by discontinuous jumps. According to Ait-Sahalia et al. (2024), large jumps are relatively rare; therefore, there may simply not be many opportunities to improve the forecasts.

NIC specifications 3 and 5 are aligned with two regimes of reaction, in which small shocks reduce and large shocks increase kurtosis. The coefficient ω_3 of the quadratic term in specification 3 and the linear coefficient $\omega_4^{\log}$ for the bounded standardized returns of specification 5 are negative. This suggests that small values of the standardized returns $z_{d,\tau}$ reduce kurtosis, while larger shocks increase kurtosis. The negative impact of small shocks and the positive impact of large shocks can potentially also explain the insignificant $\omega_4^{\log}$, because the shape of the NIC cannot be captured by the log-linear NIC.

We also estimate the five specifications for the JPM stock; results appear in Table C1 in Appendix C.3. The estimates are similar to those for AAPL, in which the main difference lies in the forecast performance. Although the piece-wise linear specification for log-kurtosis provides the best in-sample forecasts, it has the worst out-of-sample forecasts. This suggests that the piece-wise linear NIC might be over-fitting the data. Because the quartic specification seems to be the most robust, we proceed with it in the following subsection on spillovers of extreme risk.

Another relevant question for studying conditional kurtosis is the impact of various frequencies. Table C3 in Appendix reports estimates for 5-, 15-, and

Table 3.3 Estimation results of models with H-distribution for AAPL.

Specification (ξ) $\rightarrow$	1	2	3	4	5
μ	0.002 (0.003)	0.002 (0.003)	0.001 (0.003)	0.002 (0.003)	0.001 (0.003)
α_0	-0.085 (0.003)	-0.068 (0.003)	-0.075 (0.003)	-0.082 (0.004)	-0.094 (0.004)
α_1	0.972 (0.002)	0.978 (0.002)	0.976 (0.002)	0.973 (0.002)	0.973 (0.002)
α_2	-0.013 (0.001)	-0.012 (0.001)	-0.013 (0.001)	-0.013 (0.001)	-0.013 (0.001)
α_3	0.074 (0.002)	0.060 (0.003)	0.067 (0.003)	0.072 (0.003)	0.084 (0.004)
β_1	0.761 (0.034)	0.798 (0.032)	0.775 (0.033)	0.767 (0.034)	0.749 (0.035)
β_2	0.206 (0.027)	0.177 (0.026)	0.197 (0.027)	0.201 (0.027)	0.221 (0.028)
$\omega_0^{(\log)}$	4.523 (0.082)	1.240 (0.288)	1.651 (0.336)	0.847 (0.949)	0.265 (0.039)
$\omega_1^{(\log)}$		0.708 (0.066)	0.641 (0.073)	0.429 (0.630)	0.878 (0.021)
ω_2		0.008 (0.002)	0.014 (0.004)		
ω_3			-0.109 (0.031)		
$\omega_4^{\log}$				0.013 (0.011)	-0.114 (0.017)
$\omega_5^{\log}$					0.166 (0.022)
b					1.404 (0.088)
MCS p-values	0.005	0.342	0.502	0.006	1.0
$LR_{1,\xi}$ (p-val)	0.00 (1.00)	49.09 (0.00)	69.84 (0.00)	1.49 (0.475)	109.18 (0.00)

Notes: Estimates for the period from November 1, 2017 to November 8, 2019, five-minute returns. MCS performed for the period from November 1, 2017 to November 8, 2019 with an estimation period from October 29, 2015 to October 31, 2017. $\omega^{(\log)}$ indicates both kurtosis and log-kurtosis specifications. The likelihood ratios $LR_{1,\xi}$ compare specification ξ against the specification 1. Numbers in parentheses represent standard errors for estimates and the p-values for $LR_{1,\xi}$.

30-minute returns of the model with H-distribution and the quartic specification of conditional kurtosis. The persistence of conditional kurtosis increases with frequency, and the coefficient of the quartic term decreases. This means that the magnitude of extreme shocks decreases with rising frequency more slowly than the magnitude of common shocks.

We also explore potential diurnal patterns in kurtosis. Preliminary analysis based on adjusted returns suggests that there may potentially be diurnal patterns beyond the diurnal factors. However, the distributional diurnal patterns do not translate directly to kurtosis or the patterns are not sufficiently strong to be well identified in a two-year period.

3.5.2 Comparison of H-Distribution, T-Distribution, and GED

We compare the H-distribution with the t-distribution and GED in terms of AIC and BIC. The results for AAPL, JPM, FDX, and ODP with a quartic specification are shown in Table C2. In general, the H-distribution and t-distribution provide a strongly superior distributional fit compared to the GED. However, the t-distribution degenerates to a model with a constant kurtosis. Table C4 shows estimates of a quartic specification with H- and t-distribution. We explore the t-distribution with other specifications of conditional kurtosis, with little or no success in identifying conditional kurtosis. This suggests that the H-distribution might be better suited to identifying time-varying kurtosis.

Among our cited sources, the study conducted by Ergün and Jun (2010) is probably the closest to our work. The authors apply a model of Jondeau and Rockinger (2003) with a skewed t-distribution that does not allow platykurtic shapes of distribution and conclude that kurtosis does not vary over time. Table C4 shows an identical conclusion for the Student's t-distribution. Estimates of ω_1 and ω_2 of GARCHK-t are practically zero and insignificant, while ω_0 represents constant unconditional kurtosis. The ω_0 values of the GARCHK-t model are aligned with the GARCHK-H model with constant kurtosis.

Our motivation for introducing the H-distribution was to allow platykurtic shapes of distribution, based on the suspicion that enforced leptokurtic distribution

by t-distribution might be the reason for the lack of identification. However, our hypothesis was not confirmed, as the filtered conditional kurtosis is typically higher than 3. More specifically, for example, the lowest filtered kurtosis for AAPL is obtained by the piece-wise linear log-kurtosis specification at a value of 3.2. This is not necessarily proof that the conditional distribution cannot be platykurtic. The GARCHK-like models assume that kurtosis is determined by stock's own past. It is more reasonable in an economic sense to assume that high kurtosis is realized at the moment when the new shocking information that causes an extreme shock appears. With pre-determined kurtosis, the lowest possible kurtosis is typically driven by a single observation with an extreme shock that appeared after an extensive calm period. If we allow the kurtosis during the calm period to fall too low, the extreme shock will have an explosive negative log-likelihood. The model could be adapted to stochastic kurtosis in a similar fashion to models with stochastic volatility, as in the work of Nelson (1988) and Ruiz (1994). However, such an extension is beyond the scope of this work.

All specifications discussed have an exponential specification of intraday volatility. Nelson (1991) introduces exponential GARCH (E-GARCH) for daily returns and discusses the existence of moments under the assumption of shocks following normal distribution, t-distribution, or GED. When the shocks of E-GARCH are assumed to follow a t-distribution, the existence of unconditional kurtosis is not guaranteed. Empirically, values of parameters when the unconditional kurtosis does not exist are relatively common. This raises a concern that the performance of the t-distribution is driven by the distribution itself or its application with an exponential specification of intraday volatility. Our model does not fully specify the data-generating process because the daily volatility is driven by realized daily volatility, which is treated as an external predictor. Therefore, it is not clear what the relationship is between $z_{d,\tau}$ and σ_d. We take an inspiration for the specification from Engle and Sokalska (2012), where external volatility forecasts substantially simplify the model. To study the theoretical properties, it is necessary to introduce a more complex model that fully describes the relationship between intraday returns and daily realized volatility. A mixed-frequency extension of RealGARCH by P. R. Hansen et al. (2012) could be a potential candidate. Introducing such a

model is out of the scope of this chapter. Instead, to address the concern that results may be affected by the potential non-existence of unconditional kurtosis, we estimate models with an alternative volatility specification using a standard GARCH process for intraday volatility. The estimates for the t-distribution and H-distribution are reported in Appendix C.6. The conditional kurtosis of both specifications degenerates to a constant kurtosis model. Therefore, we can conclude that the lack of identification with exponential GARCHK-t is likely not related to the existence of the unconditional kurtosis. Another interesting observation is the lack of identification with the non-exponential GARCHK-H model. This shows that the H-distribution generally does not provide better identification or superior forecasts, but only empirically performs better than the t-distribution when paired with exponential GARCHK.

Table 3.4 Comparison of GARCHK model with H-distribution and Student's t-distribution.

	AAPL			JPM		
	H	H-constant	t	H	H-constant	t
μ	0.002	0.002	0.001	-0.007	-0.007	-0.001
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
α_0	-0.068	-0.085	-0.077	-0.060	-0.073	-0.067
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
α_1	0.978	0.972	0.974	0.978	0.971	0.973
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
α_2	-0.012	-0.013	-0.014	-0.010	-0.010	-0.013
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
α_3	0.060	0.074	0.094	0.056	0.067	0.082
	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)
β_1	0.798	0.761	0.775	0.832	0.797	0.841
	(0.032)	(0.034)	(0.033)	(0.030)	(0.031)	(0.028)
β_2	0.177	0.206	0.192	0.147	0.172	0.133
	(0.026)	(0.027)	(0.026)	(0.025)	(0.024)	(0.021)
ω_0	1.239	4.523	4.615	1.156	4.191	4.259
	(0.288)	(0.082)	(1.458)	(0.356)	(0.066)	(0.066)
ω_1	0.708	-	0.000	0.713	-	0.005
	(0.066)		(0.309)	(0.087)		(0.011)
ω_2	0.008	-	0.001	0.006	-	0.000
	(0.002)		(0.009)	(0.002)		(0.001)
LR: $\omega_1 = 0$, $\omega_2 = 0$	49.054		0.406	27.64		0.323

$$\log \sigma_d = \beta_1 \log \sigma_{d-1} + \beta_2 \left[\log RV_{d-1} - \mathbb{E}[\log RV_d]\right]$$

$$\log \sigma_{d,\tau} = \alpha_0 + \alpha_1 \log \sigma_{d,\tau-1} + \alpha_2 z_{d,\tau-1} + \alpha_3 |z_{d,\tau-1}|$$

$$\kappa_{d,\tau} = \omega_0 + \omega_1 \kappa_{d,\tau-1} + \omega_2 B \left(z_{d,\tau-1}\right)^4$$

Notes: Estimates for Apple and JP Morgan based on the period from November 1, 2017 to November 8, 2019, with five-minute returns. Standard errors are in parenthesis. The $\chi^2(1)$ critical values for the LR test are 5.991 and 9.210 for 5% and 1% levels, respectively.

3.6 Spillovers of Extreme Risk

3.6.1 Model with Market Shocks

To study information spillovers of extreme events, we construct a generated variable that serves as a proxy for new shocking information in the market. Consider a set of n stocks indexed $i \in \{1, 2, ..., n\}$ to be a representative set of stocks for a market. We use a model with time-constant kurtosis to filter out conditionally standardized series $z^i_{d,\tau}$. Because we want to separate information from the past of the modeled stock and the market past, we construct the proxy with the modeled stock left out. The proxy variable for the market shock for stock j is then:

$$MS^j_{d,\tau} = \frac{1}{n-1} \sum_{i\in\{1,2,...,n\},\, i\neq j} \left(B(z^i_{d,\tau})\right)^4. \tag{3.6.1}$$

An alternative construction could be:

$$MAS^j_{d,\tau} = \left(\frac{1}{n-1} \sum_{i\in\{1,2,...,n\},\, i\neq j} B(z^i_{d,\tau})\right)^4. \tag{3.6.2}$$

The key difference between (3.6.1) and (3.6.2) is what the extreme values of the two proxy measures represent. The large values of the first proxy are mostly driven by one or only a few stocks that have relatively large returns during the period d, τ, while the values of the second proxy represent the large movement of the whole market. Since we are interested in spillovers rather than major market crashes or surges, we use the first definition.

We add the constructed variable to the kurtosis process (3.3.20) as:

$$\kappa_{d,\tau} = \omega_0 + \omega_1\kappa_{d,\tau-1} + \omega_2 B\left(z_{d,\tau-1}\right)^4 + \omega^{MS} MS^j_{d,\tau-1}. \tag{3.6.3}$$

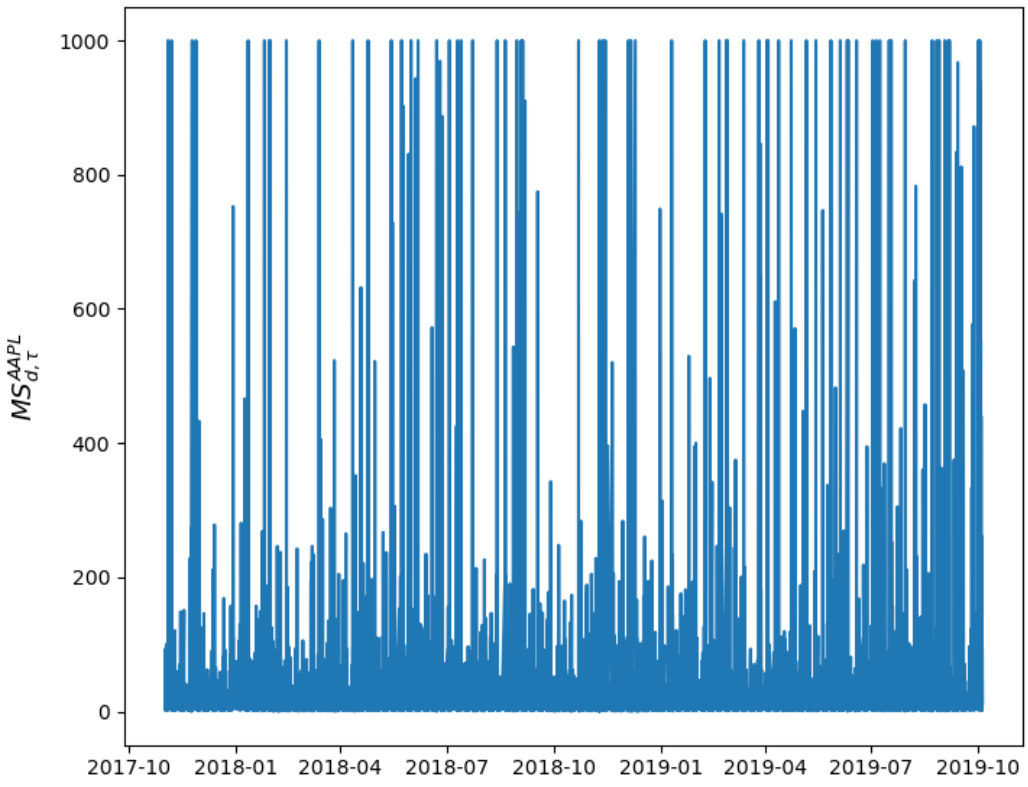
If all available information is immediately incorporated into the price of the stock, there should be no additional information in $MS^j_{d,\tau-1}$ that is not in $z^4_{d,\tau-1}$. Therefore, the proxy for market shocks should have no impact on the future kurtosis of the studied stock. If markets are not perfectly efficient, then ω^{MS} represents causal impact in the sense of Granger's causality (Granger, 1969) in kurtosis.

3.6.2 Empirical Results

The first step in the analysis of extreme risk spillovers is filtering the measures of market shocks defined by (3.6.1). Using five-minute returns and the 100 stocks, we obtain a filtered series of conditionally standardized returns $z^i_{d,\tau}$.

Figure 3.3 shows the values of the market extreme shock measure MS^{AAPL} for Apple stock, meaning that Apple shocks are left out of the calculation. Although it is difficult to read such a plot, there seem to be visible periods with a low and high frequencies of extreme shocks.

Figure 3.3 Extreme shock measure $MS^{AAPL}_{d,\tau}$.



Notes: Values of the market extreme shock measure $MS^{AAPL}_{d,\tau}$ for the AAPL stock based on the remaining 99 stocks in our dataset. Five-minute frequency.

Table 3.5 shows basic statistics of the constructed market extreme shock measures for the four main stocks analyzed. The table demonstrates that leaving out one stock has very little impact on the summary statistics of the measures. Given that the measure is based on the fourth power of the standardized returns, the measure in each period is typically driven by one or a few stocks with high

standardized returns, while the contribution of other stocks is relatively negligible. In other words, it is a measure of any extreme shock within the market rather than an extreme shock to the market as a whole. This is a desired property, since our goal is to study the delayed spillovers of extreme shocks. The statistics also confirm the discussion in the previous section, suggesting that extreme events relative to conditional variance are indeed rare, and a few influential observations drive the estimates of conditional kurtosis.

Table 3.5 Summary statistics of $MS_{d,\tau}$.

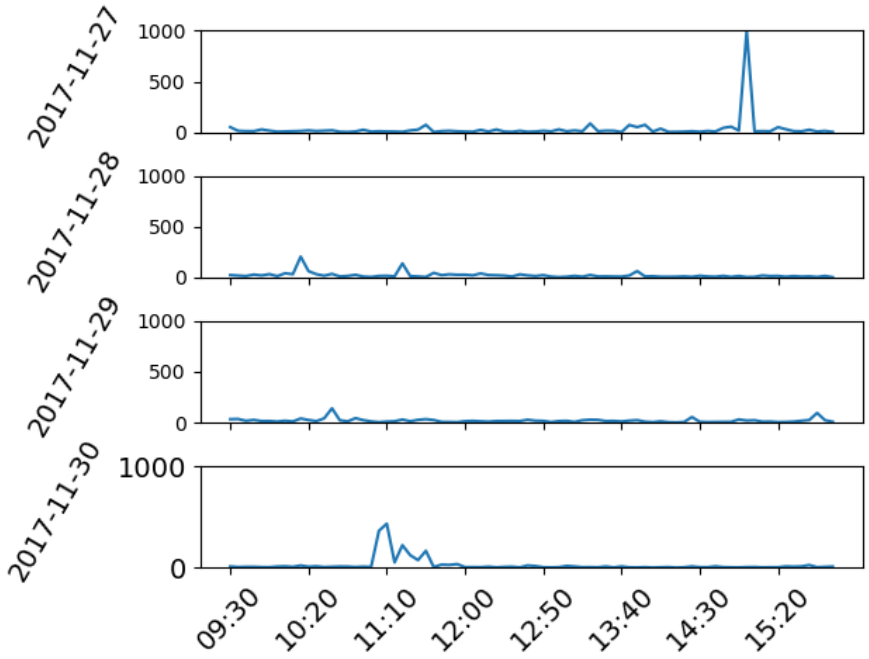
SYMBOL	AAPL	JPM	FDX	ODP
Mean	15.59	15.62	15.56	15.32
Minimum	0.03 $\times 10^{-6}$	0.02 $\times 10^{-6}$	0.02 $\times 10^{-6}$	0.02 $\times 10^{-6}$
25% percentile	4.31	4.32	4.29	4.17
50% percentile	7.30	7.31	7.28	7.10
75% percentile	13.14	13.14	13.12	12.81
Maximum*	168,566.80	168,567.15	168,567.15	168,567.15
ACF 1	0.08	0.08	0.08	0.08

Notes: Summary statistics for the market extreme shock measures for individual stocks. ACF 1 is autocorrelation of the first order. The maximum is calculated without imposing the limits $B()$. Five-minute frequency.

To better understand how the generated variable of market extreme shocks behaves, we more closely explore a small subset of the analyzed time window. Figure 3.4 shows trajectories of the market extreme shock measure on four different days for the Apple stock. The measure is low—close to zero—most of the time, with occasional spikes. November 30 demonstrates a temporary episode of increased extreme risk when some of the 99 stocks experience relatively extreme shocks for almost an hour.

Using the measure of market extreme shocks, we estimate the GARCHK-H model with specification (3.6.3) for the four main analyzed stocks. The results are presented together with GARCHK-H estimates without the market extreme shocks in Table 3.6. The addition of market shocks improves the fit according to the LR tests. The LR test in this context can be understood as a test for

Figure 3.4 Filtered proxy for market extreme shocks $MS_{d,\tau}^{AAPL}$ based on the full set of stocks with AAPL exclusion.



Granger's causality in kurtosis, since the added market shock is lagged by one period. Our observations are aligned with Mazzarisi et al. (2020), who study Granger's causality in tails via predicting the probability of returns exceeding a certain threshold. Our results are more general, because we confirm Granger's causality for kurtosis, which does not depend on an arbitrarily chosen threshold.

The addition of extreme market shocks reduces the coefficient of past shocks. In addition, the market coefficients are higher in all four cases than the coefficient for the past shock of the analyzed stock. This suggests that the main driver behind periods of high kurtosis in the market is spillovers from other stocks, and only a small fraction of shocking news is exclusively related to the analyzed stock. The $dm_{2,MS}$ statistics suggest further improvement in the out-of-sample distributional forecast precision. However, the improvement remains small.

The model relies on filtering four different latent processes: daily volatility, intraday volatility, market extreme shocks, and kurtosis. To better understand how the series relate to each other, Figure 3.5 shows the log returns, diurnal factors, filtered series, and conditional kurtosis mapped to the h parameter of the H-distribution. In the short window of six days, there are several instances when

Table 3.6 GARCHK-H with market shocks.

	AAPL		JPM		FDX		ODP	
μ	0.002	0.007	-0.007	-0.001	-0.004	0.002	-0.011	-0.011
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
α_0	-0.068	-0.065	-0.060	-0.058	-0.073	-0.071	-0.053	-0.051
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
α_1	0.978	0.980	0.978	0.980	0.970	0.971	0.969	0.971
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
α_2	-0.012	-0.012	-0.010	-0.009	-0.011	-0.010	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
α_3	0.060	0.058	0.056	0.054	0.062	0.061	0.033	0.032
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
β_1	0.798	0.829	0.832	0.842	0.779	0.782	0.863	0.863
	(0.032)	(0.027)	(0.030)	(0.027)	(0.030)	(0.028)	(0.023)	(0.023)
β_2	0.177	0.159	0.147	0.145	0.188	0.195	0.120	0.122
	(0.026)	(0.023)	(0.024)	(0.022)	(0.023)	(0.023)	(0.018)	(0.019)
ω_0	1.239	1.330	1.156	1.473	4.208	1.169	4.429	3.343
	(0.288)	(0.183)	(0.356)	(0.197)	(0.608)	(0.220)	(1.831)	(0.864)
ω_1	0.708	0.647	0.713	0.592	0.087	0.716	0.470	0.538
	(0.066)	(0.045)	(0.087)	(0.052)	(0.127)	(0.051)	(0.209)	(0.109)
ω_2	0.008	0.004	0.006	0.001	0.011	0.001	0.012	0.008
	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)	(0.001)	(0.005)	(0.004)
ω^{MS}	-	0.016	-	0.017	-	0.012	-	0.037
	-	(0.003)	-	(0.003)		(0.003)	-	(0.014)
$dm_{2,MS}$		2.91		1.96		2.22		1.71
LR: $\omega^{MS}=0$		84.79		101.96		50.77		38.43

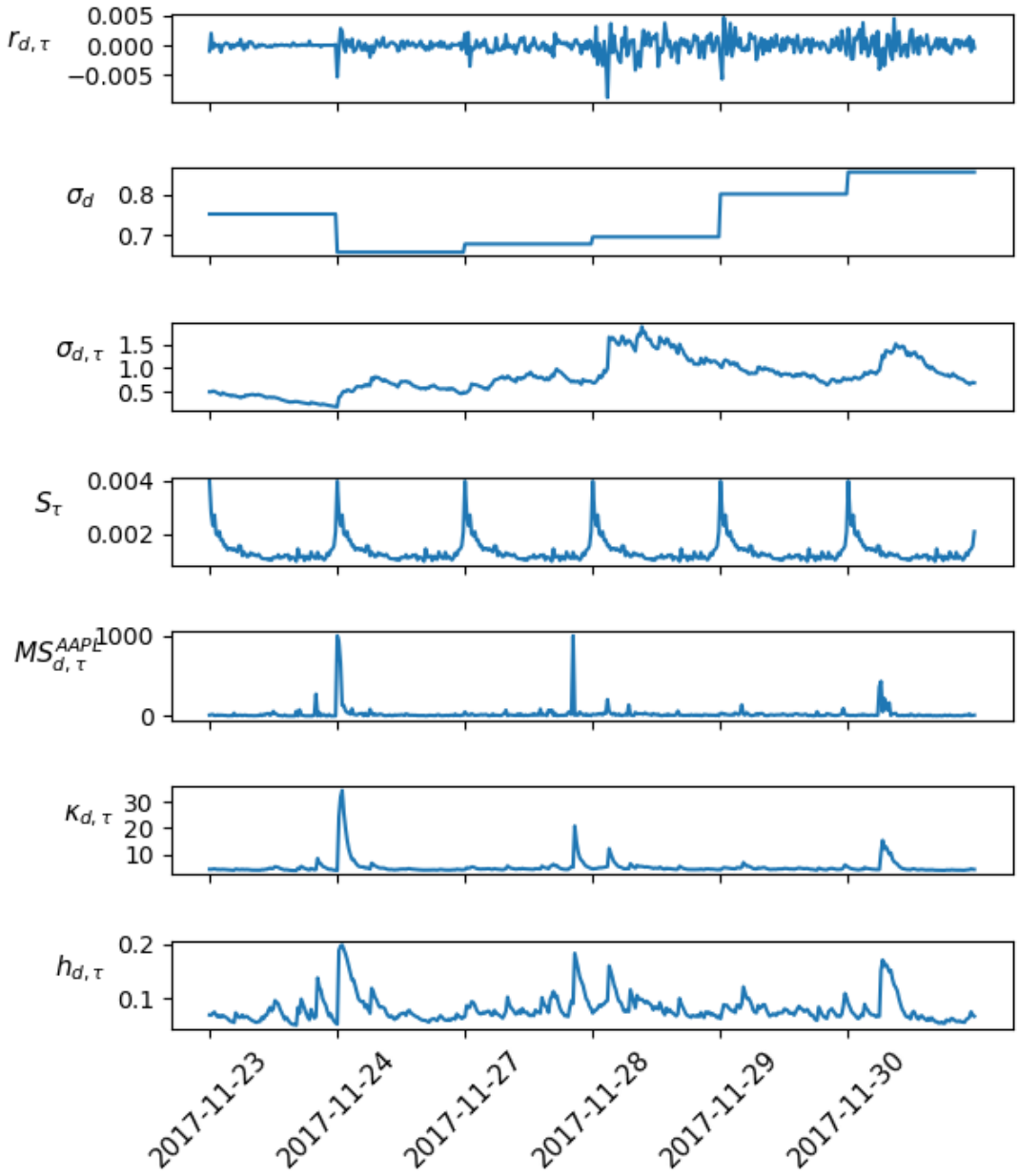
$$\kappa_{d,\tau} = \omega_0 + \omega_1 \kappa_{d,\tau-1} + \omega_2 B\left(z_{d,\tau-1}\right)^4 + \omega^{MS} MS^j_{d,\tau-1}$$

Notes: Estimates for the period from November 1, 2017 to November 8, 2019, five-minute returns. The $dm_{2,MS}$ test compares specification 2 with and without market shocks; calculated for the period from November 1, 2017 to November 8, 2019 with an estimation period from October 29, 2015 to October 31, 2017. Standard errors are reported in parentheses. The $N(0,1)$ critical values for the $dm_{2,MS}$ tests are 1.64 and 2.32 for 5% and 1% sig. levels, respectively. The $\chi^2(1)$ critical values for the LR test are 3.841 and 6.635 for 5% and 1% sig. levels, respectively.

both market extreme shock spikes together with AAPL kurtosis, although there seem to be no visible spikes in kurtosis without spikes in $MS^{AAPL}_{d,\tau}$. While it is

a short window to draw any firm conclusions, such an observation supports the idea that extreme risk spillovers are the main drivers behind periods of increased kurtosis.

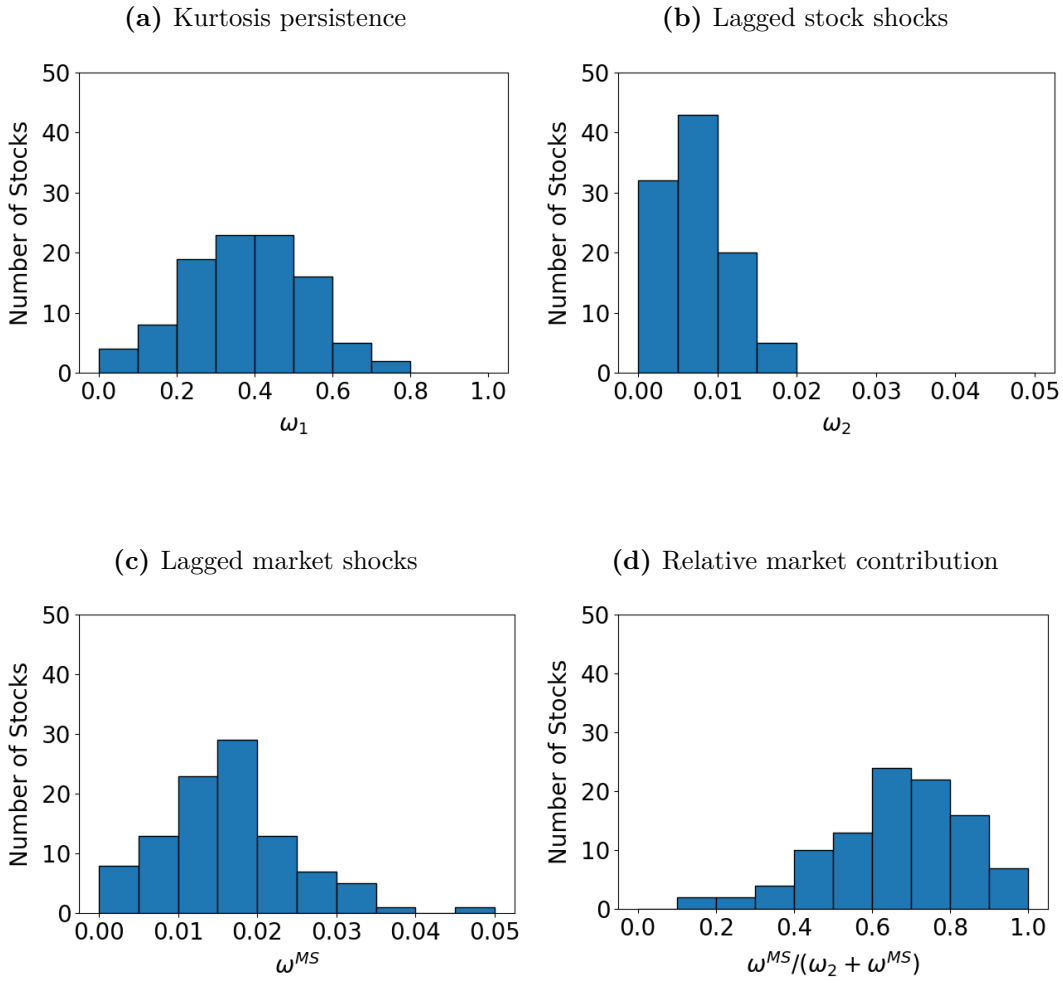
Figure 3.5 Filtered series for AAPL stock for the period November 17-22, 2017.



To answer the question of whether the extreme risk spillovers are the primary driver of kurtosis, we estimate the model for each of the 100 stocks. The estimation results are visualized in Figure 3.6 with the addition of the $\omega^{MS}/(\omega_2 + \omega^{MS})$ ratio,

which represents the relative contribution of market shocks to conditional kurtosis updates. For the majority of the stocks, the ratio is over 50%; thus, we can conclude that extreme risk spillovers are the main driver behind periods with increased kurtosis.

Figure 3.6 Histograms of conditional kurtosis estimates for all 100 stocks analyzed.



$$\kappa_{d,\tau} = \omega_0 + \omega_1 \kappa_{d,\tau-1} + \omega_2 B \left(z_{d,\tau-1}\right)^4 + \omega^{MS} MS^j_{d,\tau-1}$$

3.7 Conclusion

We develop a GARCHK-H model with a novel distribution of innovation that allows platykurtic and leptokurtic shapes and mixed-frequency volatility for modeling conditional kurtosis of intraday stock returns. This novel distribution allows platykurtic and leptokurtic shapes driven by a single parameter without drastically changing the distribution center, unlike GED.

Using the model, we analyze individual returns of selected highly liquid stocks to study time-varying kurtosis, finding the quartic specification of time-varying kurtosis to be suitable for the intraday setup. The periods of high kurtosis are relatively short-lived, and the impact of past shocks is weaker than those typically reported in the literature on modeling daily returns. Subsequently, we extend the model with a generated proxy variable for extreme shocks in the market based on returns of 100 highly liquid US-listed stocks. Using the proxy variable, we perform a test of Granger's causality in kurtosis and conclude that the market influences periods of high kurtosis of individual stocks. Further analysis reveals that spillovers of extreme risk between stocks are the main driver of a period of increased kurtosis, and only a small fraction of extreme risk is related to the individual stock.

The H-distribution performs better empirically. We identify time-varying kurtosis and improve the conditional distribution fit in comparison to models with a constant kurtosis. However, this does not demonstrate a general superiority of the distribution to the t-distribution or the GED. A theoretical comparison is limited because the relationship between daily realized volatility and intraday returns is not fully specified, and empirical performance is examined only for the considered model specifications.

The work could be extended in three main directions; extreme risk spillovers and platykurtic shapes. First, the empirical analysis could be extended from the 100 US-listed stocks to a broader set of stocks and various markets and sizes. Alternatively, the stocks could be grouped by sector, and spillovers between sectors could be studied, revealing whether extreme shock in one sector (e.g., financial services) is more likely to increase the extreme risk of an individual stock within another sector. Second, we did not confirm the need for platykurtic

shapes. However, the exclusively leptokurtic shapes we observe may be artifacts of the models used and a consequence of pre-determined kurtosis. Reformulating our model to a stochastic version in which kurtosis and returns are realized at the same time could potentially allow platykurtic shapes. Third, the model could be extended to fully specify the data-generating process and compare the theoretical properties of the process with the H-distribution to the process with the t-distribution or GED.

Conclusion

This work addresses unique challenges and questions related to conditional distributions of intraday returns. The nature of its contributions is two-fold: methodological and empirical.

The first chapter introduces a novel modeling framework for capturing distributional patterns and demonstrates their importance for distributional forecasting of intraday returns. The proposed modeling framework for distributional adjustment is compatible with models commonly used in the financial literature and practice. The superior distributional forecasting improves the precision of risk management metrics, Value-at-Risk or Expected Shortfall. However, the impact of the work may extend beyond the financial literature, as distributional adjustment may potentially be useful in any distributional modeling in which diurnal or seasonal patterns in distribution are of concern.

The second chapter studies a clustered commonality in intraday volatility of stock returns. Empirical results show the existence of clustering commonality and demonstrate the potential for robust improvement in volatility forecast precision. In addition, the empirical results shed light on commonality and market participant behavior. Although there is no direct policy recommendation arising from our results, the results may be relevant to trading rules in the future.

Finally, the third chapter proposes a novel distribution and studies the conditional kurtosis of intraday returns and spillovers of extreme risk. We provide evidence of the existence of time-varying kurtosis. The empirical results suggest that new shocking information is not immediately incorporated into a stock price, and spillovers of extreme risk are the main reason for delayed incorporation of the new information.

This dissertation consists of three separate chapters, with the properties of the conditional distribution of intraday returns being the common theme. Each of the three chapters has potential for future research, as discussed in their respective conclusions. However, the topics are closely related. Their ideas could be combined for further investigation.

Bibliography

Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, *9*(8), 1295.

Ait-Sahalia, Y., & Jacod, J. (2012). Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature*, *50*(4), 1007–1050.

Ait-Sahalia, Y., Jacod, J., & Li, J. (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics*, *168*(2), 207–222.

Ait-Sahalia, Y., Jacod, J., & Xiu, D. (2020). *Inference on risk premia in continuous-time asset pricing models* (tech. rep.). National Bureau of Economic Research.

Ait-Sahalia, Y., Kalnina, I., & Xiu, D. (2020). High-frequency factor models and regressions. *Journal of Econometrics*, *216*(1), 86–105.

Ait-Sahalia, Y., Li, C. X., & Li, C. (2024). *So many jumps, so few news* (Working Paper No. 32746). National Bureau of Economic Research. http://www.nber.org/papers/w32746

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

Ait-Sahalia, Y., & Xiu, D. (2019). A hausman test for the presence of market microstructure noise in high frequency data. *Journal of Econometrics*, *211*(1), 176–205.

Akaike, H. (2003). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Amado, C., & Teräsvirta, T. (2013). Modelling volatility by variance decomposition. *Journal of Econometrics*, *175*(2), 142–153.

Amado, C., & Teräsvirta, T. (2017). Specification and testing of multiplicative time-varying GARCH models with applications. *Econometric Reviews*, *36*(4), 421–446.

Amaya, D., Christoffersen, P., Jacobs, K., & Vasquez, A. (2015). Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics*, *118*(1), 135–167.

Amisano, G., & Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, *25*(2), 177–190.

Anagnostidis, P., & Fontaine, P. (2020). Liquidity commonality and high frequency trading: Evidence from the French stock market. *International Review of Financial Analysis*, *69*, 101428.

Anatolyev, S. (2021). Directional news impact curve. *Journal of Forecasting*, *40*(1), 94–107.

Anatolyev, S., & Baruník, J. (2019). Forecasting dynamic return distributions based on ordered binary choice. *International Journal of Forecasting*, *35*(3), 823–835.

Anatolyev, S., & Petukhov, A. (2016). Uncovering the skewness news impact curve. *Journal of Financial Econometrics*, *14*(4), 746–771.

Anatolyev, S., & Shakin, D. (2007). Trade intensity in the russian stock market: Dynamics, distribution and determinants. *Applied Financial Economics*, *17*(2), 87–104.

Andersen, T. G., & Bollerslev, T. (1997a). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *Journal of Finance*, *52*(3), 975–1005.

Andersen, T. G., & Bollerslev, T. (1997b). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, *4*(2–3), 115–158.

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., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, *89*(4), 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579–625.

Andersen, T. G., Bollerslev, T., & Meddahi, N. (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, *160*(1), 220–234.

Andersen, T. G., Su, T., Todorov, V., & Zhang, Z. (2024). Intraday periodic volatility curves. *Journal of the American Statistical Association*, *119*(546), 1181–1191.

Andersen, T. G., & Teräsvirta, T. (2009). Realized volatility. In *Handbook of financial time series* (pp. 555–575). Springer.

Angus, J. E. (1994). The probability integral transform and related results. *SIAM review*, *36*(4), 652–654.

Audrino, F., & Camponovo, L. (2015). *Oracle properties, bias correction, and inference of the adaptive lasso for time series extremum estimators*. dx.doi.org/10.2139/ssrn.2340030

Audrino, F., Huang, C., & Okhrin, O. (2019). Flexible HAR model for realized volatility. *Studies in Nonlinear Dynamics & Econometrics*, *23*(3), 20170080.

Audrino, F., & Knaus, S. D. (2016). Lassoing the HAR model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*, *35*(8-10), 1485–1521.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, *61*(4), 1645–1680.

Bali, T. G., Mo, H., & Tang, Y. (2008). The role of autoregressive conditional skewness and kurtosis in the estimation of conditional VaR. *Journal of Banking & Finance*, *32*(2), 269–282.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *64*(2), 253–280.

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, *2*(1), 1–37.

Baur, D. G., & Dimpfl, T. (2019). A quantile regression approach to estimate the variance of financial returns. *Journal of Financial Econometrics*, *17*(4), 616–644.

Bekierman, J., & Gribisch, B. (2021). A mixed frequency stochastic volatility model for intraday stock market returns. *Journal of Financial Econometrics*, *19*(3), 496–530.

Bell, W. R., & Hillmer, S. C. (1984). Issues involved with the seasonal adjustment of economic time series. *Journal of Business & Economic Statistics*, *2*(4), 291–320.

Bernardo, A. E., & Welch, I. (2004). Liquidity and financial market runs. *The Quarterly Journal of Economics*, *119*(1), 135–158.

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

Bollerslev, T., & Ghysels, E. (1996). Periodic autoregressive conditional heteroscedasticity. *Journal of Business & Economic Statistics*, *14*(2), 139–151.

Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *The Review of Financial Studies*, *31*(7), 2729–2773.

Bordignon, S., Caporin, M., & Lisi, F. (2007). Generalised long-memory GARCH models for intra-daily volatility. *Computational Statistics & Data Analysis*, *51*(12), 5900–5912.

Boudt, K., & Zhang, J. (2015). Jump robust two time scale covariance estimation and realized volatility budgets. *Quantitative Finance*, *15*(6), 1041–1054.

Brooks, C., Burke, S. P., Heravi, S., & Persand, G. (2005). Autoregressive conditional kurtosis. *Journal of Financial Econometrics*, *3*(3), 399–421.

Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. *The Review of Financial Studies*, *22*(6), 2201–2238.

Bu, R., Hizmeri, R., Izzeldin, M., Murphy, A., & Tsionas, M. (2023). The contribution of jump signs and activity to forecasting stock price volatility. *Journal of Empirical Finance*, *70*, 144–164.

Bühlmann, P., Rütimann, P., Van De Geer, S., & Zhang, C.-H. (2013). Correlated variables in regression: Clustering and sparse estimation. *Journal of Statistical Planning and Inference*, *143*(11), 1835–1858.

Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: An empirical study on performance estimation methods. *Machine Learning*, *109*(11), 1997–2028.

Chen, J. (2024). Jump clustering, information flows, and stock price efficiency. *Journal of Financial Econometrics*, nbae009.

Chinco, A., Clark-Joseph, A. D., & Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, *74*(1), 449–492.

Chordia, T., Roll, R., & Subrahmanyam, A. (2000). Commonality in liquidity. *Journal of Financial Economics*, *56*(1), 3–28.

Christensen, K., Siggaard, M., & Veliyev, B. (2023). A machine learning approach to volatility forecasting. *Journal of Financial Econometrics*, *21*(5), 1680–1727.

Coroneo, L., & Veredas, D. (2012). A simple two-component model for the distribution of intraday returns. *European Journal of Finance*, *18*(9), 775–797.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Coughenour, J. F., & Saad, M. M. (2004). Common market makers and commonality in liquidity. *Journal of Financial Economics*, *73*(1), 37–69.

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

Dang, T. L., Moshirian, F., & Zhang, B. (2015). Commonality in news around the world. *Journal of Financial Economics*, *116*(1), 82–110.

Du, Z., & Escanciano, J. C. (2017). Backtesting expected shortfall: Accounting for tail risk. *Management Science*, *63*(4), 940–958.

Eckernkemper, T., & Gribisch, B. (2021). Intraday conditional value at risk: A periodic mixed-frequency generalized autoregressive score approach. *Journal of Forecasting*, *40*(5), 883–910.

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

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance*, *48*(5), 1749–1778.

Engle, R. F., & Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, *66*(5), 1127–1162.

Engle, R. F., & Sokalska, M. E. (2012). Forecasting intraday volatility in the US equity market. multiplicative component GARCH. *Journal of Financial Econometrics*, *10*(1), 54–83.

Ergün, A. T., & Jun, J. (2010). Time-varying higher-order conditional moments and forecasting intraday VaR and expected shortfall. *Quarterly Review of Economics and Finance*, *50*(3), 264–272.

Feng, Y., & McNeil, A. J. (2008). Modelling of scale change, periodicity and conditional heteroskedasticity in return volatility. *Economic Modelling*, *25*(5), 850–867.

Franses, P. H., & Ghijsels, H. (1999). Additive outliers, garch and forecasting volatility. *International Journal of Forecasting*, *15*(1), 1–9.

Ghysels, E., Gourieroux, C., & Jasiak, J. (2004). Stochastic volatility duration models. *Journal of Econometrics*, *119*(2), 413–433.

Gourieroux, C., Jasiak, J., & Le Fol, G. (1999). Intra-day market activity. *Journal of Financial Markets*, *2*(3), 193–226.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*(3), 424–438.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Guermat, C., & Harris, R. D. F. (2002). Forecasting value at risk allowing for time variation in the variance and kurtosis of portfolio returns. *International Journal of Forecasting*, *18*(3), 409–419.

Gunnarsson, E. S., Isern, H. R., Kaloudis, A., Risstad, M., Vigdel, B., & Westgaard, S. (2024). Prediction of realized volatility and implied volatility indices using ai and machine learning: A review. *International Review of Financial Analysis*, *93*, 103221.

Hafner, C. M., & Herwartz, H. (2001). Option pricing under linear autoregressive dynamics, heteroskedasticity, and conditional leptokurtosis. *Journal of Empirical Finance*, *8*(1), 1–34.

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

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054.

Hansen, L. P., & Hodrick, R. J. (1980). Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy*, *88*(5), 829–853.

Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, *24*(2), 127–161.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877–906.

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

Harvey, C. R., & Siddique, A. (2000). Conditional skewness in asset pricing tests. *The Journal of Finance*, *55*(3), 1263–1295.

He, C., & Teräsvirta, T. (1999). Properties of moments of a family of garch processes. *Journal of Econometrics*, *92*(1), 173–192.

Headrick, T. C., Kowalchuk, R. K., & Sheng, Y. (2008). Parametric probability densities and distribution functions for tukey g-and-h transformations and their use for fitting data. *Applied Mathematical Sciences*, *2*(9), 449–462.

Heston, S. L., Korajczyk, R. A., & Sadka, R. (2010). Intraday patterns in the cross-section of stock returns. *The Journal of Finance*, *65*(4), 1369–1407.

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*, *64*(5), 2289–2325.

Hong, Y., Liu, Y., & Wang, S. (2009). Granger causality in risk and detection of extreme risk spillover between financial markets. *Journal of Econometrics*, *150*(2), 271–287.

Jacod, J., Li, Y., & Zheng, X. (2017). Statistical properties of microstructure noise. *Econometrica*, *85*(4), 1133–1174.

Jondeau, E., & Rockinger, M. (2003). Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. *Journal of Economic Dynamics and Control*, *27*(10), 1699–1737.

Jondeau, E., & Rockinger, M. (2006). The impact of news on higher moments. *Swiss Finance Institute Research Paper*, (28).

Karolyi, G. A., Lee, K.-H., & Van Dijk, M. A. (2012). Understanding commonality in liquidity around the world. *Journal of Financial Economics*, *105*(1), 82–112.

Kim, T.-H., & White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters*, *1*(1), 56–73.

Koch, A., Ruenzi, S., & Starks, L. (2016). Commonality in liquidity: A demand-side explanation. *The Review of Financial Studies*, *29*(8), 1943–1974.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*(1), 79–86.

Lalancette, S., & Simonato, J.-G. (2017). The role of the conditional skewness and kurtosis in vix index valuation. *European Financial Management*, *23*(2), 325–354.

León, Á., Rubio, G., & Serna, G. (2005). Autoregressive conditional volatility, skewness and kurtosis. *The Quarterly Review of Economics and Finance*, *45*(4-5), 599–618.

Li, Z. M., & Linton, O. (2022). A ReMeDI for microstructure noise. *Econometrica*, *90*(1), 367–389.

Liao, T. W. (2005). Clustering of time series data – a survey. *Pattern Recognition*, *38*(11), 1857–1874.

Loughran, T., & McDonald, B. (2014). Measuring readability in financial disclosures. *The Journal of Finance*, *69*(4), 1643–1671.

Martens, M., Chang, Y.-C., & Taylor, S. J. (2002). A comparison of seasonal adjustment methods when forecasting intraday volatility. *Journal of Financial Research*, *25*(2), 283–299.

Mazzarisi, P., Zaoli, S., Campajola, C., & Lillo, F. (2020). Tail granger causalities and where to find them: Extreme risk spillovers vs spurious linkages. *Journal of Economic Dynamics and Control*, *121*, 104022.

McAleer, M. (2014). Asymmetry and leverage in conditional volatility models. *Econometrics*, *2*(3), 145–150.

McAleer, M., & Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews*, *27*(1-3), 10–45.

Meng, X., & Taylor, J. W. (2020). Estimating value-at-risk and expected shortfall using the intraday low and range data. *European Journal of Operational Research*, *280*(1), 191–202.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, *8*(4), 323–361.

Morris, S., & Shin, H. S. (2004). Liquidity black holes. *Review of Finance*, *8*(1), 1–18.

Mykland, P. A., Zhang, L., & Chen, D. (2019). The algebra of two scales estimation, and the s-tsrv: High frequency estimation that is robust to sampling times. *Journal of Econometrics*, *208*(1), 101–119.

Nekhili, R., & Bouri, E. (2023). Higher-order moments and co-moments' contribution to spillover analysis and portfolio risk management. *Energy Economics*, *119*, 106596.

Nelson, D. B. (1988). *The time series behavior of stock market volatility and returns. unpublished ph.d. dissertation* [Doctoral dissertation, Massachusetts Institute of Technology].

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. (1994). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society*, *62*(6), 1349–1382.

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

Newey, W. K., & West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelation-consistent covariance matrix. *Econometrica*, *55*(3).

Orlowski, P., Schneider, P., & Trojani, F. (2024). On the nature of (jump) skewness risk premia. *Management Science*, *70*(2), 1154–1174.

Paparrizos, J., & Gravano, L. (2015). *k*-shape: Efficient and accurate clustering of time series. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, 1855–1870.

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

Qiu, Y., Zhang, X., Xie, T., & Zhao, S. (2019). Versatile HAR model for realized volatility: A least square model averaging perspective. *Journal of Management Science and Engineering*, *4*(1), 55–73.

Rodriguez, M. J., & Ruiz, E. (2012). Revisiting several popular garch models with leverage effect: Differences and similarities. *Journal of Financial Econometrics*, *10*(4), 637–668.

Rossi, E., & Fantazzini, D. (2015). Long memory and periodicity in intraday volatility. *Journal of Financial Econometrics*, *13*(4), 922–961.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, *63*(1), 289–306.

Smith, D. R. (2007). Conditional coskewness and asset pricing. *Journal of Empirical Finance*, *14*(1), 91–119.

Soltyk, S. J., & Chan, F. (2023). Modeling time-varying higher-order conditional moments: A survey. *Journal of Economic Surveys*, *37*(1), 33–57.

Stephan, J. A., & Whaley, R. E. (1990). Intraday price change and trading volume relations in the stock and stock option markets. *The Journal of Finance*, *45*(1), 191–220.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288.

Tukey, J. W. (1977). Exploratory data analysis. *Reading/Addison-Wesley*.

Wallis, K. F. (1974). Seasonal adjustment and relations between variables. *Journal of the American Statistical Association*, *69*(345), 18–31.

Wang, H., Li, G., & Tsai, C.-L. (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *69*(1), 63–78.

Wu, W. B., Huang, Y., & Huang, Y. (2010). Kernel estimation for time series: An asymptotic theory. *Stochastic Processes and their Applications*, *120*(12), 2412–2431.

Zhang, C., Zhang, Y., Cucuringu, M., & Qian, Z. (2024). Volatility forecasting with machine learning and intraday commonality. *Journal of Financial Econometrics*, *22*(2), 492–530.

Zhao, S., Lu, Q., Han, L., Liu, Y., & Hu, F. (2015). A mean-cvar-skewness portfolio optimization model based on asymmetric laplace distribution. *Annals of Operations Research*, *226*, 727–739.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429.

List of Figures

List of Tables

A Chapter 1 Attachments

A.1 Conditional Distribution of p_t Process

Lemma A.1.1. Let x_t be a continuous process with a strictly increasing marginal cdf G, and conditional cdf and pdf $G_{x_t|\mathcal{F}_{t-1}}(x)$ and $g_{x_t|\mathcal{F}_{t-1}}(x)$, respectively. Then, the process constructed as $p_t := G(x_t)$ has, respectively, conditional cdf and pdf

$$\begin{aligned} K_{p_t|\mathcal{F}_{t-1}}(p) =& G_{x_t|\mathcal{F}_{t-1}}\left(G^{-1}(p)\right), \\ k_{p_t|\mathcal{F}_{t-1}}(p) =& g_{x_t|\mathcal{F}_{t-1}}\left(G^{-1}(p)\right)\left(\left.\frac{dG^{-1}(s)}{ds}\right|_{s=p}\right). \end{aligned}$$

Proof. Since $G(x)$ is strictly increasing, the inverse exists, and

$$\mathbb{P}[x_t \leq x|\mathcal{F}_{t-1}] = \mathbb{P}[p_t \leq G^{-1}(x)|\mathcal{F}_{t-1}] = G_{x_t|\mathcal{F}_{t-1}}(G^{-1}(p)).$$

The pdf follows by differentiation, where $\frac{dG^{-1}(s)}{ds}$ is always positive due to the strict monotonicity of G. □

A.2 Marginal Distribution of GARCHK

Consider the following processes inspired by Brooks et al. (2005),

$$x_t = \sigma_t \varepsilon_t, \tag{A.2.1}$$

$$\varepsilon_t \sim t(v_t), \tag{A.2.2}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \varepsilon_{t-1^2}^2 \sigma_{t-1}^2, \tag{A.2.3}$$

$$v_t = \frac{6}{\kappa_t} + 4, \tag{A.2.4}$$

$$\kappa_t = \beta_0 + \beta_1 \kappa_{t-1} + \beta_2 \varepsilon_{t-1}^4. \tag{A.2.5}$$

With $\mathcal{F}_{t-1}$, we are able to filter-out σ_t^2 and v_t. Let $F^{Student}(x|v)$ and $f^{Student}(x|v)$ be the cdf and pdf of Students' t-distribution with v degrees of freedom, respectively. This yields a marginal cdf and pdf of x_t as

$$G(x) = \mathbb{E}\left[F^{Student}\left(\frac{x}{\sigma_t}\middle| v_t\right)\right], \tag{A.2.6}$$

$$f(x) = \mathbb{E}\left[\frac{1}{\sigma_t} f^{Student}\left(\frac{x}{\sigma_t}\middle| v_t\right)\right]. \tag{A.2.7}$$

The marginal cdf and pdf of x_t can be approximated by a Monte Carlo simulation as

$$G(x) \approx \frac{1}{nsim} \sum_{i=1}^{nsim} F^{Student}\left(\frac{x}{\sigma_T^i}\middle| v_T^i\right), \tag{A.2.8}$$

$$g(x) \approx \frac{1}{nsim} \sum_{i=1}^{nsim} \frac{1}{\sigma_T^i} f^{Student}\left(\frac{x}{\sigma_T^i}\middle| v_T^i\right), \tag{A.2.9}$$

where (σ_T^1, v_T^1), (σ_T^2, v_T^2), ..., $(\sigma_T^{nsim}, v_T^{nsim})$ are $nsim$ vectors of time-varying parameters simulated for T periods. T should be large enough to make the impact of the initial conditions v_0 and σ_0 negligible.

A.3 Computational Complexity

The framework relies heavily on numerical procedures. Let us explore the computation times of DA-EGARCH to intuitively understand the time needed to estimate the model.

Table C1 shows the computational complexity of using kernel density estimation as described in Section 1.2.3 for the first-step adjustment and the time needed for estimating the Bandwidths using cross-validation. We used the first 100, 1000, 5000, and 10,000 observations for AAPL in our dataset. Since the actual time depends on the actual data, the used optimizer, implementation, initial guess, etc., the results are rather qualitative than quantitative. Ignoring the computational complexity of calculating the standard normal cdf and pdf, the kernel density estimate adjustment has an asymptotic computational complexity T^2.

Table C1 Computational complexity of kernel-based $r_{d,\tau}$ to p_t PIT.

Sample Size (T)	Complexity (T^2)	Adjustment (ms*)	Bandwidths (ms*)
100	10,000	2	158
1,000	1,000,000	82	10,753
5,000	25,000,000	2,074	424,122
10,000	100,000,000	8,760	2,321,625

*milliseconds

Column bandwidths represent the selection of bandwidths by cross-validation.

Next, we focus on the second step of the estimation procedure, fitting the PITed EGARCH to p_t. The main difference from simply fitting EGARCH is the need to calculate the marginal distribution of the EGARCH process. Using the algorithm described in Section 1.2.4, we create an equidistant grid $0 < p_0 < p_1 < ... < p_m < 1$ and use the following parameters

- $m = 10,000$,
- $error = 10^{-6}$,
- number of conditional variance trajectories $n_\sigma = 1,000$,
- number of simulated periods: 200.

Ignoring the computational complexity of computing the standard normal cdf, pdf, and logarithm, the asymptotic computational complexity of the algorithm calculating marginal pdf, cdf, and quantile functions is linear in sample size, when the number of gridpoints m is fixed. The time needed for likelihood calculation and MLE is mostly driven by computations of the marginal distribution. Therefore, both times grow linearly with sample size.

Table C2 Computational complexity of DA-EGARCH maximum likelihood estimation (MLE).

Sample Size (T)	Marginal Distribution (ms*)	MLE (ms*)
100	40	6,920
1000	477	53,175
5000	1,660	218,375
10000	3,303	512,982

*milliseconds

Marginal Distribution (ms) column represents the time needed to calculate the marginal pdf, cdf, and quantile function of the EGARCH process for likelihood evaluation in each iteration of MLE optimization.

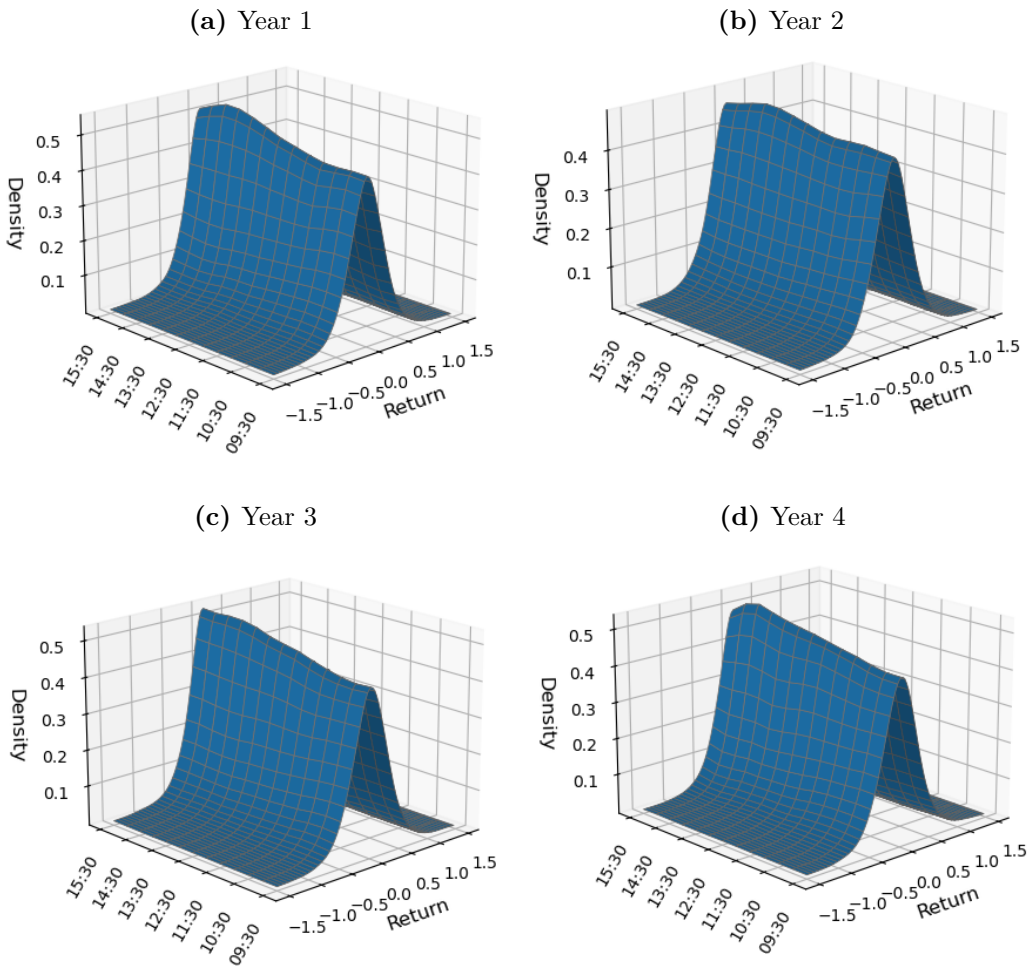
A.4 Stability of Diurnal Patterns

The key assumption of the DA-EGARCH model is the stability of the diurnal patterns over time. Andersen et al. (2024) documents a slow change of diurnal patterns in volatility over time. Similarly, there could be slow changes in the distributional diurnal patterns. Figure A.1 shows the marginal intraday distributions estimated for the four years covered by our data set. The densities are averaged across the 50 stocks to reveal potential slow changes in the whole market. Generally, the diurnal patterns in the distribution beyond volatility are visually similar across years. The most pronounced pattern of lower density around the center of the distribution in the morning, compared to the afternoon, is identical for each year. The actual values are different, which may be driven by uncertainty arising from estimation and potentially slow changes in the distribution over time. However, the results suggest that extending the model with slowly drifting distributions could potentially further improve the precision of the forecasts.

The stability of the diurnal patterns is also relevant for interpreting the results of the moving window exercise. A moving window allows for a slow change in diurnal distributional patterns as the model is regularly re-estimated. However, the speed of the change depends on the window length. Additionally, all observations within the window are weighted equally, whereas estimation of slowly drifting distributional diurnal patterns could benefit from allocating more weight to more recent observations than to distant past observations. In the exercise with daily updated estimates based on 150-day long moving windows, the DA-EGARCH model strongly outperformed the MGARCH-t model in terms of forecasting the tail of the conditional distribution. On the other hand, in the exercise with annually updated estimates and a two-year estimation window, the forecasting performance of the DA-EGARCH model of the tail was close to that of the MGARCH-t model. This suggests a potential for further future improvement by extending the diurnal adjustment with a slow drift of patterns over time. Studying time-varying diurnal patterns is out of the scope of the chapter and is left for future research.

Figure A.1 demonstrates that the general form of the diurnal pattern is not changing over time. However, the actual values of the marginal CDF may vary over time. To explore the potential stability more closely, we examine the

Figure A.1 Annual averages of non-parametric estimates of marginal densities of standardized log-returns.

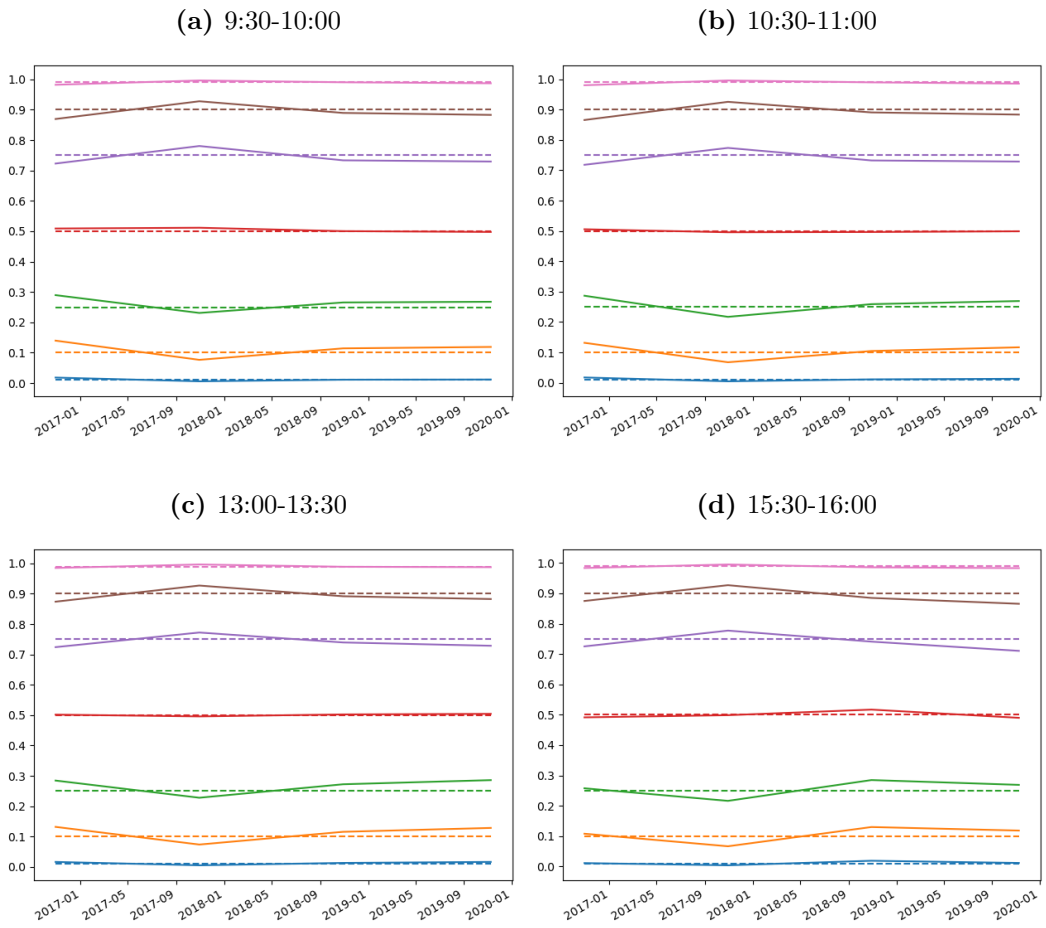


Timeseries of 50 US-traded stocks split into four annual windows covering the period from October 29, 2015, to November 8, 2019. Returns from each window are standardized, and then the non-parametric kernel density estimator of diurnal patterns is applied to each stock individually. Estimated densities are averaged across stocks.

$0.01, 0.1, 0.25, 0.5, 0.75, 0.9$, and 0.99 quantiles in four chosen intraday periods. For an assessment, we use identity $\gamma = F(F^{-1}(\gamma, \tau), \tau)$; $F^{-1}(\gamma, \tau)$ is estimated using the entire dataset, while $F(r, \tau)$ is annually re-estimated for each subperiod. The results averaged across 50 stocks are shown in Figure A.2. We average the 50 stocks to demonstrate a general trend. The results look very similar for individual stocks. If the marginal distribution is stable over the years, the solid and dashed lines should overlap without a clear trend. The solid lines spread out and contract in parallel. This suggests that potential drifts in diurnal volatility patterns are the main contributor to potential instability over time.

These results do not invalidate our work because we use a moving window but rather suggest the most promising extension; incorporating slow drift in diurnal patterns could further improve forecasting performance.

Figure A.2 Stability of Diurnal Adjustment - Quantile Assesment



The dashed lines represent various values of γ and solid lines with corresponding color represent $\widehat{F}^*(\widehat{F}^{-1}(\gamma,\tau),\tau)$, where $\widehat{F}^*$ is annually re-estimated, and $\widehat{F}$ is estimated once using the whole period of from Oct 29, 2015, to Nov 8, 2019. The shown results are averages of the 50 stocks.

B Chapter 2 Attachments

B.1 Full List of Symbols, Sectors and Industries

Table C1 Stocks, industries and sectors analyzed.

Sector	Industry	Symbols
Communication Services	Advertising Agencies	OMC, IPG
	Internet Content & Information	GOOG, GOOGL
	Electronic Gaming & Multimedia	EA
	Entertainment	NFLX, DIS, LYV, NWS, NWSA
	Telecom Services	CMCSA, CHTR
Consumer Discretionary	Specialty Retail	ORLY, AZO, BBY
	Auto Parts	GPC, APTV, LKQ, BWA
	Resorts & Casinos	LVS, MGM, CZR
	Footwear & Accessories	NKE, DECK
	Restaurants	MCD, CMG, DRI, DPZ
	Apparel Manufacturing	RL
	Packaging & Containers	PKG, IP, AVY
	Auto & Truck Dealerships	KMX
	Personal Services	ROL
	Travel Services	BKNG, RCL, CCL, EXPE, NCLH
	Lodging	MAR, HLT
	Home Improvement Retail	HD, LOW
	Furnishings, Fixtures & Appliances	MHK
	Auto Manufacturers	GM, F
	Apparel Retail	ROST, LULU
	Residential Construction	DHI, LEN, NVR, PHM
	Internet Retail	AMZN, EBAY
	Leisure	HAS
Consumer Staples	Beverages – Wineries & Distilleries	BF.B
	Grocery Stores	KR
	Farm Products	ADM, BG
	Confectioners	MDLZ, HSY
	Packaged Foods	KHC, GIS, K, MKC, HRL, CAG CPB
	Household & Personal Products	PG, CL, KMB, EL, CHD, CLX
	Discount Stores	COST, DG, DLTR
	Beverages – Non-Alcoholic	KO, PEP, MNST, KDP
	Tobacco	PM, MO

Table C2 Stocks, industries and sectors analyzed.

Sector	Industry	Symbols
Energy	Oil & Gas Midstream	OKE, KMI
	Oil & Gas Integrated	CVX
	Oil & Gas Refining & Marketing	PSX, MPC
	Oil & Gas Exploration & Production	COP, EOG, FANG, OXY, HES, EQT DVN, APA
	Oil & Gas Equipment & Services	HAL
Financials	Capital Markets	MS, GS, MKTX
	Banks - Regional	PNC, MTB, FITB, HBAN, RF, CFG KEY
	Asset Management	BX, BLK, KKR, AMP, RJF, NTRS PFG, BEN, IVZ
	Credit Services	MA, AXP, PYPL, COF, DFS
	Financial Data & Stock Exchanges	MCO, ICE, CME, MSCI, NDAQ, CBOE FDS
	Insurance – Property & Casualty	PGR, CB, ALL, HIG, CINF, L AIZ
	Insurance – Life	AFL, MET, PRU
	Insurance – Diversified	BRK.B, AIG, ACGL
	Banks – Diversified	JPM, BAC, C, BK
	Insurance Brokers	MMC, AON, AJG, BRO

Table C3 Stocks, industries and sectors analyzed.

Sector	Industry	Symbols
Healthcare	Drug Manufacturers – General	LLY, JNJ, ABBV, MRK, AMGN, PFE BMY, GILD, BIIB
	Drug Manufacturers – Specialty & Generic	CTLT
	Diagnostics & Research	DHR, A, IDXX, MTD, LH, DGX CRL
	Medical Distribution	MCK, COR, CAH, HSIC
	Biotechnology	REGN, INCY
	Medical Instruments & Supplies	ISRG, BDX, RMD, COO, ALGN, HOLX BAX
	Healthcare Plans	CI, CVS, HUM, CNC, MOH
	Medical Care Facilities	HCA, DVA
	Medical Devices	ABT, BSX, MDT, EW, DXCM, PODD
Industrials	Airlines	DAL, LUV
	Farm & Heavy Construction Machinery	CAT, DE, PCAR
	Trucking	ODFL
	Consulting Services	EFX
	Engineering & Construction	PWR
	Conglomerates	HON, MMM
	Specialty Industrial Machinery	ETN, PH, ITW, EMR, CMI, AME IR, ROK, DOV, PNR, IEX, NDSN GNRC, AOS
	Aerospace & Defense	GE, LMT, BA, GD, NOC, HII
	Industrial Distribution	GWW, FAST, POOL
	Specialty Business Services	CTAS, CPRT, GPN
	Electrical Equipment & Parts	HUBB
	Security & Protection Services	ALLE
	Railroads	CSX, NSC
	Waste Management	RSG
	Integrated Freight & Logistics	FDX, JBHT, EXPD, CHRW
	Building Products & Equipment	JCI, BLDR, MAS
Materials	Steel	NUE
	Agricultural Inputs	CF, MOS, FMC
	Chemicals	CE
	Specialty Chemicals	LIN, APD, ECL, PPG, LYB, IFF ALB, EMN
	Gold	NEM
	Building Materials	MLM
	Copper	FCX

Table C4 Stocks, industries and sectors analyzed.

Sector	Industry	Symbols
Real Estate	REIT - Office	ARE, BXP
	REIT - Hotel & Motel	HST
	REIT - Industrial	PLD, PSA, EXR
	REIT - Healthcare Facilities	DOC
	REIT - Specialty	AMT, EQIX, DLR, CCI, IRM
	REIT - Retail	O, KIM, REG, FRT
	Real Estate Services	CSGP
	REIT - Residential	AVB, EQR, ESS, MAA, CPT
Technology	Software - Application	CRM, NOW, INTU, ADP, CDNS, ADSK ROP, FICO, PAYX, ANSS, PTC, PAYC
	Semiconductors	NVDA, AVGO, AMD, QCOM, MU, ADI INTC, NXPI, MCHP, MPWR, ON, QRVO
	Communication Equipment	CSCO, MSI, HPE, JNPR
	Information Technology Services	ACN, IBM, FI, FIS, CTSH, IT BR, CDW, LDOS, EPAM, JKHY
	Software - Infrastructure	MSFT, ORCL, ADBE, PANW, FTNT, GDDY FFIV, AKAM
	Scientific & Technical Instruments	GRMN, KEYS
	Semiconductor Equipment & Materials	AMAT, LRCX, KLAC
	Solar	FSLR, ENPH
	Computer Hardware	ANET, HPQ, NTAP
	Electronic Components	APH, GLW, JBL
	Consumer Electronics	AAPL
Utilities	Utilities - Independent Power Producers	NRG
	Utilities - Regulated Gas	ATO, NI
	Utilities - Diversified	AES
	Utilities - Regulated Water	AWK
	Utilities - Regulated Electric	NEE, DUK, AEP, D, PEG, EXC ED, EIX, ETR, DTE, PPL, AEE FE, ES, CNP, CMS, LNT, EVRG PNW

B.2 Tables

Table C5 Pair-wise counts of MSE model superiority out of 363 stocks – setup All.

	HAR-D	HAR-D, Market	Lasso, Market	Adaptive Lasso $\gamma = 0.5$	Adaptive Lasso $\gamma = 1$	Adaptive Lasso $\gamma = 2$	Lasso, Sectors	Lasso, Industries	CRL, 5 Clusters	CRL, 10 Clusters	CRL, 20 Clusters
HAR-D	0 (0.00%)	44 (12.12%)	58 (15.98%)	62 (17.08%)	69 (19.01%)	76 (20.94%)	46 (12.67%)	54 (14.88%)	63 (17.36%)	48 (13.22%)	53 (14.60%)
HAR-D Market	319 (87.88%)	0 (0.00%)	263 (72.45%)	78 (21.49%)	96 (26.45%)	115 (31.68%)	44 (12.12%)	71 (19.56%)	69 (19.01%)	63 (17.36%)	53 (14.60%)
Lasso Market	305 (84.02%)	100 (27.55%)	0 (0.00%)	66 (18.18%)	79 (21.76%)	99 (27.27%)	32 (8.82%)	59 (16.25%)	63 (17.36%)	51 (14.05%)	50 (13.77%)
Adaptive Lasso Sectors, $\gamma = 0.5$	301 (82.92%)	285 (78.51%)	297 (81.82%)	0 (0.00%)	269 (74.10%)	273 (75.21%)	75 (20.66%)	124 (34.16%)	146 (40.22%)	143 (39.39%)	105 (28.93%)
Adaptive Lasso Sectors, $\gamma = 1$	294 (80.99%)	267 (73.55%)	284 (78.24%)	94 (25.90%)	0 (0.00%)	247 (68.04%)	63 (17.36%)	105 (28.93%)	112 (30.85%)	112 (30.85%)	91 (25.07%)
Adaptive Lasso Sectors, $\gamma = 2$	287 (79.06%)	248 (68.32%)	264 (72.73%)	90 (24.79%)	116 (31.96%)	0 (0.00%)	50 (13.77%)	95 (26.17%)	98 (27.00%)	98 (27.00%)	70 (19.28%)
Lasso Sectors	317 (87.33%)	319 (87.88%)	331 (91.18%)	288 (79.34%)	300 (82.64%)	313 (86.23%)	0 (0.00%)	166 (45.73%)	207 (57.02%)	203 (55.92%)	162 (44.63%)
Lasso Industries	309 (85.12%)	292 (80.44%)	304 (83.75%)	239 (65.84%)	258 (71.07%)	268 (73.83%)	197 (54.27%)	0 (0.00%)	215 (59.23%)	216 (59.50%)	180 (49.59%)
CRL, 5 Clusters	300 (82.64%)	294 (80.99%)	300 (82.64%)	217 (59.78%)	251 (69.15%)	265 (73.00%)	156 (42.98%)	148 (40.77%)	0 (0.00%)	186 (51.24%)	140 (38.57%)
CRL, 10 Clusters	315 (86.78%)	300 (82.64%)	312 (85.95%)	220 (60.61%)	251 (69.15%)	265 (73.00%)	160 (44.08%)	147 (40.50%)	177 (48.76%)	0 (0.00%)	138 (38.02%)
CRL, 20 Clusters	310 (85.40%)	310 (85.40%)	313 (86.23%)	258 (71.07%)	272 (74.93%)	293 (80.72%)	201 (55.37%)	183 (50.41%)	223 (61.43%)	225 (61.98%)	0 (0.00%)

The numbers represent absolute numbers of stocks for which the model in a row outperforms the model in the column in terms of out-of-sample MSE.

Out-of-sample evaluation period spans August 30, 2017 – November 8, 2019.

B.3 Microstructure Noise

To assess the potential presence of microstructure noise, we employ a robust Hausman test developed by Ait-Sahalia and Xiu (2019), denoted as H_{3n}[1]. The null hypothesis of the test assumes no presence of the microstructure noise. The test relies on several hyperparameters, which are set to identical values as in Ait-Sahalia and Xiu (2019).

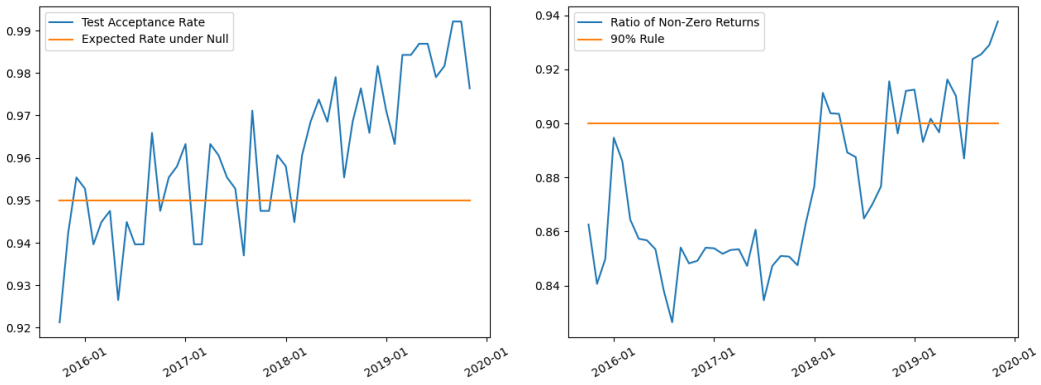
We perform tests with 0.05 significance monthly for each stock within our dataset. Figure B.1, panel (a), shows the ratio of stocks for which the tests do not reject the null hypothesis of no presence of microstructure noise. The expected acceptance rate is 95%; therefore, stocks with rejected nulls are likely false positives.

The test relies on non-zero returns. Ait-Sahalia, Kalnina and Xiu (2020) uses a sequential testing procedure based on the H_{3n} test to choose the highest acceptable sampling frequency for each individual stock. They gradually increase the frequency, and in each iteration, they test for the presence of microstructure noise and check whether at least 90% of returns are non-zero. The procedure is stopped when the test is rejected, or the number of non-zero returns is below the 90% threshold. The check for a sufficient number of non-zero returns is motivated by achieving reasonable power of the test. We avoid choosing different frequencies for different stocks because we forecast 30-minute realized log-variance using other stocks' realized log-variances, and a sampling frequency lower than 3 minutes could negatively impact the efficiency of realized log-variance due to a small number of observations. The 90% non-zero returns rule is arbitrary and, as the authors acknowledge, it is a conservative choice. We, instead of applying the strict 90% rule, which many of the stocks fail to meet, calculate the percentage of non-zero returns within our dataset for each month shown in Figure B.1, panel b. If a lack of power drove the test results, we should see a higher acceptance rate during the period with more zero returns. The results show the exact opposite. Therefore, we conclude that our results are likely not significantly impacted by microstructure noise.

[1]A formal introduction of the test would require an introduction of an extensive list of definitions and additional notation. For a technical description, we refer to the original source.

Figure B.1 Tests for Microstructure Noise

(a) Acceptance Rate **(b)** Non-Zero Return Occurrence Rate



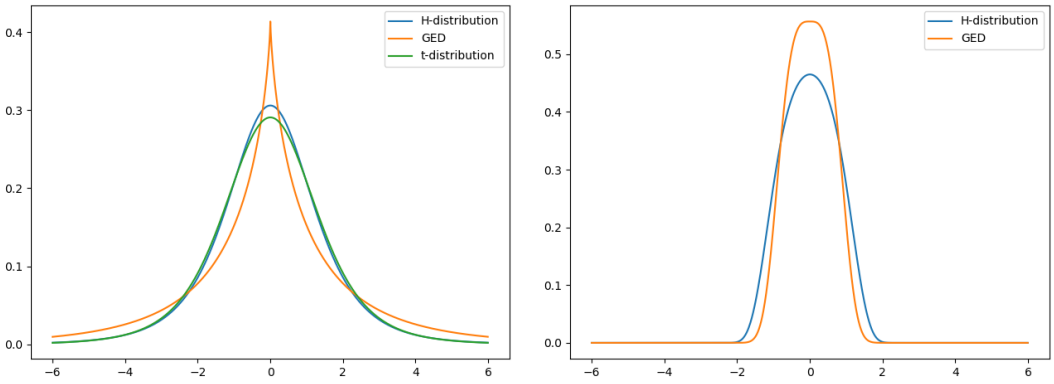
Panel (a) shows the ratio of stocks for which the Hausman test with 0.05 significance does not reject the absence of microstructure noise. Tests are conducted monthly. Panel (b) shows an average ratio of non-zero returns within a month. Calculated using 3-minute returns, covering the period from November 2015 to October 2019.

C Chapter 3 Attachments

C.1 Comparison of Densities

Figure C.1 Comparison of platy-/lepto-kurtic distributions

(a) Leptokurtic - $Kurtosis = 10$ **(b)** Platykurtic - $Kurtosis = 2.3$



Notes: Comparison of shapes of the H-distribution, generalized error distribution (GED), and Student's t-distribution. All three distributions are standardized to unit variance.

Unlike the t-distribution and H-distribution, the GED also changes the 'peakedness' of the distribution. A leptokurtic GED has a pointed central part of the distribution, while in the case of a platykurtic distribution it flattens with a uniform distribution in the limit as shape parameter $v \to \infty$ (A. Harvey & Lange, 2017).

C.2 H-Distribution PDF, CDF and Quantiles

Proof of proposition 3.3.0.1. We find the inversion for $h > 0$. The inversion for $h < 0$ directly follows, and $h = 0$ case is identity.

Let

$$k(x) = x \exp(x),$$

and assume $k(x) \geq 0$, then the inversion is the Lambert W function (principal branch W_0)

$$k^{-1} = W(k).$$

We define

$$\text{sign}(y) = \begin{cases} 1 \text{ if } y > 0, \\ -1 \text{ if } y \leq 0, \end{cases}$$

to distinguish the sign, and assume $h > 0$. Tukey's transformation can be then rewritten as

$$\begin{aligned} y &= z \exp\left\{\frac{h_\tau z^2}{2}\right\}, \\ \text{sign}(y)|y| &= \text{sign}(y)|z| \exp\left\{\frac{h_\tau z^2}{2}\right\}, \\ |y| &= |z| \exp\left\{\frac{h_\tau z^2}{2}\right\} \\ &= \sqrt{k(hz^2)}\sqrt{\frac{1}{h}}. \end{aligned}$$

By taking squares and using the Lambert W to solve for z^2

$$\begin{aligned} y^2 &= k(hz^2)\frac{1}{h}, \\ W\left(y^2 h\right) &= hz^2, \\ z^2 &= W\left(y^2 h\right)\frac{1}{h}. \end{aligned}$$

The equation has two roots of z, in which the one matching the sign is the required inversion, i.e.,

$$z = \text{sign}(y)\sqrt{\frac{W\left(y^2 h\right)}{h}}.$$

□

Proof of proposition 3.3.0.2. Assume that X has symmetric distribution around 0 with pdf $f(x)$ and cdf $F(x)$.

First, consider $h > 0$. Using the inverse 3.3.0.1, the CDF of $K_h(X)$ is

$$\begin{aligned} G(y) &= \mathbb{P}\left[X \exp\left(\frac{hX^2}{2}\right) \leq y\right], \\ &= \mathbb{P}\left[X \leq \text{sign}(y)\sqrt{\frac{W\left(y^2 h\right)}{h}}\right], \\ &= F\left(\text{sign}(y)\sqrt{\frac{W\left(y^2 h\right)}{h}}\right). \end{aligned}$$

The derivative of the Lambert function obtained using an implicit differentiation is

$$\frac{dW(k)}{dk} = \frac{1}{k + e^{W(k)}} \text{ for } k \geq 0,$$

therefore,

$$\frac{\partial \sqrt{\frac{W(y^2h)}{h}}}{\partial y} = \sqrt{\frac{1}{W\left(y^2h\right)}} \frac{y\sqrt{h}}{y^2h + \exp\left(W(y^2h)\right)}.$$

The density is then

$$f\left(\sqrt{\frac{W(y^2h)}{h}}\right) \sqrt{\frac{1}{W\left(y^2h\right)}} \frac{|y|\sqrt{h}}{y^2h + \exp\left(W(y^2h)\right)}.$$

The quantile function is then given as

$$y_p = \begin{cases} F^{-1}(p) \exp\left(\frac{1}{2}h\left(F^{-1}(p)\right)^2\right) & \text{if } p > 0.5 \\ -F^{-1}(p) \exp\left(\frac{1}{2}h\left(F^{-1}(p)\right)^2\right) & \text{if } p \leq 0.5 \end{cases}.$$

Because under the assumption of symmetricity $\text{sign}(y) = 1 \Leftrightarrow p > 0.5$ and $\text{sign}(z) = -1 \Leftrightarrow p \leq 0.5$ we obtain the desired quantile function. For negative $h < 0$, the CDF is

$$\begin{aligned} G(y) &= \mathbb{P}\left[\text{sign}(y)\sqrt{\frac{W(X^2|h|)}{|h|}} \leq y\right], \\ &= \mathbb{P}\left[X \leq y \exp\left(\frac{y^2|h|}{2}\right)\right], \\ &= F\left(y \exp\left(\frac{y^2|h|}{2}\right)\right). \end{aligned}$$

and the PDF directly follows as

$$g(y) = f\left(y \exp\left(\frac{y^2|h|}{2}\right)\right)\left[\exp\left(\frac{y^2|h|}{2}\right) + y^2 \exp\left(\frac{y^2|h|}{2}\right)|h|\right].$$

□

C.3 GARCHK-H Estimates: JPM NIC

Table C1 Estimation results of models with the H-distribution for JPM

Specification (ξ) $\rightarrow$	1	2	3	4	5
μ	-0.007 (0.004)	-0.007 (0.004)	-0.007 (0.004)	-0.007 (0.004)	-0.003 (0.004)
α_0	-0.073 (0.003)	-0.060 (0.003)	-0.060 (0.003)	-0.071 (0.004)	-0.093 (0.003)
α_1	0.971 (0.002)	0.978 (0.002)	0.979 (0.002)	0.973 (0.002)	0.974 (0.002)
α_2	-0.010 (0.001)	-0.010 (0.001)	-0.008 (0.001)	-0.010 (0.001)	-0.008 (0.001)
α_3	0.067 (0.002)	0.056 (0.003)	0.056 (0.003)	0.065 (0.003)	0.089 (0.003)
β_1	0.797 (0.031)	0.832 (0.030)	0.839 (0.029)	0.803 (0.031)	0.746 (0.039)
β_2	0.172 (0.024)	0.147 (0.024)	0.140 (0.023)	0.167 (0.024)	0.214 (0.030)
$\omega_0^{(\log)}$	4.191 (0.066)	1.156 (0.356)	1.103 (0.334)	0.678 (1.187)	0.168 (0.009)
$\omega_1^{(\log)}$		0.713 (0.087)	0.728 (0.079)	0.520 (0.829)	0.929 (0.007)
ω_2		0.006 (0.002)	0.005 (0.002)		
ω_3			-0.009 (0.026)		
$\omega_4^{\log}$				0.010 (0.009)	-0.077 (0.005)
$\omega_5^{\log}$					0.189 (0.009)
b					1.750 (0.110)
MCS p-val	0.245	0.876	1.000	0.253	0.102
$LR_{1,\xi}$ (p-val)	0.00 (1.00)	27.64 (0.00)	33.73 (0.00)	0.81 (0.667)	65.24 (0.00)

Notes: Estimates for the period from November 1, 2017 to November 8, 2019, five-minute returns. MCS performed for the period from November 1, 2017 to November 8, 2019 with an estimation period from October 29, 2015 to October 31, 2017. $\omega^{(\log)}$ indicates both kurtosis and log-kurtosis specifications. The likelihood ratios $LR_{1,\xi}$ compare specification ξ against the specification 1. Numbers in parentheses represent standard errors for estimates and the p-values for $LR_{1,\xi}$.

C.4 GARCHK-t, -H, and -GED Comparison

Table C2 Comparison of GARCHK with H, t, and GED distribution fits.

	Model	AIC	BIC	k	T
AAPL	H-kurtosis	96 210	96 305	11	39 234
	t-distribution	96 246	96 341	11	
	GED	96 567	96 661	11	
	H-distribution with $MS_{d,\tau}^{AAPL}$	96 128	96 231	12	
JPM	H-kurtosis	100 574	100 668	11	39 234
	t-distribution	100 590	100 684	11	
	GED	100 868	100 962	11	
	H-distribution with $MS_{d,\tau}^{JPM}$	100 474	100 577	12	
FDX	H-kurtosis	100 131	100 226	11	39 234
	t-distribution	100 171	100 266	11	
	GED	100 392	100 487	11	
	H-distribution with $MS_{d,\tau}^{FDX}$	100 083	100 186	12	
ODP	H-kurtosis	103 704	103 799	11	39 234
	t-distribution	103 765	103 859	11	
	GED			11	
	H-distribution with $MS_{d,\tau}^{ODP}$	103 668	103 771	12	

$$\log \sigma_d = \beta_1 \log \sigma_{d-1} + \beta_2 \left[\log RV_{d-1} - \mathbb{E}[\log RV_d]\right]$$

$$\log \sigma_{d,\tau} = \alpha_0 + \alpha_1 \log \sigma_{d,\tau-1} + \alpha_2 z_{d,\tau-1} + \alpha_3 |z_{d,\tau-1}|$$

$$\kappa_{d,\tau} = \omega_0 + \omega_1 \kappa_{d,\tau-1} + \omega_2 B \left(z_{d,\tau-1}\right)^4$$

Notes: Comparison of AIC and BIC of models for the four main analyzed stocks estimated over the period from November 1, 2017 to November 8, 2019 with five-minute returns. k and T represent the number of parameters and observations, respectively. The GED values for the ODP stock are missing, because the estimation procedure failed to find a meaningful optimum. All models have quartic specification. The H-distribution with $MS_{d,\tau}^{j}$ has an additional $\omega^{MS} MS_{d,\tau}^{j}$ term in the kurtosis specification.

C.5 GARCHK-H Estimates: Lower Frequencies

Table C3 Model estimates under various frequencies.

	AAPL			JPM		
	5 min	15 min	30 min	5 min	15 min	30 min
μ	0.002	0.007	0.009	0.000	-0.006	-0.003
	(0.003)	(0.006)	(0.009)	(0.004)	(0.007)	(0.010)
α_0	-0.068	-0.102	-0.183	-0.062	-0.049	-0.055
	(0.003)	(0.010)	(0.028)	(0.003)	(0.005)	(0.008)
α_1	0.978	0.933	0.817	0.977	0.979	0.970
	(0.002)	(0.009)	(0.035)	(0.002)	(0.004)	(0.008)
α_2	-0.012	-0.019	-0.026	-0.010	-0.014	-0.015
	(0.001)	(0.003)	(0.006)	(0.001)	(0.002)	(0.003)
α_3	0.060	0.070	0.087	0.058	0.045	0.047
	(0.003)	(0.007)	(0.014)	(0.003)	(0.004)	(0.007)
β_1	0.798	0.690	0.604	0.810	0.903	0.906
	(0.032)	(0.047)	(0.053)	(0.031)	(0.034)	(0.046)
β_2	0.177	0.274	0.373	0.166	0.083	0.070
	(0.026)	(0.040)	(0.048)	(0.025)	(0.028)	(0.034)
ω_0	1.239	1.707	2.963	1.359	1.969	2.600
	(0.288)	(0.625)	(2.207)	(0.436)	(0.536)	(1.111)
ω_1	0.708	0.635	0.464	0.664	0.490	0.378
	(0.066)	(0.132)	(0.191)	(0.106)	(0.136)	(0.126)
ω_2	0.008	0.010	0.011	0.006	0.017	0.016
	(0.002)	(0.002)	(0.001)	(0.002)	(0.006)	(0.010)

$$\log \sigma_d = \beta_1 \log \sigma_{d-1} + \beta_2 \left[\log RV_{d-1} - \mathbb{E}[\log RV_d]\right]$$

$$\log \sigma_{d,\tau} = \alpha_0 + \alpha_1 \log \sigma_{d,\tau-1} + \alpha_2 z_{d,\tau-1} + \alpha_3 |z_{d,\tau-1}|$$

$$\kappa_{d,\tau} = \omega_0 + \omega_1 \kappa_{d,\tau-1} + \omega_2 B \left(z_{d,\tau-1}\right)^4$$

Notes: Comparison of estimates for the period from November 1, 2017 to November 8, 2019, various frequencies. Quartic specification with the H-distribution. Standard errors are reported in parentheses.

C.6 GARCHK-H Estimates: Non-Exponential Volatility

Table C4 Comparison of non-exponential GARCHK model with H-distribution and Student's t-distribution.

	AAPL			JPM		
	H	H-constant	t	H	H-constant	t
μ	0.001	0.002	0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
α_0	0.009	0.008	0.010	0.014	0.014	0.014
	(0.001)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
α_1	0.902	0.903	0.894	0.904	0.903	0.904
	(0.004)	(0.002)	(0.004)	(0.005)	(0.004)	(0.004)
α_2	0.051	0.050	0.054	0.049	0.050	0.049
	(0.003)	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)
β_1	0.816	0.818	0.819	0.845	0.845	0.845
	(0.022)	(0.034)	(0.026)	(0.024)	(0.024)	(0.024)
β_2	0.167	0.167	0.168	0.141	0.141	0.141
	(0.019)	(0.027)	(0.023)	(0.025)	(0.024)	(0.020)
ω_0	4.426	4.570	4.421	4.138	4.145	4.085
	(1.242)	(0.823)	(2.053)	(1.932)	(0.668)	(0.631)
ω_1	0.001	-	0.000	0.000	-	0.001
	(0.027)		(0.056)	(0.016)		(0.015)
ω_2	0.000	-	0.001	0.000	-	0.000
	(0.001)		(0.001)	(0.002)		(0.001)
LR: $\omega_1 = 0$, $\omega_2 = 0$	0.926		0.049	0.337		0.293

$$\log \sigma_d = \beta_1 \log \sigma_{d-1} + \beta_2 \left[\log RV_{d-1} - \mathbb{E}[\log RV_d]\right]$$

$$\sigma_{d,\tau} = \alpha_0 + \alpha_1 \sigma_{d,\tau-1} + \alpha_2 z_{d,\tau-1}$$

$$\kappa_{d,\tau} = \omega_0 + \omega_1 \kappa_{d,\tau-1} + \omega_2 B \left(z_{d,\tau-1}\right)^4$$

Notes: Estimates for Apple and JP Morgan based on the period from November 1, 2017 to November 8, 2019, with five-minute returns. Standard errors are in parenthesis. The $\chi^2(1)$ critical values for the LR test are 5.991 and 9.210 for 5% and 1% levels, respectively.