A Coupled Component DCS-EGARCH Model for Intraday and Overnight Volatility

Oliver Linton∗ Jianbin Wu†
University of Cambridge Nanjing University

December 6, 2019

Abstract

We propose a semi-parametric coupled component exponential GARCH model for intraday and overnight returns that allows the two series to have different dynamical properties. We adopt a dynamic conditional score model with t-distributed innovations that captures the very heavy tails of overnight returns. We propose a several-step estimation procedure that captures the nonparametric slowly moving components by kernel estimation and the dynamic parameters by maximum likelihood. We establish the consistency, asymptotic normality, and semiparametric efficiency of our semiparametric estimation procedures. We extend the modelling to the multivariate case where we allow time varying correlation between stocks. We apply our model to the study of Dow Jones industrial average component stocks and CRSP size-based portfolios over the period 1993-2017. We show that the ratio of overnight to intraday volatility has actually increased in importance for Dow Jones stocks during the last two decades. This ratio has also increased for large stocks in the CRSP database, but decreased for small stocks in CRSP.

Keywords: DCS, GAS, GARCH, size-based portfolios, Testing
JEL Classification: C12, C13

1 Introduction

The balance between intraday and overnight returns is of considerable interest as it potentially sheds light on many issues in finance: the efficient markets hypothesis, the calendar time versus trading time models, the process by which information is impacted into stock prices, the relative merits of auction versus continuous trading, the effect of high frequency trading on market quality, and the globalization and connectedness of international markets. We propose a bivariate time series model for intraday and overnight returns that respects their temporal ordering and permits the two

∗Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: obl20@cam.ac.uk.
†Corresponding author. School of Economics, Nanjing Universtiy, Nanjing, 210093, China. Email: wujianbin@nju.edu.cn.
processes to have different marginal properties, and to feedback into each other, and allows for both short run and long components. In particular, our volatility model for each return series has a long run component that slowly evolves over time, and is treated nonparametrically, and a parametric dynamic volatility component that allows for short run deviations from the long run process, where those deviations depend on previous intraday and overnight shocks. We adopt a dynamic conditional score (DCS) model, Harvey (2013) and Harvey and Luati (2014), that links the news impact curves of the innovations to the shock distributions, which we assume to be t-distributions with unknown degrees of freedom (which may differ between intraday and overnight). In practice, the overnight return distribution is more heavy tailed than the intraday return, and in fact very heavily tailed. Our model allows for a difference in the tail thickness of the conditional distributions. The short run dynamic process allows for leverage effects and separates the overnight shock from the intraday shock. We also introduce a multivariate model that allows for time varying correlations between stocks and between overnight and intraday returns.

We apply our model to the study of 26 Dow Jones industrial average component stocks over the period 1993-2017, a period that saw several substantial institutional changes. There are several purposes for our application. First, many authors have argued that the introduction of computerized trading and the increased prevalence of High Frequency Trading (HFT) strategies in the period post 2005 has lead to an increase in volatility, see Boehmer, Fong, and Wu (2015) and Linton, O’Hara, and Zigrand (2013). To address this, a direct comparison of volatility before and after would be problematic here because of the Global Financial Crisis (GFC), which raised volatility during the same period that HFT was becoming more prevalent. There are a number of studies that have investigated this question with natural experiments methodology (Brogaard, 2011), but the conclusions one can draw from such work are event specific. We model the volatility process with a view to addressing this hypothesis in a more general way. One implication of this hypothesis is that ceteris paribus the ratio of intraday to overnight volatility should have increased during this period because trading is not taking place during the market close period. We would like to evaluate whether this has occurred. One could just compare the daily return volatility from the intraday segment with the daily return volatility from the overnight segment, as many studies such as French and Roll (1986) have done. However, this would ignore both fast and slow variation in volatility through business cycle and other causal factors. Also, overnight raw returns are very heavy tailed and so sample (unconditional) variances are not very reliable. We use our dynamic two component model, which allows for both fast and slow dynamic components to volatility, as is now common practice (Engle and Lee 1999, Engle and Rangel, 2008, Hafner and Linton, 2010, Rangel and Engle 2012, and Han and Kristensen, 2015). Our model also allows dynamic feedback between overnight and intraday volatility, which is of interest in itself. Our model generates heavy tails in observed returns, but the parameter estimates we employ are robust to this phenomenon. Our methodology therefore allows us to compare the long run components of volatility over this period without over reliance on Gaussian-type theory. We show that for the Dow Jones stocks the long run component of overnight volatility has actually increased in importance during this period relative to the long run component of intraday volatility (although intraday volatility is still generally higher
than overnight volatility). We provide a formal test statistic that confirms quantitatively the strength of this effect; our test can be interpreted as carrying out a difference in difference analysis but in ratio form. Imbens and Wooldridge (2007). This finding seems to be hard to reconcile with the view that trading has increased volatility. We also document the short run dynamic processes. Notably, we find, unlike Blanc, Chicheportiche, and Bouchaud (2014), that overnight returns significantly affect future intraday volatility. We also find that overnight return shocks have t-distributions with degrees of freedom roughly equal to three, which emphasizes the potential fragility of Gaussian-based estimation routines that earlier work has been based on. We also estimate a multivariate model and document that there has been an upward trend in the long run component of contemporary overnight correlation between stocks as well as in the long run component of contemporary intraday correlation between stocks. However, the trend development for the overnight correlations started later than for intraday, and started happening only after 2005, whereas the intraday correlations appear to have slowly increased more or less from the beginning of the period.

We also apply our model to size-sorted portfolios of CRSP stocks over the period 1993-2017. We find that the ratio of overnight to intraday volatility has indeed increased for large stocks, but has decreased for small stocks especially in the 1990s. Notably, the slope increases monotonically from the smallest-cap to the largest-cap decile, and the ratio of overnight to intraday volatility is typically high during recent crises.

A second practical purpose for our model is to improve forecasts of intraday volatility or close to close volatility. Our model allows us to condition on the open price to forecast intraday volatility or to update the close to close volatility forecast and also to take into account the full dynamic consequences of the overnight shock and previous ones. We compare forecast performance of our model with a procedure based only on close to close returns and find in most cases superior performance.

We work only with the return series, although for some stocks intraday transaction and quote records are available for the duration of our study, which would permit the computation of realized volatility measures, which are for some the preferred measure of intraday volatility. This however would pose some additional questions in terms of the joint modelling of discrete time returns and realized volatility, and puts an imbalance between the measurement of intraday and overnight volatility. Furthermore, it would be problematic to implement some of those techniques on the small CRSP stocks in the early part of the sample period, so it is not a silver bullet. Instead we do make use of alternative measures of market (SP500) volatility - the VIX (which includes overnight volatility) and the Rogers and Satchell (1991) intraday volatility measure - to conduct a robustness check. We find that these measures confirm the finding regarding the rise of overnight volatility relative to intraday after 2004.

Related Literature. Overnight returns have recently attracted much attention in empirical finance. Many find overnight and intraday returns behave entirely differently, and overnight returns tend to outperform intraday returns. Specifically, Cooper, Cliff, and Gulen (2008) suggest that the U.S. equity premium over the period 1993-2006 is solely due to overnight returns. Kelly and Clark

---

1Our main findings involve averages of the daily volatility series and so the efficiency gain of realized volatility may not be so large in this context.
find the overnight returns are on average larger than the intraday returns. Berkman, Koch, Tuttle, and Zhang (2012) find a significant positive mean overnight return and a significant negative mean intraday return. They suggest stocks that have recently attracted the attention of retail investors tend to have higher net retail buying at the open, leading to high overnight returns that followed by intraday reversals. Aboody, Even-Tov, Lehavy, and Trueman (2018) suggest overnight returns can serve as a measure of firm-specific investor sentiment, and find short-term persistence in overnight returns. Polk, Lou, and Skouros (2019) link investor heterogeneity to the strong persistence of the overnight and intraday returns. They find an overnight versus intraday tug of war in strategy risk premium, and the risk premium is earned entirely overnight for the largest stocks. Besides the difference in expected returns, overnight returns are found less volatile (French and Roll 1986; Lockwood and Linn 1990; Aretz and Bartram 2015), but more leptokurtic than intraday returns in the U.S. market (Ng and Masulis 1995; Blanc, Chicheportiche, and Bouchaud 2014).

Tsiakas (2008) proposed a stochastic volatility model for daytime returns with feedback from night to day and leverage effects built in. He assumed Gaussian innovations; he did not model the overnight returns. In the literature on realized volatility, many authors have considered how to incorporate overnight returns into daily variance modeling and forecasting, by scaling the intraday measure (e.g., Martens 2002 and Fleming, Kirby, and Ostdieck 2003), or by combining daytime realized volatility and the squared overnight return with optimally chosen weight parameters (e.g., Hansen and Lunde 2005). However, these authors also did not model the overnight returns either. Andersen, Bollerslev, and Huang (2011) decomposed the total daily return variability into the continuous sample path variance, the discontinuous intraday jumps, and the overnight variance. For this overnight variance, they used an augmented GARCH-t type structure with the immediately preceding daytime realized volatility as an additional explanatory variable. Blanc, Chicheportiche, and Bouchaud (2014) employ a quadratic ARCH model with flexible dynamics for both intraday and overnight returns; they also allow for feedback from overnight to intraday returns and leverage effects. They use a t distributed shock to drive each process and to define an estimation algorithm. They impose a pooling assumption on the model parameters across 280 S&P500 stocks that are continually in the index over 2000-2009, and assume stationarity over the period in question.

Our paper is closely related to the Generalized Autoregressive Score models or the Beta-t-(E)GARCH model. Creal, Koopman, and Lucas (2013) introduced a general class of time series models called Generalized Autoregressive Score models (GAS). Simultaneously, Harvey and Chakravarty (2008) developed a score driven model specifically for volatilities, called the Beta-t-(E)GARCH model, built on exactly the same philosophy. Harvey (2013) settles on the dynamic conditional score model terminology, and we follow that nomenclature. This paper is also related to the work of Engle and Rangel (2008) and Hafner and Linton (2010) about incorporating long run volatilities. Engle and Rangel (2008) introduced nonparametric slowly varying trends into GARCH models; Hafner and Linton (2010) propose a multivariate extension and develop the distribution theory for inference.
# The Model and its Properties

We let $r^D_t$ denote intraday returns and $r^N_t$ denote overnight returns on day $t$. We take the ordering that night precedes day so that $r^D_t = \ln(P^C_t / P^O_t)$ and $r^N_t = \ln(P^O_t / P^C_{t-1})$, where $P^O_t$ denotes the open price on day $t$ and $P^C_t$ denotes the close price on day $t$. Daily close to close returns satisfy $r_t = r^D_t + r^N_t$. The timeline is illustrated below:

$$
\ldots \rightarrow P^C_{t-1} \rightarrow P^O_t \rightarrow P^C_t \rightarrow \ldots
$$

<table>
<thead>
<tr>
<th>Night $t$</th>
<th>Day $t$</th>
<th>Night $t+1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r^N_t$</td>
<td>$r^D_t$</td>
<td></td>
</tr>
<tr>
<td>$r^N_{t+1}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We do not distinguish between weekend, holiday weekends, and ordinary midweek overnight periods, although we comment on this issue in the concluding section below.

Our model allows intraday returns to depend on overnight returns with the same $t$, but overnight returns just depend on lagged variables. We suppose that

$$
\begin{pmatrix}
1 & \delta \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
r^D_t \\
r^N_t
\end{pmatrix}
= \begin{pmatrix}
\mu_D \\
\mu_N
\end{pmatrix}
+ \Pi
\begin{pmatrix}
r^D_{t-1} \\
r^N_{t-1}
\end{pmatrix}
+ \begin{pmatrix}
u^D_t \\
u^N_t
\end{pmatrix},
\tag{1}
$$

where $u^D_t$ and $u^N_t$ are conditional mean zero shocks. Under the EMH, $\delta = 0$ and $\Pi = 0$, but we allow these coefficients to be nonzero to pick up what could be misspricing effects or short run effects such as might arise from the market microstructure, [Scholes and Williams (1977)].

We suppose that the error process has conditional heteroskedasticity, with both long run and short run effects, both modelled in exponential form, [Nelson (1991)]. Specifically, we suppose that

$$
u_t = \begin{pmatrix}
u^D_t \\
\nu^N_t
\end{pmatrix} = \begin{pmatrix}
\exp(\lambda^D_t) \exp(\sigma^D(t/T) \varepsilon^D_t) \\
\exp(\lambda^N_t) \exp(\sigma^N(t/T) \varepsilon^N_t)
\end{pmatrix},
\tag{2}
$$

where $\varepsilon^D_t$ and $\varepsilon^N_t$ are i.i.d. mean zero shocks from $t$ distributions with $v^D$ and $v^N$ degrees of freedom, respectively, while $\sigma^D(\cdot)$ and $\sigma^N(\cdot)$ are unknown but smooth functions that will represent the slowly varying (long-run) scale of the process, and $T$ is the number of observations. Suppose that for $j = D, N$:

$$
\sigma^j(s) = \sum_{i=1}^{\infty} \theta^j_i \psi^j_i(s), \quad s \in [0, 1]
\tag{3}
$$

for some orthonormal basis $\{\psi^j_i(s)\}_{i=1}^{\infty}$ with $\int_0^1 \psi^j_i(s) \, ds = 0$ and

$$
\int \psi^j_i(s) \psi^j_k(s) \, ds = \begin{cases}
1 & \text{if } i = k \\
0 & \text{if } i \neq k.
\end{cases}
$$

We suppose $\sigma^D(\cdot)$ and $\sigma^N(\cdot)$ integrate to zero to achieve identification. It is similar to the [Hafner and Linton (2010)] normalization. The only difference is that we are normalizing the log variance to integrate to zero for convenience because our model is in exponential form. This normalization also has the advantage of delivering orthogonality between the score for the parameters in the short run.
component and the score for \( \theta \), as we will see later in Theorem 1, see Linton (1993). In the following, \( j \) is always used to denote \( D, N \) without further mentioning.

Regarding the short run dynamic part of (2), we adopt a dynamic conditional score approach, Creal, Koopman, and Lucas (2011) and Harvey and Luati (2014). The conditional (scale) score

\[
m_i^t = \frac{(1 + v_j)(e_i^t)^2}{v_j \exp(2\lambda_i^t) + (e_i^t)^2} - 1, \quad v_j > 0
\]

where \( e_i^t = \exp(-\sigma^d(t/T))u_i^t \). Note that \( m_i^t \) is a bounded function of \( e_i^t \) with \( E(m_i^D|F_{t-1}^N) = 0 \) and \( E(m_i^N|F_{t-1}^D) = 0 \), where \( F_{t-1}^N \) is the information set at the open of day \( t-1 \) and \( F_{t-1}^D \) is the information set at the close of day \( t-1 \). We suppose that \( \lambda_i^D \) and \( \lambda_i^N \) are linear combinations of past values of the shocks determined by \( m_i^t \), \( j = D, N \):

\[
\lambda_i^D = \omega_D(1 - \beta_D) + \beta_D \lambda_{i-1}^D + \gamma_D m_{i-1}^D + \rho_D m_i^N
+ \gamma^*_D(m_{i-1}^D + 1)\text{sign}(e_{i-1}^D) + \rho^*_D(m_i^N + 1)\text{sign}(e_i^N)
\]

\[
\lambda_i^N = \omega_N(1 - \beta_N) + \beta_N \lambda_{i-1}^N + \gamma_N m_{i-1}^N + \rho_N m_i^D
+ \gamma^*_N(m_{i-1}^N + 1)\text{sign}(e_{i-1}^D) + \rho^*_N(m_i^D + 1)\text{sign}(e_i^N)
\]

We suppose that this stochastic process has a compatible initialization, for simplicity we assume below that the process started in the infinite past. This gives two dynamic processes for the short run scale of the overnight and intraday return. The parameters \( \rho_D, \rho^*_D \) capture the effect of overnight shocks on intraday volatility, while \( \rho_N, \rho^*_N \) capture the effects of intraday shocks on overnight volatility; we call \( \rho_D, \rho^*_D, \rho_N, \rho^*_N \) feedback parameters that couple together the processes \( \lambda_i^D, \lambda_i^N \), whereas \( \gamma_D, \gamma^*_D, \gamma_N, \gamma^*_N \) are capturing the effect of shocks from previous same type of period on same type of period volatility. We allow for leverage effects, Nelson (1991) and Glosten, Jagannathan, and Runkle (1993), through the parameters \( \gamma^*_D, \rho^*_D, \rho^*_N, \) and \( \gamma^*_N \). The parameters \( \beta_D, \beta_N \) measure the persistence of the volatility processes. We set the intercepts this way so that \( \omega_D \) is the unconditional mean of \( \lambda_i^D \) and \( \omega_N \) is the unconditional mean of \( \lambda_i^N \); we may consider \( \exp(\omega_D - \omega_N) \) to measure the relative mean volatility contribution of the daily process and the overnight process. Let

\[
\phi = (\omega_D, \beta_D, \gamma_D, \gamma^*_D, \rho_D, \rho^*_D, v_D, \omega_N, \beta_N, \gamma_N, \gamma^*_N, \rho_N, \rho^*_N, v_N)^T \in \Phi \subset \mathbb{R}^{14}
\]

be the finite dimensional parameters of interest. The two unknown functions \( \sigma^D(\cdot) \) and \( \sigma^N(\cdot) \) complete the semiparametric model for the process \( \{u_i\} \).

Harvey (2013) argues that the quadratic innovations that feature in GARCH models naturally fit with the Gaussian distribution for the shock, but once one allows heavier tail distributions like the t-distribution, it is anomalous to or not obvious why to focus on quadratic innovations, and indeed this focus leads to a lack of robustness because large shocks are fed substantially into the volatility update. He argues that it is more natural to link the shock to volatility to the distribution of the rescaled return shock, which in the case of the t distribution has the advantage that large shocks are automatically

\footnote{The shock variable \( m_i^t \) can be expressed as \( m_i^t = (v_j + 1)b_i^t - 1 \), where \( b_i^t \) has a beta distribution, \textit{beta}(1/2, \text{v}_j/2).}
down weighted, and in such a way driven by the shape of the error distribution. The DCS model has the incidental advantage that there are analytic expressions for moments, autocorrelation functions, multi-step forecasts, and the mean squared forecast errors.

Before introducing our estimation procedure we comment on some properties of our model that are useful in applications and in theoretical understanding. In the Supplementary Material contained in [Linton and Wu (2018)] we prove that if $|\beta_j| < 1$, $j = D, N$, then $\epsilon_t^j$ and $\lambda_t^j$ are strongly stationary and $\beta$-mixing with exponential decay. [He, Teräsvirta, and Malmsten (2002)] derive formulae for moments of a family of exponential GARCH models, and similar calculations can be reproduced here. We discuss next the invertibility of the process $(\lambda_t^D, \lambda_t^N)^\top$. This property is important for the asymptotic theory. To establish invertibility, we write the dynamics of $(\lambda_t^D, \lambda_t^N)^\top$ as a stochastic recurrence equation (SRE) in terms of $\epsilon_t$,

$$
\begin{pmatrix}
\lambda_{t+1}^D \\
\lambda_{t+1}^N
\end{pmatrix} = f_t \begin{pmatrix}
\left( \begin{array}{c}
\lambda_t^D \\
\lambda_t^N
\end{array} \right), \epsilon_t
\end{pmatrix}
$$

for some mapping $f_t$. Following [Straumann and Mikosch (2006)] and [Wintenberger (2013)], the model is invertible if this SRE is stable so that the effect of any initialization $(\lambda_0^D, \lambda_0^N)^\top$ vanishes asymptotically at an exponential rate. Sufficient conditions for invertibility are the contraction condition and the existence of logarithmic moments [Bougerol (1993)]. The existence of logarithmic moments can be easily obtained since $\lambda_t^D, \lambda_t^N$ and their scores are functions of bounded variables $m_{t-k}$. For the contraction condition, it is easy to have

$$
\begin{align*}
\frac{\partial \lambda_t^D}{\partial \lambda_{t-1}^D} &= \beta_D + (\gamma_D + \gamma_D^* \text{sign}(u_{t-1}^D)) \frac{\partial m_{t-1}^D}{\partial \lambda_{t-1}^D} + (\rho_D + \rho_D^* \text{sign}(u_{t}^N)) \frac{\partial m_{t}^N}{\partial \lambda_{t}^N} \frac{\partial \lambda_{t-1}^D}{\partial \lambda_{t-1}^N} \\
&= \beta_D + a_{t-1}^{DD} + a_{t-1}^{DN} a_{t-1}^{ND}, \\
\frac{\partial \lambda_t^N}{\partial \lambda_{t-1}^N} &= \beta_D + a_{t-1}^{DD} + a_{t}^{DN} a_{t-1}^{ND}, \quad \frac{\partial \lambda_t^D}{\partial \lambda_{t-1}^N} = a_{t}^{DN} (\beta_N + a_{t-1}^{NN}), \quad \frac{\partial \lambda_t^N}{\partial \lambda_{t-1}^D} = a_{t}^{ND}.
\end{align*}
$$

The Jacobian matrix of the mapping $f_t$ is thus

$$
\begin{bmatrix}
\beta_D + a_{t-1}^{DD} + a_t^{DN} a_{t-1}^{ND} & a_t^{DN} (\beta_N + a_{t-1}^{NN}) \\
a_{t-1}^{ND} & \beta_N + a_{t-1}^{NN}
\end{bmatrix}
= A_t B_{t-1},
$$

for certain matrices $A_t, B_t$ defined below in (11).

Applying Theorem 3.1 in [Bougerol (1993)], if for some integer $p \geq 1$, $E \log(\sup_{\lambda_t^D, \lambda_t^N} || \prod_{i=1}^{p} A_{p-i+1} B_{p-i} ||_\infty) < 0$ holds, then $(\lambda_t^D, \lambda_t^N)^\top$ is invertible (the norm $||.||_\infty$ is defined below in (12)). Taking $p = 1$, it is sufficient to have $E \log(\sup || A_1 B_0 ||_\infty) < 0$. When $\rho_D, \rho_D^*, \rho_N, \rho_N^* = 0$, this condition becomes $|\beta_D + a_{t-1}^{DD}| < 1$ and $|\beta_N + a_{t-1}^{NN}| < 1$, equivalent to that in [Harvey and Lange (2018)] for the univariate model. Otherwise, our condition is in general more restrictive than theirs. It is possible that $E \log(\sup || A_1 B_0 ||_\infty) < 0$ does not hold in some cases, in which case we should take larger $p$. The

---

\footnote{This type of argument is similar to the argument in limited dependent variable models such as binary choice where a linear function of covariates is connected to the observed outcome by a link function determined by the distributional assumption.}
condition is often satisfied, e.g., with \( p = 2 \). With invertibility, the assumption that \( h^j_t \) starts from the infinite past in assumption A.3 below can be relaxed.

Finally, we note that although the conditional distribution of returns is symmetric about the mean, the unconditional distribution implied by our model may be asymmetric because of the conditional mean process and the asymmetric news impact curve that we allow for. He, Silvennoinen, and Teräsvirta \citeyear{2008}. Thiele \citeyear{2019} consider some DCS models with asymmetric t-distributions, which might be a possible direction for future work.

\section{Estimation}

Suppose that we know \( \delta, \mu, \Pi \) and hence \( u^j_t, j = D, N \). In practice these can be replaced by root-\( T \) consistent estimators, and we shall not detail the properties of the mean estimators in the sequel as these are well known, and we shall drop them from the notation for convenience in the theoretical analysis. We next describe how we estimate the unknown quantities \( \phi \) and \( \sigma^j(\cdot) \). For any \( \alpha > 0 \), we have for \( j = N, D \),

\[
E \left( \left| u^j_t \right|^\alpha \right) = E \left( \left| \varepsilon^j_t \right|^\alpha \right) E \left( \exp(\alpha \lambda^j_t) \exp(\alpha \sigma^j(t/T)) = c^j(\phi; \alpha) \times \exp(\alpha \sigma^j(t/T)),
\]

where \( c^j \) is a constant that depends in a complicated way on the parameter vector \( \phi \) and on \( \alpha \).

Therefore, we can estimate \( \sigma^N(s), \sigma^D(s) \) as follows with kernel technology. Let \( K(u) \) be a kernel with support \([-1, 1]\) and \( h \) a bandwidth, and let \( K_h(.) = K(.-/h)/h \). Then let

\[
\tilde{\sigma}^j(s) = \frac{1}{\alpha} \log \left( \frac{1}{T} \sum_{t=1}^{T} K_h(s - t/T) |u^j_t|^\alpha \right) \quad (7)
\]

for any \( s \in (0, 1) \). In fact, we employ a boundary modification for \( s \in [0, h] \cup [1 - h, 1] \), whereby \( K \) is replaced by a boundary kernel, which is a function of two arguments \( K(u, c) \), where the parameter \( c \) controls the support of the kernel; thus left boundary kernel \( K(u, c) \) with \( c = s/h \) has support \([-1, c] \) and satisfies \( \int_{-1}^{c} K(u, c)du = 1 \), \( \int_{-1}^{c} u K(u, c)du = 0 \), and \( \int_{-1}^{c} u^2 K(u, c)du < \infty \). Similarly for the right boundary. The purpose of the boundary modification is to ensure that the bias property holds throughout \([0, 1]\). Härdle and Linton \citeyear{1994}. One may apply more sophisticated adjustments such as Jones, Linton, and Nielsen \citeyear{1995} that preserves positivity but reduces the bias in the boundary region. For identification, we recenter \( \tilde{\sigma}^j(t/T) \) as

\[
\tilde{\sigma}^j(t/T) = \tilde{\sigma}^j(t/T) - \frac{1}{T} \sum_{t=1}^{T} \tilde{\sigma}^j(t/T).
\]  

Note that \( \tilde{\sigma}^j(u) \) can be written as \( \tilde{\sigma}^j(u) = \sum_{i=1}^{\infty} \tilde{\theta}^j_i \psi^j_i(s) \) for some coefficients \( \tilde{\theta}^j_i \) (that satisfy \( \sum_{i=1}^{\infty} |\tilde{\theta}^j_i| < \infty \)) determined uniquely by the estimator \( \tilde{\sigma}^j(u) \), that is, we can represent the kernel estimator as a sieve estimator with a potentially infinite number of coefficients, see Linton and Wu \citeyear{2018}. We will use this representation for notational convenience, that is, we will represent \( \tilde{\sigma}^j(.) \) in terms of \( \{\tilde{\theta}^j_i\}_{i=1}^{\infty} \) or just \( \tilde{\theta} \) for shorthand. In practice, the bandwidth may be chosen by some rule of thumb method.
Let \( \tilde{c}_t^N = \exp(-\tilde{\sigma}_t^N(t/T))u_t^N \) and \( \tilde{c}_t^D = \exp(-\tilde{\sigma}_t^D(t/T))u_t^D \), and let \( \tilde{\theta} \) denote \( \{\tilde{\sigma}^j(s), s \in [0, 1], j = N, D\} \). Define the global log-likelihood function for \( \phi \) (apart from an unnecessary constant and conditional on the estimated values of \( \theta \))

\[
 l_T(\phi; \tilde{\theta}) = \frac{1}{T} \sum_{t=1}^T l_t(\phi; \tilde{\theta}) = \frac{1}{T} \sum_{t=1}^T \left( l_t^N(\phi; \tilde{\theta}) + l_t^D(\phi; \tilde{\theta}) \right),
\]

\[
 l_t^j(\phi; \tilde{\theta}) = -\lambda_t^j(\phi; \tilde{\theta}) - \frac{v_j + 1}{2} \ln \left( 1 + \frac{(\tilde{c}_t^j)^2}{v_j \exp(2\lambda_t^j(\phi; \tilde{\theta}))} \right) + \ln \Gamma \left( \frac{v_j + 1}{2} \right) - \frac{1}{2} \ln v_j - \ln \Gamma \left( \frac{v_j}{2} \right),
\]

(9)

where \( \Gamma \) is the gamma function and \( \lambda_t^j(\phi; \tilde{\theta}) \) are defined in \( (5) \) and \( (6) \). For practical purposes, \( \lambda_{10}^j \) may be set equal to the unconditional mean, \( \lambda_{10}^j = \omega_j \). We estimate \( \phi \) by maximizing \( l_T(\phi; \tilde{\theta}) \) with respect to \( \phi \in \Phi \). Let \( \tilde{\phi} \) denote these estimates.

Given estimates of \( \phi \) and the preliminary estimates of \( \sigma^D(\cdot), \sigma^N(\cdot) \), we calculate

\[
 \tilde{\eta}_t^N = \exp(-\tilde{\lambda}_t^N)u_t^N; \quad \tilde{\eta}_t^D = \exp(-\tilde{\lambda}_t^D)u_t^D,
\]

where \( \tilde{\lambda}_t^j = \lambda_t^j(\phi; \tilde{\theta}) \). We then update the estimates of \( \sigma^D(\cdot), \sigma^N(\cdot) \) with the local likelihood function in \cite{Severini:1992} given \( \tilde{\eta}_t^j \) and \( \tilde{v}_j \), i.e., we maximize the objective function

\[
 \tilde{L}_T(\gamma; \tilde{\lambda}^j, s) = -\frac{1}{T} \sum_{i=1}^T K_h(s - t/T) \left[ \gamma + \frac{\tilde{v}_j + 1}{2} \ln \left( 1 + \frac{\tilde{\eta}_t^j \exp(-\gamma)^2}{\tilde{v}_j} \right) \right]
\]

(10)

with respect to \( \gamma \in \mathbb{R} \), for \( j = D, N \) separately, where \( \tilde{\lambda}^j = (\tilde{\lambda}_1^j, \ldots, \tilde{\lambda}_T^j)^\top \). Here, we also use a boundary kernel for \( s \in [0, h] \cup [1 - h, 1] \). In practice we use Newton-Raphson iterations making use of the analytic derivatives of the objective functions, which are given in \( (11) \) in the Supplementary material of \cite{Linton:2013, Harvey:2013} gives some discussion about computational issues.

To summarize, the estimation algorithm is as follows.

**Algorithm**

**Step 1.** Estimate \( \delta, \mu^j, \Pi \) by least squares and \( \tilde{\sigma}^j(u), u \in [0, 1], j = N, D \) from \( (7) \) and \( (8) \).

**Step 2.** Estimate \( \phi \) by optimizing \( l_T(\phi; \tilde{\theta}) \) with respect to \( \phi \in \Phi \) (by Newton-Raphson) to give \( \tilde{\phi} \).

**Step 3.** Given the initial estimates \( \tilde{\theta} \) and \( \tilde{\phi} \), we replace \( \lambda_t^j \) with \( \tilde{\lambda}_t^j = \lambda_t^j(\tilde{\phi}; \tilde{\theta}) \). Then let \( \tilde{\sigma}^j(s) \) optimize \( \tilde{L}_T(\sigma^j(s); \tilde{\lambda}, s) \) with respect to \( \sigma^j(s) \). Rescale \( \tilde{\sigma}^j(t/T) = \sigma^j(t/T) - \frac{1}{T} \sum_{i=1}^T \hat{\sigma}^j(t/T) = \sum_{i=1}^\infty \hat{\theta}_t^j \psi_t^j(s) \). Update \( \phi \) by optimizing \( l_T(\phi; \tilde{\theta}) \) with respect to \( \phi \in \Phi \) to give \( \hat{\phi} \).

**Step 4.** Repeat Steps 2-3 to update \( \hat{\theta} \) and \( \hat{\phi} \) until convergence. We define convergence in terms of the distance measure

\[
 \Delta_r = \sum_{j=D,N} \int \left[ \tilde{\sigma}_{[r]}(u) - \tilde{\sigma}_{[r-1]}(u) \right]^2 du + \left( \hat{\phi}_{[r]} - \hat{\phi}_{[r-1]} \right)^\top \left( \hat{\phi}_{[r]} - \hat{\phi}_{[r-1]} \right),
\]

that is, we stop when \( \Delta_r \leq \epsilon \) for some prespecified small \( \epsilon \).
4 Large Sample Properties of Estimators

In this section we give the asymptotic distribution theory of the estimators considered above. The proofs of all results are given in Linton and Wu (2018). Let \( h'_t = \lambda'_t + \sigma'(t/T) \), and let:

\[
A_t = \begin{bmatrix}
1 & a_{t}^{DN} \\
0 & 1
\end{bmatrix}, \quad B_{t-1} = \begin{bmatrix}
(\beta_D + a_{t-1}^{DP}) & 0 \\
0 & (\beta_N + a_{t-1}^{NN})
\end{bmatrix},
\]

We use the maximum row sum matrix norm, \( \| \cdot \|_\infty \), defined by

\[
\| A \|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{ij}|.
\]

**Assumptions A**

1. \( \| E (A_t \otimes A_t) \|_\infty < \infty, \| E B_t E A_t \|_\infty < 1, \| E (B_{t-1} A_{t-1} \otimes B_{t-1} A_{t-1}) \|_\infty < \| E B_t E A_t \|_\infty \), and the top-Lyapunov exponent of the sequence of \( A_t B_{t-1} \) is strictly negative. The top Lyapunov exponent is defined as Theorem 4.26 of Douc, Moulines, and Stoffer (2014).

2. \( 0 \leq |\beta_j| < 1 \).

3. \( h'_t \) starts from the infinite past. The parameter \( \phi_0 \) is an interior point of \( \Phi \subset \mathbb{R}^4 \), where \( \Phi \) is the parameter space of \( \phi_0 \).

4. The functions \( \sigma^j \) are twice continuously differentiable on \( [0, 1], j = D, N \).

5. \( E |u_t^j|^{2+\delta} < \infty \) for some \( \delta > 0 \), \( j = D, N \).

6. The function \( l(\phi) = E(l_T(\phi; \theta_0)) \) is uniquely maximized at \( \phi = \phi_0 \).

7. The kernel function \( K \) is bounded, symmetric about zero with compact support, that is \( K(s) = 0 \) for all \( |s| > C_1 \) with some \( C_1 < \infty \). Moreover, it is Lipschitz, that is \( |K(s) - K(s')| \leq L|s - s'| \) for some \( L < \infty \) and all \( s, s' \in \mathbb{R} \). Denote \( \|K\|_2 = \int K(s)^2ds \).

8. \( h(T) \to 0 \) as \( T \to \infty \) such that \( T^{1/2-\delta}h \to \infty \) for some small \( \delta > 0 \).
Assumptions A3-A7 are used to derive the properties of $\tilde{\sigma}^j(s)$, in line with Vogt and Linton (2014) and Vogt (2012). But we only require that $E|u_t^j|^\alpha < \infty$, since we use $\tilde{\sigma}^j(s) = \log(T^{-1} \sum_{t=1}^T K_h(s - t/T)|u_t^j|^\alpha)/\alpha$. This is in line with the fact that the fourth-order moment of overnight returns may not exist for some datasets. The mixing condition in Vogt and Linton (2014) is replaced by Assumption A2, because of our tight model structure.

Assumption A1 is required to derive the stationarity of score functions, where $\|E(A_t \otimes A_t)\|_\infty < \infty$ can be verified easily, since $b_N^t$ in $A_t$ follows a beta distribution.

Lemma 1 in Linton and Wu (2018) gives the uniform convergence rate of the initial estimator $\tilde{\sigma}^j(s)$, which is close to $T^{-2/5}$ when $h = O(T^{-1/5})$. The proof mainly follows Theorem 3 in Vogt and Linton (2014). We note that our initial estimator is robust to the specification of the short run dynamic process in the sense that Lemma 1 continues to hold under the weak dependence assumptions for whatever stationary mixing process is assumed for $\lambda_j^t$.

We next present an important orthogonality condition that allows us to establish a simple theory for the parametric component.

**Theorem 1** Suppose that Assumptions A1-A4 hold. Then, for each $k$ and $i$, for $k \in \{1, \ldots, \infty\}$ and $i \in \{1, \ldots, 14\}$, we have

$$
\frac{1}{T} \sum_{t=1}^T E \left[ \frac{\partial l_t(\phi_0; \theta_0)}{\partial \theta_k} \frac{\partial l_t(\phi_0; \theta_0)}{\partial \phi_i} \right] = o(T^{-1/2}).
$$

The proof of Theorem 1 is provided in Linton and Wu (2018). Theorem 1 implies that the score functions with respect to $\theta$ and $\phi$ are asymptotically orthogonal. The intuition behind this is that $\sigma^j$ is a function of a deterministic variable, $t/T$, while $\lambda_j^t$ is a stationary process independent of time $t$. The cross product of their score functions can be somehow separated, see Linton (1993) for a similar result. The asymptotic orthogonality implies that the particular asymptotic property of $\tilde{\phi}$ and $\hat{\phi}$ in Theorem 2 follows.

Define the asymptotic information matrix

$$
I(\phi_0) = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^T E \left[ \frac{\partial l_t(\phi_0; \theta_0)}{\partial \phi} \frac{\partial l_t(\phi_0; \theta_0)}{\partial \phi}^\top \right].
$$

**Theorem 2** Suppose that Assumptions A1-A8 hold. Then

$$
\sqrt{T} \left( \tilde{\phi} - \phi_0 \right) = \sqrt{T} \left( \hat{\phi} - \phi_0 \right) + o_P(1) \implies N(0, I(\phi_0)^{-1}).
$$

**Theorem 3** Suppose that Assumptions A1-A8 hold. Then for $s \in (0, 1)$

$$
\sqrt{T} h \begin{pmatrix}
\tilde{\sigma}^D(s) - \sigma_0^D(s) \\
\tilde{\sigma}^N(s) - \sigma_0^N(s)
\end{pmatrix} \implies N \left( 0, ||K||^2 \begin{pmatrix}
\frac{1}{2} + \frac{3}{2\nu_D} & 0 \\
0 & \frac{1}{2\nu_N} + \frac{3}{2\nu_N}
\end{pmatrix} \right). \tag{13}
$$

Theorem 2 shows that $\tilde{\phi}$ and $\hat{\phi}$ have the same asymptotic property and are efficient. The form of the limiting variance in (13) is consistent with the known Fisher information for the estimation of scale parameters of a t-distribution with known location and degrees of freedom (these quantities
are estimated at a faster rate), which makes this part of the procedure also efficient in the sense considered in [Tibshirani (1984)]. The proofs of Theorem 2 and 3 are provided in [Linton and Wu (2018)]. The information matrix, $\mathcal{I}(\phi_0)$, can be computed explicitly. We can conduct inference with Theorem 2 and Theorem 3 using plug-in estimates of the unknown quantities. In the application we present various Wald statistics for testing hypotheses about $\phi$ such as: the absence of leverage effects, the absence of feedback effects, and the equality of intraday and overnight parameters.

**Test of constancy of the ratio of long run components.** We next provide a test of the constancy of the ratio of long run overnight to intraday volatility. We consider the null hypothesis to be

$$H_0: \exp(\sigma^N_0(s)) = \rho \exp(\sigma^D_0(s)) \text{ for some } \rho \in \mathbb{R}_+, \text{ for all } s \in (0, 1),$$

versus the general alternative that the ratio $\exp(\sigma^N_0(s)) / \exp(\sigma^D_0(s))$ is time varying. By Theorem 3 and the delta method, $\exp(\hat{\sigma}^D(s))$ and $\exp(\hat{\sigma}^N(s))$ converge jointly to a normal distribution, and are asymptotically mutually independent. Therefore, consider the t-ratio

$$\hat{t}(s) = \frac{\sqrt{T}h(\hat{\rho}(s) - \rho)}{\sqrt{\hat{\omega}(s)}},$$

$$\hat{\rho}(s) = \frac{\exp(\hat{\sigma}^N(s))}{\exp(\hat{\sigma}^D(s))}, \quad \hat{\rho} = \int_0^1 \frac{\exp(\hat{\sigma}^N(s))}{\exp(\hat{\sigma}^D(s))} ds,$$

$$\hat{\omega}(s) = \rho^2 ||K||_2^2 \left( \frac{\hat{v}_N + 3}{2\hat{v}_N} + \frac{\hat{v}_D + 3}{2\hat{v}_D} \right).$$

Large values of $|\hat{t}(s)|$ are inconsistent with the null hypothesis. It follows from Theorem 3 that for $s \in (0, 1)$, $\hat{t}(s) \Rightarrow N(0, 1)$ under the null hypothesis. We may carry out the pointwise test statistic or confidence interval based on this. We also consider an integrated version of this, specifically let

$$\tau = \frac{\int \hat{t}(s)^2 dW_T(s) - a_T}{b_T},$$

where $W_T(.)$ is some weighting function, for example Lebesgue measure on $[0, 1]$, and $a_T, b_T$ are constants. This test statistic is similar to for example [Fan and Li (1996)]. Under the null hypothesis, $E(\hat{t}(s)^2) \approx 1$ so we take $a_T = 1$. Under the null hypothesis

$$\text{var} \left( \int \hat{t}(s)^2 dW_T(s) \right) = E \left( \int \hat{t}(s)^2 \hat{t}(r)^2 dW_T(s)dW_T(r) \right) - 1.$$  

In the special case that $W_T$ is the measure that puts equal mass on the points $s_1, \ldots, s_M$ with $M = O(Th)$ so that $\hat{t}(s_l)$ and $\hat{t}(s_k)$ are asymptotically independent for $l \neq k$, we may take $b_T = \sqrt{2}$, because $E(\hat{t}(s)^4) \approx 3$. Under the null hypothesis, $\tau \Rightarrow N(0, 1)$, while under the alternative hypothesis $\tau \to \infty$ with probability one. This testing strategy is well suited to detect general alternatives to the null hypothesis of constancy of the volatility ratio.

### 5 A Multivariate model

We next consider an extension to a multivariate model. We keep a similar structure to the univariate model except that we allow the slowly moving component to be matrix valued.
We consider two approaches to modelling the conditional mean. Suppose that
\[ r_t = \begin{pmatrix} r^D_t \\ r^N_t \end{pmatrix}; \quad \mu = \begin{pmatrix} \mu^D \\ \mu^N \end{pmatrix}, \]
where \( r^D_t \) and \( r^N_t \) are \( n \times 1 \) vectors containing all the intraday and overnight returns respectively, and let
\[ \mathbf{D} r_t = \mu + \Pi r_{t-1} + u_t, \]
where \( u^D_t \) and \( u^N_t \) are mean zero shocks, while
\[ \mathbf{D} = \begin{pmatrix} I_n & \text{diag}(\Delta) \\ \text{diag}(\Pi) & I_n \end{pmatrix}; \quad \Pi = \begin{pmatrix} \text{diag}(\Pi_{11}) & \text{diag}(\Pi_{12}) \\ \text{diag}(\Pi_{21}) & \text{diag}(\Pi_{22}) \end{pmatrix}, \]
and \( \Delta, \Pi_{11}, \Pi_{12}, \Pi_{21}, \text{and } \Pi_{22} \) are \( n \times 1 \) vectors. This dynamic model is similar to that considered in the univariate section. In the application we also consider an alternative modelling approach when we have also market returns. In this case, we specify \( r^D_t \) using a microstructure-adjusted Market Model
\[ \begin{align*}
    r^D_t &= a^D_t + \beta^D N_{mt} r^N_{mt} + \beta^D r^N_{mt-1} + u^D_t, \\
    r^N_t &= a^N_t + \beta^N N_{mt} r^D_{mt} + \beta^N r^D_{mt-1} + u^N_t,
\end{align*} \tag{17} \tag{18} \]
where \( r^D_{mt} \) and \( r^N_{mt} \) are the market intraday and overnight returns and \( r^D_{it} \) and \( r^N_{it} \) are the returns of stock \( i \). Including a lagged return in the market model to account for microstructure goes back to Scholes and Williams (1977).

We now consider the specification of the variance equation for the errors \( u_t \). We suppose that
\[ u_t = \begin{pmatrix} \Sigma^D(t/T)^{1/2} \text{diag}(\exp(\lambda^D_t)) \\ 0 \\ \Sigma^N(t/T)^{1/2} \text{diag}(\exp(\lambda^N_t)) \end{pmatrix} \begin{pmatrix} \varepsilon^D_t \\ \varepsilon^N_t \end{pmatrix}, \]
where: \( \varepsilon^D_t \) are i.i.d. shocks (mutually independent across \( i, j, \) and \( t \), identically distributed over \( t \)) from univariate t distributions with \( v_{ij} \) degrees of freedom, while \( \lambda^D_t \) are \( n \times 1 \) vectors. We assume that \( \Sigma^D(\cdot) \) and \( \Sigma^N(\cdot) \) are smooth matrix functions but are otherwise unknown. They allow slowly evolving correlation between stocks in the day or night, and for those correlations to vary by stock and over time.

We can write the covariance matrices in terms of the correlation matrices and the variances as follows
\[ \Sigma^j(s) = \text{diag}(\exp(\sigma^j(s))) R^j(s) \text{diag}(\exp(\sigma^j(s))), \quad j = D, N; \tag{19} \]
with \( \text{diag}(\exp(\sigma^j(s))) \) being the volatility matrix and \( R^j(s) \) being the correlation matrix with unit diagonal elements and off-diagonal elements \( R^j_{il}(s) \) with \( -1 \leq R^j_{il}(s) \leq 1 \). For identification, we still assume \( \int_0^1 \sigma^j_i(s) ds = 0 \), for \( i \in \{1, \ldots, n\} \) and \( j = D, N \).

As with the univariate model, define \( e^j_t = \text{diag}(\exp(\lambda^j_t)) e^j_t \in \mathbb{R}^n \), and suppose that:
\[ m^j_{it} = \frac{(1 + v_{ij})(e^j_{it})^2}{v_{ij} \exp(2\lambda^j_{it}) + (e^j_{it})^2} - 1, \]

13
\[ \lambda^D_{it} = \omega_{it}D(1 - \beta_{it}D) + \beta_{it}D\lambda^D_{it-1} + \gamma_{it}Dm^D_{it-1} + \rho_{it}Dm^N_{it} \]
\[ + \gamma^*_D(m^D_{it-1} + 1)\text{sign}(u^D_{it-1}) + \rho^*_D(m^N_{it} + 1)\text{sign}(u^N_{it}), \]
\[ \lambda^N_{it} = \omega_{it}N(1 - \beta_{it}N) + \beta_{it}N\lambda^N_{it-1} + \gamma_{it}Nm^N_{it-1} + \rho_{it}Nm^D_{it-1} \]
\[ + \rho^*_N(m^D_{it-1} + 1)\text{sign}(u^D_{it-1}) + \gamma^*_N(m^N_{it} + 1)\text{sign}(u^N_{it-1}). \]

For each \( i \) define the parameter vector \( \phi_i = (\omega_{iD}, \beta_{iD}, \gamma_{iD}, \rho_{iD}, \rho^*_D, v_{iD}, \omega_{iN}, \beta_{iN}, \gamma_{iN}, \rho_{iN}, \rho^*_N, v_{iN})^T \in \Phi \subset \mathbb{R}^{14} \), and let \( \phi = (\phi_1^T, \ldots, \phi_n^T) \) denote all the dynamic parameters.

Define \( \epsilon_i \) the vector with the \( i^{th} \) element 1 and all others 0, so that \( \epsilon_i^T = \epsilon_1^T \text{diag}(\exp(-\lambda_i)) \Sigma^j(t/T)^{-1/2} u_i^j \).

The normalized global log-likelihood function is
\[ l_T(\phi, \Sigma(\cdot)) = \frac{1}{T} \sum_{t=1}^{T} l_T^N(\phi, \Sigma(\cdot)) + l_T^D(\phi, \Sigma(\cdot)) \]
\[ l_T^i(\phi, \Sigma(\cdot)) = \sum_{i=1}^{n} \left( - \prod_{i=1}^{n} \lambda^j_{it} - v_{ij} + \frac{1}{2} \ln \left( 1 + \left( \epsilon_i^T \text{diag}(\exp(-\lambda_i^j)) \Sigma^j(t/T)^{-1/2} u_i^j \right)^2 \right) \right) \]
\[ - \frac{1}{2} \ln \det \Sigma^j(t/T) + \sum_{i=1}^{n} \left( \ln \Gamma \left( \frac{v_{ij} + 1}{2} \right) - \frac{1}{2} \ln v_{ij} - \ln \Gamma \left( \frac{v_{ij}}{2} \right) \right). \]

Our estimation algorithm is as follows. We first define an initial estimator for \( \Sigma^j(t/T) \) and then obtain an estimator of \( \phi \), and then we update them. Suppose that we know \( \Delta, \Pi \) and \( \mu \). To give an estimator of \( \Sigma^j(t/T) \) that is robust to heavy tails, we estimate the volatility parameter
\[ \widetilde{\sigma}^j_i(s) = \frac{1}{\alpha} \log \left( \frac{1}{T} \sum_{t=1}^{T} K_h(s - t/T) |u^j_{it}|^\alpha \right), \quad (20) \]
and then rescale \( \widetilde{\sigma}^j_i(t/T) \) as
\[ \widetilde{\sigma}^j_i(t/T) = \widetilde{\sigma}^j_i(t/T) - \frac{1}{T} \sum_{t=1}^{T} \widetilde{\sigma}^j(t/T). \quad (21) \]

Supposing that the heavy tails issue is less severe in the estimation of correlation, which seems reasonable, we estimate the correlation parameter by standard procedures
\[ \widetilde{R}_{ik}(s) = \frac{\sum_{t=1}^{T} K_h(s - t/T) u^j_{it} u^j_{jt}}{\sqrt{\sum_{t=1}^{T} K_h(s - t/T) u^j_{it} u^j_{it}} \sum_{t=1}^{T} K_h(s - t/T) u^j_{jt} u^j_{jt}} \quad (22) \]
for \( s \in (0, 1) \), and boundary modification as previously detailed. Alternatively, we can use a robust correlation estimator. Omitting the superscript \( j = D, N \) here, we may compute the pairwise Kendall tau
\[ \hat{\tau}_{ik}(s) = \frac{\sum_{t=1}^{T-1} \sum_{j=i}^{T-1} K_h(s - \frac{t}{T}) K_h(s - \frac{j}{T}) (I \{(u_{i,k} - u_{j,k})(u_{i,l} - u_{j,l}) > 0\} - I \{(u_{i,k} - u_{j,k})(u_{i,l} - u_{j,l}) < 0\})}{\sum_{t=1}^{T-1} \sum_{j=i}^{T-1} K_h(s - \frac{t}{T}) K_h(s - \frac{j}{T}) (I \{(u_{i,k} - u_{j,k})(u_{i,l} - u_{j,l}) > 0\} + I \{(u_{i,k} - u_{j,k})(u_{i,l} - u_{j,l}) < 0\})}. \]
Then applying the relation between Kendall tau and the linear correlation coefficient for the elliptical distribution suggested by Lindskog, Mcneil, and Schmock (2003) and Battey and Linton (2014), we obtain the robust linear correlation estimator, \( \hat{\rho}_{k,l}(s) = \sin(\frac{\pi}{2} \hat{\tau}_{k,l}(s)) \). In some cases, the matrix of pairwise correlations must be adjusted to ensure that the resulting matrix is positive definite.

We have

\[
\tilde{\Sigma}^j(s) = \text{diag}(\exp(\tilde{\sigma}^j(s))) \tilde{R}^j(s) \text{diag}(\exp(\tilde{\sigma}^j(s))) \text{, } j = D, N. \tag{23}
\]

Letting \( \tilde{c}^j_i = \tilde{\Sigma}^j(\frac{t}{T})^{-1/2}u^j_i \), we obtain \( \tilde{\phi}_i \) by maximizing the univariate log-likelihood function of \( \tilde{c}^j_i \) in \( \Theta \) for each \( i = 1, \ldots, n \). To update the estimator for each \( \Sigma^j(\frac{t}{T}) \), denote \( \Theta = (\Sigma^j)^{-1/2} \). We first obtain \( \hat{\Theta} \) with the local likelihood function given \( \tilde{\lambda}^j_i \) and \( \tilde{v}^j_i \), i.e., maximize the local objective function

\[
L^j_I(\Theta; \tilde{\lambda}, s) = \frac{1}{T} \sum_{t=1}^{T} K_h(s - t/T) \left[ \log |\Theta| - \sum_{i=1}^{n} \left( \tilde{v}^j_{ij} + 1 \right) \ln \left( 1 + \frac{(\tilde{u}^j_i \text{diag} \left( \exp(-\tilde{\lambda}^j_i) \right) \Theta u^j_i)^2}{\tilde{v}^j_{ij}} \right) \right]
\]

with respect to vech(\( \Theta \)), and let \( \hat{\Sigma}^j(t/T) = \hat{\Theta}^{-2} \). The derivatives of the objective function are given in (13) and (14) in the Supplementary Material contained in Linton and Wu (2018).

Our multivariate model can be considered as a diagonal DCS EGARCH model with a slowly moving correlation matrix. Assuming diagonality on the short run component \( \lambda^j_t \) enables us to estimate the model easily and rapidly. In particular, the computation time of the initial estimator is only of order \( n \), with \( n \) being the number of assets considered; it is thus feasible even with quite large \( n \). The extension to models with non-diagonal short run components is possible, but only feasible with small \( n \). We do not provide the distribution theory here for space reasons but it follows by similar arguments to given for the univariate case. The invertibility conditions of \( \lambda^j_t \) are the same as those in the univariate model.

Blanc, Chicheportiche, and Bouchaud (2014) impose a pooling assumption in their modelling, which translates here to the restriction that \( \phi_i = \phi_1 \) for all \( i = 1, \ldots, n \). This improves efficiency when the restriction is true. We can test the restriction by a standard Wald procedure or Likelihood ratio statistic. In the application we find these pooling restrictions are strongly rejected by the data.

6 Empirical application

In this section, we first apply our coupled-component GARCH model to the Dow Jones stocks, and we report detailed estimates and examine the out-of-sample forecast performance. We also apply our model to the size-based portfolios with stocks in the CRSP database, and briefly describe the results in subsection 6.7.

6.1 Data and preliminary analysis

We investigate 26 components of the Dow Jones industrial average index during the period of 4 January 1993 to 29 December 2017. The 26 stocks are AAPL, MSFT, XOM, JNJ, INTC, WMT, CVX, UNH, CSCO, HD, PFE, BA, VZ, PG, KO, MRK, DIS, IBM, GE, MCD, MMM, NKE, UTX,
CAT, AXP, and TRV. The data are obtained from Datastream, and the prices have been adjusted for corporate actions. We define overnight returns as the log price change between the close of one trading day and the opening of the next trading day. We do not incorporate weekend and holiday effects into our model as they are not the focus of this paper. In addition, although the weekend effect is documented by studies such as French (1980) and Rogalski (1984), and further supported by Cho, Linton, and Whang (2007) with a stochastic dominance approach, many studies suggest the disappearance of the weekend effect, including Mehdian and Perry (2001) and Steeley (2001). In addition, Sullivan, Timmermann, and White (2001) claim that many calendar effects arise from data-snooping.

Many studies, such as Cooper, Cliff, and Gulen (2008) and Berkman, Koch, Tuttle, and Zhang (2012), find significantly higher overnight returns. Cooper, Cliff, and Gulen (2008) even suggest that the U.S. equity premium during their research period is solely due to overnight returns. Figure A.1 in Linton and Wu (2018) plots the cumulative returns for these 26 stocks. There is no clear dominance of positive overnight returns from these Dow Jones stocks.

Berkman, Koch, Tuttle, and Zhang (2012) find significant positive mean overnight returns of +10 basis points (bp) per day, along with -7 bp for intraday returns from the 3000 largest U.S. stocks. Following Berkman, Koch, Tuttle, and Zhang (2012), we first compute the cross-sectional mean returns for each day, then compute the time-series mean and standard deviation of these values. The mean intraday return is 2.05 bp with a standard error of 1.12 bp, while the mean overnight return is 1.68 bp with a standard error of 0.71 bp. The difference between overnight and intraday means is not statistically significant.

Table A.1 in Linton and Wu (2018) provides summary statistics for intraday and overnight returns. Compared with intraday returns, overnight returns exhibit more negative skewness and leptokurtosis. Specifically, 9 of these 26 stocks exhibit negative intraday skewness, while 25 of these 26 stocks have negative overnight skewness. The largest sample kurtosis for overnight returns is 935.78, which suggests the non-existence of the population kurtosis. We find that the per hour variance of intraday returns is roughly 12 times the per hour variance of overnight returns, which is somewhat less than the range of 13-100 times found by French and Roll (1986).

### 6.2 Results of the univariate model

We estimate the univariate coupled-component GARCH model for each Dow Jones stocks. The estimates and their robust standard errors in the mean equations are reported in Table A.2 in Linton

---

4These stocks are constituents of the Dow Jones index according to the constituent list in May 2018. The V, GS, and DWDP.K are excluded because they do not have prices available in 1993. The JPM is excluded because its open price from 2 September 1993 to 4 January 1995 is missing in Datastream.

Suppose that hourly stock returns satisfy \( r_{ht} \sim \mu_h, \sigma_{h}^2, \kappa_{h3}, \kappa_{h4} \), which is consistent with French and Roll (1986). Daily (based on a 6-hour trading day) and weekend (66 hours from Friday close to Monday open) returns should then satisfy

\[
r_{Dh} \sim 6 \mu_h, 6 \sigma_h^2, \frac{\kappa_{h3}}{\sqrt{6}}, \frac{\kappa_{h4}}{6} ; \quad r_{Wt} \sim 66 \mu_h, 66 \sigma_h^2, \frac{\kappa_{h3}}{\sqrt{66}}, \frac{\kappa_{h4}}{66}
\]

In fact, overnight returns including weekend returns are very leptokurtic.
and Wu (2018). We multiply returns by 100 to give more readable coefficients. \( \Pi_{ij} \) refers to the element of the \( i \)th row \( j \)th column in the coefficient matrix \( \Pi \). For the prediction of intraday returns, 12 of the 28 stocks have significant \( \Pi_{11} \) values that are all negative, and 7 of 28 stocks have significant \( \delta \) values which are all positive. This outcome suggests that both overnight and intraday returns tend to have a negative effect on the subsequent intraday return. However, we do not find clear patterns for predicting overnight returns. The constant terms, \( \mu_D \) and \( \mu_N \), are positive for most Dow Jones stocks.

Table A.3 in Linton and Wu (2018) gives the estimates of dynamic parameters. Parameters \( \beta_D \) and \( \beta_N \) are significantly different from 1, and \( \rho_D, \gamma_D, \rho_N \) and \( \gamma_N \) are positive and significant. In addition, we find significant leverage effects, with negative and significant \( \rho_D^*, \gamma_D^*, \rho_N^*, \) and \( \gamma_N^* \), which suggest higher volatility after negative returns.

We are also concerned about the difference between overnight and intraday parameters. Table A.4 in Linton and Wu (2018) reports Wald tests with the null hypothesis that the intraday and overnight parameters are equal within each stock. The parameter \( \omega_D \), which determines the unconditional short-run scale, is significantly larger than \( \omega_N \). The overnight degree-of-freedom parameter is around 3, which is significantly smaller than the intraday counterpart at approximately 8. Both are in line with previous studies suggesting that overnight returns are more leptokurtic but less volatile. With other pairs of intraday and overnight parameters, \( \beta_j, \gamma, \rho, \gamma_j^* \rho_j^* \), the null hypothesis is seldom rejected. However, the joint null hypothesis, \((\beta_D, \gamma, \rho_D, \gamma_D^*, \rho_D^*) = (\beta_N, \gamma, \rho_N, \gamma_N^*, \rho_N^*)\), is rejected by many stocks. It is noteworthy that the null hypothesis \( H_0 : \gamma_N = \rho_D \) is not rejected by our data, which is inconsistent with Blanc, Chicheportiche, and Bouchaud (2014). They suggest that past overnight returns weakly affect future intraday volatilities, except for the very next one, but have a substantial impact on future overnight volatilities. This inconsistency is probably because the dynamic conditional score model shrinks the impact of extreme overnight observations. After this shrinkage, the effect of overnight innovations on parameter estimation becomes closer to the intraday innovations.

Many papers have argued that the introduction of high-frequency trading in the period post 2005 has led to an increase in volatility. Figure A.2 in Linton and Wu (2018) plots the intraday and overnight volatilities, \( \sqrt{\frac{\nu_j}{\nu_j - 2}} \exp(2\lambda_j^t + 2\sigma_j^t(\frac{t}{T})) \), for \( j = D, N \). The five dashed vertical lines from left to right indicate the dates: 10 March 2000 (dot-com bubble), 11 September 2001 (the September 11 attacks), 16 September 2008 (financial crisis), 6 May 2010 (flash crash), and 1 August 2011 (August 2011 stock markets fall). The intraday volatility (red lines) significantly dominates the overnight volatility (black lines) in the first half of the study period, but this domination gradually disappears, especially after the 2008 financial crisis. In addition, the intraday volatilities after 2005 are in general smaller than those before 2005, except for the financial crisis period. This finding is contrary to the typical argument that high-frequency trading increases volatilities. To further investigate this point, we plot the ratios of overnight to intraday volatility in Figure 1. All stocks exhibit upward trends over the 25-year period considered here, and many of them experience peaks around August 2011, corresponding to the August 2011 stock markets fall.
This figure shows the dynamic ratio of overnight to intraday volatility, based on the univariate coupled-component model, with one subplot for each stock. The five dashed vertical lines from left to right represent the dates: 10 March 2000 (dot-com bubble), 11 September 2001 (the September 11 attacks), 16 September 2008 (financial crisis), 6 May 2010 (flash crash), and 1 August 2011 (August 2011 stock markets fall), respectively. Intraday and overnight volatiles are defined as $\sqrt{\nu_j/\nu_j - 2} \exp(2\lambda_j t + 2\sigma_j^2 t)$, for $j = D, N$.

Figure 1: Ratios of overnight to intraday volatility: univariate model
6.3 Constancy of the ratio of overnight to intraday volatility

The long-run intraday and overnight components, $\sigma^D(t/T)$ and $\sigma^N(t/T)$, and their 95% point-wise confidence intervals are depicted in Figure A.3 in Linton and Wu (2018). Most stocks arrive at their first peaks around 10 March 2000, corresponding to the dot-com bubble event, while some arrive at around September 2011, right after 9-11. The intraday components reach their second peaks during the financial crisis in September 2008, while overnight components continue to rise until around 2011. Roughly speaking, the intraday components are larger than the overnight ones before the first peaks, but smaller after the financial crisis of September 2008. However, it is imperative to remember that the long-run components are constructed with rescaling $\int_0^1 \sigma(s) ds = 0$. In general, the intraday volatility is still larger.

We test the constancy of the ratio of long run overnight to intraday volatility. Figure A.4 in Linton and Wu (2018) displays the test statistics $\hat{t}(s)$ and the 95% point-wise confidence intervals for $s \in [0,1]$. Consistent with the results above, the equal ratio null hypothesis is mostly rejected before the first peaks (in 2000) and after the second peaks (in 2010).

Cumulatively, this evidence indicates that the overnight volatility has increased in importance during the 25-year period considered here, relative to the intraday volatility for the Dow Jones stocks.

6.4 Volatility forecast comparison

We also compare our coupled-component GARCH model with its one-component version for the open-to-close returns to assess the improvement in volatility forecast from using overnight returns. We construct 10 rolling windows, each containing 5652 in-sample and 50 out-of-sample observations. In each rolling window, the parameters in the short-run variances are estimated with the in-sample data once and stay the same during the one-step out-of-sample forecast. In the one-step-ahead forecast of the long-run covariance matrices, the single-side weight function is used. For instance, to forecast the long-run covariance matrix of period $\tau$ ($s = \tau/T$), we set the two-sided weight function $K_h(s - t/T) = 0$, for $t \geq \tau$, and then rescale $K_h(s - t/T)$ to obtain a sum of 1. Table A.5 in Linton and Wu (2018) reports Giacomini and White (2006) model pair-wise comparison tests with the out-of-sample quasi-Gaussian and student t log-likelihood loss functions. For most stocks, the coupled-component GARCH model dominates the one-component model. Some dominances are statistically significant. We omit the comparison for overnight variance forecast between the one-component and the coupled-component model since it is not plausible to estimate a GARCH model with overnight returns alone.

6.5 Diagnostic tests

Ljung-Box tests on the absolute and squared standardized residuals are used to verify whether the coupled-component GARCH model is adequate to capture the heteroskedasticity, shown in Table A.6 in Linton and Wu (2018). With the absolute form, strong heteroskedasticity exists in both intraday and overnight returns but disappears in the standardized residuals, implying that our model captures the heteroskedasticity well. However, we are sometimes unable to detect the heteroskedasticity in
overnight returns with squared values. In general, the use of the absolute form is more robust when
the distribution is heavy tailed.

Figure A.5 in Linton and Wu (2018) displays the quantile-quantile (Q-Q) plots of the intraday
innovations, comparing these with the student t distribution with $\hat{\nu}_D$ degrees of freedom. The
points in the Q-Q plots approximately lie on a line, showing that the intraday innovations closely
approximate the t distribution. Figure A.6 in Linton and Wu (2018) displays the Q-Q plots of the
overnight innovations. Many stocks have several outliers in the lower left corners. Our model only
partly captures the negative skewness and leptokurtosis of overnight innovations.

6.6 Results of the multivariate model

The long-run correlations between intraday or overnight returns are presented in Figure A.7 in Linton
and Wu (2018). Each subplot presents the averaged correlations between that individual stock and
the remaining stocks. The correlations exhibit an obvious upward trend during the sample period
of 1998-2016. In the 1990s, the overnight correlations and intraday correlations are both around
0.2, albeit with fluctuations. In the period 2000 to 2007, intraday correlations start to increase
and are larger than the overnight correlations. However, during the period 2008 to 2016, overnight
correlations increase substantially to around 0.7 in 2011 and remain higher than 0.5, while intraday
correlations peak in around 2008 but the correlations are seldom larger than 0.5. Both correlations
start to decrease in 2017.

The eigenvalues of the dynamic covariance matrices and their scaled values (the eigenvalues
divided by the sum of eigenvalues) are presented in Figure A.8 in Linton and Wu (2018). The dynamic
of eigenvalues reinforces the previous remark that the stock markets experienced high intraday risk
in the 9-11 attacks in 2001 and in the 2008 financial crisis, while stock markets experienced high
overnight risk in around 2011. The largest eigenvalue represents a strong common component,
illustrating that a large proportion of the market financial risk can be explained by a single factor.
The largest eigenvalue increases substantially during our research period. The second and third
largest eigenvalues still account for a considerable proportion of risk in the volatile period from 2000
to 2002, but become rather insignificant in the volatile period from 2008 to 2011. The largest intraday
eigenvalue proportion reaches its peak in 2008, while the largest overnight eigenvalue proportion
remains consistently high until 2011. Remarkably, the largest eigenvalue explains nearly 50% of
intraday risk in the 2008 financial crisis and 70% of overnight risk in the August 2011 stock markets
fall. The overnight eigenvalue proportion is much higher than its intraday counterpart in the period
2008 to 2016. Generally speaking, the market risk in the crisis period from 2008 to 2011 can be
largely explained by a single-factor structure, in particular, the overnight risk. This is in line with
the finding of Li, Viktor, and George (2017) that stocks returns tend to obey an exact one-factor
structure at times of market-wide jump events.

One concern is that our initial correlation estimator is based on the Pearson product moment
correlation. This Pearson estimator may perform poorly because of the heavy tails of overnight
innovations. Therefore, we also try the robust correlation estimator in the initial step, yet the
results remain unchanged (Figure A.9 in Linton and Wu (2018)). This figure plots the largest scaled
eigenvalue of the estimated covariance matrix to assess the difference between using robust (in black) and non-robust (in red) correlation estimators in the initial step. We use dashed lines for the initial estimators and solid lines for the updated estimators. Despite the large difference of initial estimators, particularly for overnight returns, the updated estimators are roughly similar. Like the eigenvalues, the updated covariances themselves are also robust to a different initial estimator.

6.7 Results with CRSP stocks

We also investigate the overnight and intraday volatilities of 10 size-based portfolios with stocks in the CRSP database from January 1993 to December 2017. The portfolios are constructed with the CRSP assignments. We estimate the univariate coupled-component GARCH model with the intraday and overnight returns in each size-based portfolio. Parameters $\beta_D$ and $\beta_N$ are significantly different from 1, and $\rho_D$, $\gamma_D$, $\rho_N$ and $\gamma_N$ are significantly positive (Table A.10 in Linton and Wu (2018)). The leverage effects are also significant, suggesting higher volatility after negative returns. The overnight degrees of freedom are larger than 4, less heavy tailed than that of individual Dow Jones stocks.

Figure 2 presents the ratio of overnight to intraday volatility of those portfolios. The ratio exhibits a downward trend in small-cap portfolios (in particular before 2000) and an upward trend in large-cap portfolios. Notably, the trend changes monotonically from the smallest-cap portfolio to the largest-cap portfolio. The main explanation for this phenomenon is perhaps the variation in international linkage. Stocks with higher international correlations show considerably higher overnight to intraday volatility ratio, and likewise with larger market capitalization. The changes of minimal tick size also make a contribution to the downward trend of small stocks during 1990s. See the detailed discussion in Linton and Wu (2018).

7 Conclusion

The empirical results show that the ratio of overnight to intraday volatility for especially large stocks has increased during the last 25 years when accounting for both slowly changing and rapidly changing components. This is contrary to what is often argued with regard to the change in market structure and the effects of high frequency trading. Portfolios of small stocks on the other hand seem to exhibit a different trend.

We found various other results. First, we found in the multivariate model that (slowly moving) correlations between assets have increased during our sample period. In addition, overnight correlations increase more substantially than intraday correlations during recent crises. We also found that the information in overnight returns is valuable for updating the forecast of the close to close volatility.

\footnote{We also construct beta-sorted and deviation-sorted portfolios with the assignments provided in the CRSP database. However, this pattern is not found in the beta-sorted or standard deviation-sorted portfolios. Nearly all beta-sorted and standard deviation-sorted portfolios exhibit increasing overnight to intraday volatility ratios.}
This figure plots the ratio of overnight to intraday volatility for portfolios formed on size. Decile 1 is the portfolio with the smallest market capitalizations and decile 10 is the portfolio with the largest market capitalizations. The intraday and overnight volatilities are \( \sqrt{\nu_j} \nu_j - 2 \exp(2\lambda_j t + 2\sigma_j (\frac{1}{t})) \) for \( j = D, N \), respectively. The five dashed vertical lines from left to right indicate the dates 10 March 2000 (dot-com bubble), 11 September 2001 (the September 11 attacks), 16 September 2008 (financial crisis), 6 May 2010 (flash crash), and 1 August 2011 (August 2011 stock markets fall), respectively.

Figure 2: Ratio of overnight to intraday volatility of size-based portfolios
In our modelling we have not separated midweek overnight components from weekend components. We may extend the model to allow multiple different components reflecting weekend different from intraweek overnight, but at the cost of estimating many more parameters. We are also considering how to extend the model to allow stocks traded in different time zones, Lin, Engle, and Ito (1994).

8 Acknowledgements

We would like to thank Greg Connor, Jinyong Hahn, Andrew Harvey, Hashem Pesaran, Piet Sercu, Haihan Tang and Chen Wang for useful comments and suggestions. We are also grateful to two referees and the editor for helpful comments. Linton acknowledges the financial support from the Cambridge INET. Wu acknowledges the financial support from the National Natural Science Foundation of China [grant number 71803080].

References


