ARTICLE IN PRESS
Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A coupled component DCS-EGARCH model for intraday and overnight volatility

Oliver Linton [a], Jianbin Wu [b,*]

[a] Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom of Great Britain and Northern Ireland
[b] School of Economics, Nanjing University, Nanjing, 210093, China

ARTICLE INFO

Article history:
Received 24 September 2018
Received in revised form 20 September 2019
Accepted 29 December 2019
Available online xxxx

JEL classification:
C12
C13

Keywords:
DCS
GAS
GARCH
Size-based portfolios
Testing

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.

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 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; 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

* Corresponding author.
E-mail addresses: obl20@cam.ac.uk (O. Linton), wujianbin@nju.edu.cn (J. Wu).

https://doi.org/10.1016/j.jeconom.2019.12.015

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 et al. (2015) and Linton et al. (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; 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 et al. (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.[1] 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 et al. (2008) suggest that the U.S. equity premium over the period 1993–2006 is solely due to overnight returns. Kelly and Clark (2011) find the overnight returns are on average larger than the intraday returns. Berkman et al.

[1] Our 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.

Please cite this article as: O. Linton and J. Wu, A coupled component DCS-EGARCH model for intraday and overnight volatility. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2019.12.015.

(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 et al. (2018) suggest overnight returns can serve as a measure of firm-specific investor sentiment, and find short-term persistence in overnight returns. Polk et al. (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 et al., 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 modelling and forecasting, by scaling the intraday measure (e.g., Martens, 2002 and Fleming et al., 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 et al. (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 et al. (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 et al. (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.

2. The model and its properties

We let r_t^D denote intraday returns and r_t^N denote overnight returns on day t. We take the ordering that night precedes day so that $r_t^D = \ln(P_t^C/P_t^O)$ and $r_t^N = \ln(P_t^O/P_{t-1}^C)$, where P_t^O denotes the open price on day t and P_t^C denotes the close price on day t. Daily close to close returns satisfy $r_t = r_t^D + r_t^N$. The timeline is illustrated below

$$\cdots \longrightarrow P_{t-1}^C \;\Bigg|\; \underbrace{\underbrace{\xrightarrow{\text{Night } t}}_{r_t^N} \; P_t^O \; \underbrace{\xrightarrow{\text{Day } t}}_{r_t^D}}_{r_t} \; P_t^C \;\Bigg|\; \underbrace{\xrightarrow{\text{Night } t+1}}_{r_{t+1}^N} \; P_{t+1}^O \longrightarrow \cdots$$

We do not distinguish between weekend, holiday weekends, and ordinary midweek over night 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

$$\begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} r_t^D \\ r_t^N \end{pmatrix} = \begin{pmatrix} \mu_D \\ \mu_N \end{pmatrix} + \Pi \begin{pmatrix} r_{t-1}^D \\ r_{t-1}^N \end{pmatrix} + \begin{pmatrix} u_t^D \\ u_t^N \end{pmatrix}, \tag{1}$$

where u_t^D and u_t^N 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 mispricing 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

$$u_t = \begin{pmatrix} u_t^D \\ u_t^N \end{pmatrix} = \begin{pmatrix} \exp(\lambda_t^D) \exp(\sigma^D(t/T)) \varepsilon_t^D \\ \exp(\lambda_t^N) \exp(\sigma^N(t/T)) \varepsilon_t^N \end{pmatrix}, \tag{2}$$

where: ε_t^D and ε_t^N are i.i.d. mean zero shocks from t distributions with υ_D and υ_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 $j = D, N$:

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

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

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

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 θ, 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 et al. (2011) and (Harvey and Luati, 2014). The conditional (scale) score function associated with the t-distributed shocks is $(1 - x^2(\nu + 1)/(\nu\sigma^2 + x^2))/\sigma^2$, and so we take as innovation processes

$$m_t^j = \frac{(1+\upsilon_j)(e_t^j)^2}{\upsilon_j \exp(2\lambda_t^j) + (e_t^j)^2} - 1, \quad \upsilon_j > 0$$

where $e_t^j = \exp(-\sigma^j(t/T))u_t^j$. Note that m_t^j is a bounded function of e_t^j with $E(m_t^D|\mathcal{F}_t^N) = 0$ and $E(m_t^N|\mathcal{F}_{t-1}^D) = 0$, where $\mathcal{F}_{t-1}^N$ is the information set at the open of day $t-1$ and $\mathcal{F}_{t-1}^D$ is the information set at the close of day $t-1$. We suppose that λ_t^D and λ_t^N are linear combinations of past values of the shocks determined by m_t^j, $j = D, N$:

$$\begin{aligned} \lambda_t^D &= \omega_D(1-\beta_D) + \beta_D\lambda_{t-1}^D + \gamma_D m_{t-1}^D + \rho_D m_t^N \\ &\quad + \gamma_D^*(m_{t-1}^D + 1)\text{sign}(e_{t-1}^D) + \rho_D^*(m_t^N + 1)\text{sign}(e_t^N) \end{aligned} \tag{4}$$

$$\begin{aligned} \lambda_t^N &= \omega_N(1-\beta_N) + \beta_N\lambda_{t-1}^N + \gamma_N m_{t-1}^N + \rho_N m_{t-1}^D \\ &\quad + \rho_N^*(m_{t-1}^D + 1)\text{sign}(e_{t-1}^D) + \gamma_N^*(m_{t-1}^N + 1)\text{sign}(e_{t-1}^N). \end{aligned} \tag{5}$$

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 ρ_D, ρ_D^* capture the effect of overnight shocks on intraday volatility, while ρ_N, ρ_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 λ_t^D, λ_t^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; Glosten et al., 1993), through the parameters $\gamma_D^*, \rho_D^*, \rho_N^*$, and γ_N^*.[2] The parameters β_D, β_N measure the persistence of the volatility processes. We set the intercepts this way so that ω_D is the unconditional mean of λ_t^D and ω_N is the unconditional mean of λ_t^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

$$\phi = (\omega_D, \beta_D, \gamma_D, \gamma_D^*, \rho_D, \rho_D^*, \upsilon_D, \omega_N, \beta_N, \gamma_N, \gamma_N^*, \rho_N, \rho_N^*, \upsilon_N)^{\intercal} \in \Phi \subset \mathbb{R}^{14}$$

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 $\{u_t\}$.

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 down weighted, and in such a way driven by the shape of the error distribution.[3] 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 Lemma 2 we prove that if $|\beta_j| < 1, j = D, N$, then e_t^j and λ_t^j are strongly stationary and β-mixing with exponential decay. He et al. (2002) derive formulae for moments of a family of exponential

[2] The shock variable m_t^j can be expressed as $m_t^j = (\upsilon_j + 1)b_t^j - 1$, where b_t^j has a beta distribution, $beta\ (1/2, \upsilon_j/2)$.

[3] 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.

GARCH models, and similar calculations can be reproduced here. We discuss next the invertibility of the process $(\lambda_t^D, \lambda_t^N)^\intercal$. This property is important for the asymptotic theory. To establish invertibility, we write the dynamics of $(\lambda_t^D, \lambda_t^N)^\intercal$ as a stochastic recurrence equation(SRE) in terms of e_t,

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

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 $(\lambda_0^D, \lambda_0^N)^\intercal$ 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 λ_t^D, λ_t^N and their scores are functions of bounded variables m_{t-k}^j. For the contraction condition, it is easy to have

$$\begin{aligned} \frac{\partial \lambda_t^D}{\partial \lambda_{t-1}^D} &= \beta_D + \left(\gamma_D + \gamma_D^* \text{sign}(u_{t-1}^D)\right) \frac{\partial m_{t-1}^D}{\partial \lambda_{t-1}^D} + \left(\rho_D + \rho_D^* \text{sign}(u_t^N)\right) \frac{\partial m_t^N}{\partial \lambda_t^N} \frac{\partial \lambda_t^N}{\partial \lambda_{t-1}^D} \\ &= \beta_D + a_{t-1}^{DD} + a_t^{DN} a_{t-1}^{ND}, \end{aligned}$$

$$\frac{\partial \lambda_t^N}{\partial \lambda_{t-1}^D} = \beta_D + a_{t-1}^{DD} + a_t^{DN} a_{t-1}^{ND}, \quad \frac{\partial \lambda_t^D}{\partial \lambda_{t-1}^N} = a_t^{DN}(\beta_N + a_{t-1}^{NN}), \quad \frac{\partial \lambda_t^N}{\partial \lambda_{t-1}^D} = a_{t-1}^{ND}.$$

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 (9).

Applying Theorem 3.1 in Bougerol (1993), if for some integer $p \geqq 1$, $E\log(\sup_{\lambda_0^D, \lambda_0^N} \| \prod_{i=1}^p A_{p-i+1} B_{p-i} \|_\infty) < 0$ holds, then $(\lambda_t^D, \lambda_t^N)^\intercal$ is invertible (the norm $\|.\|_\infty$ is defined below in (10)). 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 condition is often satisfied, e.g., with $p = 2$. With invertibility, the assumption that h_t^j 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 et al., 2008). Thiele (2019) considers some DCS models with asymmetric t-distributions, which might be a possible direction for future work.

3. Estimation

Suppose that we know δ, μ, Π and hence $u_t^j, j = D, N$. 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 ϕ and $\sigma^j(.)$. For any $\alpha > 0$, we have for $j = N, D$,

$$E\left(\left|u_t^j\right|^\alpha\right) = E\left(\left|\varepsilon_t^j\right|^\alpha\right) E\left(\exp\left(\alpha\lambda_t^j\right)\right) \exp(\alpha\sigma^j(t/T)) = c^j(\phi; \alpha) \times \exp(\alpha\sigma^j(t/T)),$$

where c^j is a constant that depends in a complicated way on the parameter vector ϕ and on α. 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

$$\widetilde{\sigma}^j(s) = \frac{1}{\alpha} \log\left(\frac{1}{T} \sum_{t=1}^T K_h(s - t/T) \left|u_t^j\right|^\alpha\right) \tag{6}$$

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 $\int_{-1}^c K(u, c)du = 1$, $\int_{-1}^c uK(u, c)du = 0$, 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, 1994). One may apply more sophisticated adjustments such as Jones et al. (1995) that preserves positivity but reduces the bias in the boundary region. For identification, we recenter

$\widetilde{\sigma}^j(t/T)$ as

$$\widetilde{\sigma}^j(t/T) = \widetilde{\sigma}^j(t/T) - \frac{1}{T}\sum_{t=1}^{T}\widetilde{\sigma}^j(t/T). \tag{7}$$

Note that $\widetilde{\sigma}^j(u)$ can be written as $\widetilde{\sigma}^j(u) = \sum_{i=1}^{\infty}\widetilde{\theta}_i^j\psi_i^j(s)$ for some coefficients $\widetilde{\theta}_i^j$ (that satisfy $\sum_{i=1}^{\infty}|\widetilde{\theta}_i^j| < \infty$) determined uniquely by the estimator $\widetilde{\sigma}^j(u)$, that is, we can represent the kernel estimator as a sieve estimator with a potentially infinite number of coefficients, see Appendix B.2. We will use this representation for notational convenience, that is, we will represent $\widetilde{\sigma}^j(.)$ in terms of $\{\widetilde{\theta}_i^j\}_{i=1}^{\infty}$ or just $\widetilde{\theta}$ for shorthand. In practice, the bandwidth may be chosen by some rule of thumb method.

Let $\widetilde{e}_t^N = \exp(-\widetilde{\sigma}^N(t/T))u_t^N$ and $\widetilde{e}_t^D = \exp(-\widetilde{\sigma}^D(t/T))u_t^D$, and let $\widetilde{\theta}$ denote $\{\widetilde{\sigma}^j(s), s \in [0,1], j = N, D\}$. Define the global log-likelihood function for ϕ (apart from an unnecessary constant and conditional on the estimated values of θ)

$$l_T(\phi;\widetilde{\theta}) = \frac{1}{T}\sum_{t=1}^{T} l_t(\phi;\widetilde{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\left(l_t^N(\phi;\widetilde{\theta}) + l_t^D(\phi;\widetilde{\theta})\right), \tag{8}$$

$$l_t^j(\phi;\widetilde{\theta}) = -\lambda_t^j(\phi;\widetilde{\theta}) - \frac{v_j+1}{2}\ln\left(1 + \frac{(\widetilde{e}_t^j)^2}{v_j\exp(2\lambda_t^j(\phi;\widetilde{\theta}))}\right) + \ln\Gamma\left(\frac{v_j+1}{2}\right) - \frac{1}{2}\ln v_j - \ln\Gamma\left(\frac{v_j}{2}\right),$$

where Γ is the gamma function and $\lambda_t^j(\phi;\widetilde{\theta})$ are defined in (4) and (5). For practical purposes, $\lambda_{1|0}^j$ may be set equal to the unconditional mean, $\lambda_{1|0}^j = \omega_j$. We estimate ϕ by maximizing $l_T(\phi;\widetilde{\theta})$ with respect to $\phi \in \Phi$. Let $\widetilde{\phi}$ denote these estimates.

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

$$\widetilde{\eta}_t^N = \exp(-\widetilde{\lambda}_t^N)u_t^N \;;\; \widetilde{\eta}_t^D = \exp(-\widetilde{\lambda}_t^D)u_t^D,$$

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

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

with respect to $\gamma \in \mathbb{R}$, for $j = D, N$ separately, where $\widetilde{\lambda}^j = (\widetilde{\lambda}_1^j, \ldots, \widetilde{\lambda}_T^j)^{\intercal}$. 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 (25) in Appendix B. Harvey (2013) gives some discussion about computational issues. To summarize, the estimation algorithm is as follows.

Algorithm.

STEP 1. Estimate δ, μ^j, Π by least squares and $\widetilde{\sigma}^j(u), u \in [0,1], j = N, D$ from (6) and (7)

STEP 2. Estimate ϕ by optimizing $l_T(\phi;\widetilde{\theta})$with respect to $\phi \in \Phi$ (by Newton–Raphson) to give $\widetilde{\phi}$.

STEP 3. Given the initial estimates $\widetilde{\theta}$ and $\widetilde{\phi}$, we replace λ_t^j with $\widetilde{\lambda}_t^j = \lambda_t^j(\widetilde{\phi};\widetilde{\theta})$. Then let $\widehat{\sigma}^j(s)$ optimize $\widetilde{L}_T^j(\sigma^j(s);\widetilde{\lambda}, s)$ with respect to $\sigma^j(s)$. Rescale $\widehat{\sigma}^j(t/T) = \widehat{\sigma}^j(t/T) - \frac{1}{T}\sum_{t=1}^{T}\widehat{\sigma}^j(t/T) = \sum_{i=1}^{\infty}\widehat{\theta}_i^j\psi_i^j(s)$. Update ϕ by optimizing $l_T(\phi;\widehat{\theta})$with respect to $\phi \in \Phi$ to give $\widehat{\phi}$.

STEP 4. Repeat Steps 2–3 to update $\widehat{\theta}$ and $\widehat{\phi}$ until convergence. We define convergence in terms of the distance measure

$$\Delta_r = \sum_{j=D,N}\int\left[\widehat{\sigma}^{j,[r]}(u) - \widehat{\sigma}^{j,[r-1]}(u)\right]^2 du + \left(\widehat{\phi}^{[r]} - \widehat{\phi}^{[r-1]}\right)^{\intercal}\left(\widehat{\phi}^{[r]} - \widehat{\phi}^{[r-1]}\right),$$

that is, we stop when $\Delta_r \le \epsilon$ for some prespecified small ϵ.

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 Appendix B. Let $h_t^j = \lambda_t^j + \sigma^j(t/T)$, and let:

$$A_t = \begin{bmatrix} 1 & a_t^{DN} \\ 0 & 1 \end{bmatrix}, \quad B_{t-1} = \begin{bmatrix} \left(\beta_D + a_{t-1}^{DD}\right) & 0 \\ a_{t-1}^{ND} & \left(\beta_N + a_{t-1}^{NN}\right) \end{bmatrix}, \tag{9}$$

$$a_{t-1}^{DD} = -2\left(\gamma_D + \gamma_D^*\mathrm{sign}(u_{t-1}^D)\right)(v_D+1)\, b_{t-1}^D\left(1 - b_{t-1}^D\right)$$

$$a_t^{DN} = -2\left(\rho_D + \rho_D^*\mathrm{sign}(u_t^N)\right)(v_N+1)\, b_t^N\left(1 - b_t^N\right)$$

$$a_{t-1}^{NN} = -2\left(\gamma_N + \gamma_N^* \text{sign}(u_{t-1}^N)\right)(\upsilon_N + 1)\, b_{t-1}^N \left(1 - b_{t-1}^N\right)$$
$$a_{t-1}^{ND} = -2\left(\rho_N + \rho_N^* \text{sign}(u_{t-1}^D)\right)(\upsilon_D + 1)\, b_{t-1}^D \left(1 - b_{t-1}^D\right)$$
$$b_t^D = \frac{(e_t^D)^2}{\upsilon_D \exp(2\lambda_t^D) + (e_t^D)^2} \quad ; \quad b_t^N = \frac{(e_t^N)^2}{\upsilon_N \exp(2\lambda_t^N) + (e_t^N)^2}.$$

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

$$\|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} \left|a_{ij}\right|. \tag{10}$$

Assumptions A

1. $\|E(A_t \otimes A_t)\|_\infty < \infty$, $\|EB_tEA_t\|_\infty < 1$, $\|E(B_{t-1}A_{t-1} \otimes B_{t-1}A_{t-1})\|_\infty < \|EB_tEA_t\|_\infty$, *and the top-Lyapunov exponent of the sequence of* A_tB_{t-1} *is strictly negative. The top Lyapunov exponent is defined as Theorem 4.26 of* Douc et al. (2014).
2. $0 \le \left|\beta_j\right| < 1$.
3. h_t^j *starts from the infinite past. The parameter* ϕ_0 *is an interior point of* $\Phi \subset \mathbb{R}^{14}$, *where* Φ *is the parameter space of* ϕ_0.
4. *The functions* σ^j *are twice continuously differentiable on* $[0, 1]$, $j = D, N$.
5. $E|u_t^j|^{(2+\delta)\alpha} < \infty$ *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* $|K(s) - K(s')| \le L|s - s'|$ *for some* $L < \infty$ *and all* $s, s' \in \mathbb{R}$. *Denote* $\|K\|_2^2 = \int K(s)^2 ds$.
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 $\widetilde{\sigma}^j(s)$, in line with Vogt and Linton (2014) and Vogt (2012). But we only require that $E|u_t^j|^{\alpha(2+\delta)} < \infty$, since we use $\widetilde{\sigma}^j(s) = \log(T^{-1}\sum_{t=1}^T K_h(s - t/T)|u_t^j|^\alpha)/\alpha$. 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_t^N in A_t follows a beta distribution.

Lemma 1 in Appendix A gives the uniform convergence rate of the initial estimator $\widetilde{\sigma}^j(s)$, 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 λ_t^j.

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 Appendix B. Theorem 1 implies that the score functions with respect to θ and ϕ are asymptotically orthogonal. The intuition behind this is that σ^j is a function of a deterministic variable, t/T, while λ_t^j 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

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

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

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

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

$$\sqrt{Th}\begin{pmatrix}\widehat{\sigma}^D(s) & -\ \sigma_0^D(s)\\ \widehat{\sigma}^N(s) & \sigma_0^N(s)\end{pmatrix} \Longrightarrow N\left(0, \|K\|_2^2\begin{pmatrix}\frac{(\upsilon_D+3)}{2\upsilon_D} & 0\\ 0 & \frac{(\upsilon_N+3)}{2\upsilon_N}\end{pmatrix}\right). \tag{11}$$

Theorem 2 shows that $\tilde{\phi}$ and $\hat{\phi}$ have the same asymptotic property and are efficient. The form of the limiting variance in (11) 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 Theorems 2 and 3 are provided in Appendix B. The information matrix, $\mathcal{I}(\phi_0)$, can be computed explicitly. We can conduct inference with Theorems 2 and 3 using plug-in estimates of the unknown quantities. In the application we present various Wald statistics for testing hypotheses about ϕ 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

$$H_0 : \exp\left(\sigma_0^N(s)\right) = \rho \exp\left(\sigma_0^D(s)\right) \text{ for some } \rho \in \mathbb{R}_+, \quad \text{for all } s \in (0,1),$$

versus the general alternative that the ratio $\exp\left(\sigma_0^N(s)\right)/\exp\left(\sigma_0^D(s)\right)$ is time varying. By Theorem 3 and the delta method, $\exp(\widehat{\sigma}^D(s))$ and $\exp(\widehat{\sigma}^N(s))$ converge jointly to a normal distribution, and are asymptotically mutually independent. Therefore, consider the t-ratio

$$\widehat{t}(s) = \frac{\sqrt{Th}\left(\widehat{\rho}(s) - \widehat{\rho}\right)}{\sqrt{\widehat{\omega}(s)}},$$

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

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

Large values of $|\widehat{t}(s)|$ are inconsistent with the null hypothesis. It follows from Theorem 3 that for $s \in (0,1)$, $\widehat{t}(s) \Longrightarrow N(0,1)$ 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

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

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, $E(\widehat{t}(s)^2) \simeq 1$ so we take $a_T = 1$. Under the null hypothesis

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

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 $\widehat{t}(s_l)$ and $\widehat{t}(s_k)$ are asymptotically independent for $l \neq k$, we may take $b_T = \sqrt{2}$, because $E(\widehat{t}(s)^4) \simeq 3$. Under the null hypothesis, $\tau \Longrightarrow N(0,1)$, 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

$$r_t = \begin{pmatrix} r_t^D \\ r_t^N \end{pmatrix}; \ \mu = \begin{pmatrix} \mu_D \\ \mu_N \end{pmatrix},$$

where r_t^D, r_t^D 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_t^D and u_t^N are mean zero shocks, while

$$\mathbf{D} = \begin{pmatrix} I_n & \operatorname{diag}(\Delta) \\ 0 & I_n \end{pmatrix}; \quad \Pi = \begin{pmatrix} \operatorname{diag}(\Pi_{11}) & \operatorname{diag}(\Pi_{12}) \\ \operatorname{diag}(\Pi_{21}) & \operatorname{diag}(\Pi_{22}) \end{pmatrix},$$

and $\Delta, \Pi_{11}, \Pi_{12}, \Pi_{21}$, and Π_{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_{it}^j using a microstructure-adjusted Market Model

$$r_{it}^D = a_i^D + \beta_i^{DD} r_{mt}^D + \beta_i^{DN} r_{mt}^N + u_{it}^D$$

$$r_{it}^N = a_i^N + \beta_i^{NN} r_{mt}^N + \beta_i^{ND} r_{mt-1}^D + u_{it}^N,$$

where r_{mt}^D and r_{mt}^N are the market intraday and overnight returns and r_{it}^D and r_{it}^N 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

$$u_t = \begin{pmatrix} \Sigma^D(t/T)^{\frac{1}{2}} \operatorname{diag}\left(\exp(\lambda_t^D)\right) & 0 \\ 0 & \Sigma^N(t/T)^{\frac{1}{2}} \operatorname{diag}\left(\exp(\lambda_t^N)\right) \end{pmatrix} \begin{pmatrix} \varepsilon_t^D \\ \varepsilon_t^N \end{pmatrix},$$

where: ε_{it}^j 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 λ_t^j are $n \times 1$ vectors. We assume that $\Sigma^D(.)$ and $\Sigma^N(.)$ 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

$$\Sigma^j(s) = \operatorname{diag}\left(\exp(\sigma^j(s))\right) R^j(s) \operatorname{diag}\left(\exp(\sigma^j(s))\right), \quad j = D, N,$$

with $\operatorname{diag}\left(\exp(\sigma^j(s))\right)$ being the volatility matrix and $R^j(s)$ being the correlation matrix with unit diagonal elements and off-diagonal elements $R_{il}^j(s)$ with $-1 \le R_{il}^j(s) \le 1$. For identification, we still assume $\int_0^1 \sigma_i^j(s)ds = 0$, for $i \in \{1, \ldots, n\}$ and $j = D, N$.

As with the univariate model, define $e_t^j = \operatorname{diag}(\exp(\lambda_t^j))\varepsilon_t^j \in \mathbb{R}^n$, and suppose that:

$$\begin{aligned}
m_{it}^j &= \frac{(1 + v_{ij})(e_{it}^j)^2}{v_{ij}\exp(2\lambda_{it}^j) + (e_{it}^j)^2} - 1, \\
\lambda_{it}^D &= \omega_{iD}(1 - \beta_{iD}) + \beta_{iD}\lambda_{it-1}^D + \gamma_{iD} m_{it-1}^D + \rho_{iD} m_{it}^N \\
&\quad + \gamma_{iD}^*(m_{it-1}^D + 1)\operatorname{sign}(u_{it-1}^D) + \rho_{iD}^*(m_{it}^N + 1)\operatorname{sign}(u_{it}^N), \\
\lambda_{it}^N &= \omega_{iN}(1 - \beta_{iN}) + \beta_{iN}\lambda_{it-1}^N + \gamma_{iN} m_{it-1}^N + \rho_{iN} m_{it-1}^D \\
&\quad + \rho_{iN}^*(m_{it-1}^D + 1)\operatorname{sign}(u_{it-1}^D) + \gamma_{iN}^*(m_{it-1}^N + 1)\operatorname{sign}(u_{it-1}^N).
\end{aligned}$$

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

Define ι_i the vector with the ith element 1 and all others 0, so that $\varepsilon_{it}^j = \iota_i^\intercal \operatorname{diag}(\exp(-\lambda_t^j))\Sigma^j(\frac{t}{T})^{-1/2} u_t^j$. The normalized global log-likelihood function is

$$\begin{aligned}
l_T(\phi, \Sigma(\cdot)) &= \frac{1}{T}\sum_{t=1}^T l_t^N(\phi, \Sigma(\cdot)) + l_t^D(\phi, \Sigma(\cdot)) \\
l_t^j(\phi, \Sigma(\cdot)) &= \sum_{i=1}^n \left(-\prod_{i=1}^n \lambda_{it}^j - \frac{v_{ij}+1}{2} \ln\left(1 + \frac{(\iota_i^\intercal \operatorname{diag}\left(\exp(-\lambda_t^j - \sigma^j(t/T))\right)\left(\Sigma^j\left(\frac{t}{T}\right)\right)^{-1/2} u_t^j)^2}{v_{ij}}\right)\right) \\
&\quad - \frac{1}{2}\log\det\Sigma^j\left(\frac{t}{T}\right) + \sum_{i=1}^n \left(\ln\Gamma\left(\frac{v_{ij}+1}{2}\right) - \frac{1}{2}\ln v_{ij} - \ln\Gamma\left(\frac{v_{ij}}{2}\right)\right).
\end{aligned}$$

Our estimation algorithm is as follows. We first define an initial estimator for $\Sigma^j(t/T)$ and then obtain an estimator of ϕ, and then we update them. Suppose that we know Δ, Π and μ. To give an estimator of $\Sigma^j(t/T)$ that is robust to heavy tails, we estimate the volatility parameter

$$\widetilde{\sigma}_i^j(s) = \frac{1}{\alpha}\log\left(\frac{1}{T}\sum_{t=1}^T K_h(s - t/T)\left|u_{it}^j\right|^\alpha\right), \tag{12}$$

and then rescale $\widetilde{\sigma}^j(t/T)$ as

$$\widetilde{\sigma}_i^j(t/T) = \widetilde{\sigma}_i^j(t/T) - \frac{1}{T}\sum_{t=1}^T \widetilde{\sigma}_i^j(t/T). \tag{13}$$

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

$$\widetilde{R}_{ik}^j(s) = \frac{\sum_{t=1}^T K_h(s - t/T) u_{ik}^j u_{ik}^j}{\sqrt{\sum_{t=1}^T K_h(s - \frac{t}{T}) u_{it}^j u_{it}^j \sum_{t=1}^T K_h(s - \frac{t}{T}) u_{kt}^j u_{kt}^j}} \tag{14}$$

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

$$\hat{\tau}_{k,l}(s) = \frac{\sum_{i=1}^{T}\sum_{j=i}^{T-1} K_h(s-\frac{i}{T})K_h(s-\frac{j}{T})\left(I\left\{\left(u_{i,k}-u_{j,k}\right)\left(u_{i,l}-u_{j,l}\right)>0\right\}-I\left\{\left(u_{i,k}-u_{j,k}\right)\left(u_{i,l}-u_{j,l}\right)<0\right\}\right)}{\sum_{i=1}^{T}\sum_{j=i}^{T-1} K_h(s-\frac{i}{T})K_h(s-\frac{j}{T})\left(I\left\{\left(u_{i,k}-u_{j,k}\right)\left(u_{i,l}-u_{j,l}\right)>0\right\}+I\left\{\left(u_{i,k}-u_{j,k}\right)\left(u_{i,l}-u_{j,l}\right)<0\right\}\right)}.$$

Then applying the relation between Kendall tau and the linear correlation coefficient for the elliptical distribution suggested by Lindskog et al. (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

$$\widetilde{\Sigma}^j(s) = \operatorname{diag}\left(\exp(\widetilde{\sigma}^j(s))\right)\widetilde{R}^j(s)\operatorname{diag}\left(\exp(\widetilde{\sigma}^j(s))\right),\; j = D, N.$$

Letting $\widetilde{e}_t^j = \widetilde{\Sigma}^j(\frac{t}{T})^{-1/2}u_t^j$, we obtain $\widetilde{\phi}_i$ by maximizing the univariate log-likelihood function of $\widetilde{e}_{it}^j$ in (8) 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 $\widehat{\Theta}$ with the local likelihood function given $\widetilde{\lambda}_t^j$ and $\widetilde{v}_j$, i.e., maximize the local objective function

$$L_T^j(\Theta;\widetilde{\lambda}, s) = \frac{1}{T}\sum_{t=1}^{T} K_h(s-t/T)\left[\log|\Theta| - \sum_{i=1}^{n}\left(\frac{\widetilde{v}_{ij}+1}{2}\ln\left(1+\frac{(\iota_i^{\intercal}\operatorname{diag}\left(\exp(-\widetilde{\lambda}_t^j)\right)\Theta u_t^j)^2}{\widetilde{v}_{ij}}\right)\right)\right]$$

with respect to $\operatorname{vech}(\Theta)$, and let $\widehat{\Sigma}^j(t/T) = \widehat{\Theta}^{-2}$. 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 λ_t^j 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 λ_t^j are the same as those in the univariate model.

Blanc et al. (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 Section 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.[4] 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 et al. (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 et al. (2001) claim that many calendar effects arise from data-snooping.

Berkman et al. (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 et al. (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.

[4] These 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.

Compared with intraday returns, overnight returns exhibit more negative skewness and leptokurtosis (Table A.1 in Linton and Wu (2018)). 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) .[5]

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 and Wu (2018). Π_{ij} refers to the element of the ith row jth column in the coefficient matrix Π. For the prediction of intraday returns, 12 of the 28 26stocks have significant Π_{11} values that are all negative, and 7 of 28 stocks have significant δ 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, μ_D and μ_N, are positive for most Dow Jones stocks.

Parameters β_D and β_N are significantly different from 1, and ρ_D, γ_D, ρ_N and γ_N are positive and significant; see Table A.3 in Linton and Wu (2018). In addition, we find significant leverage effects, with negative and significant ρ_D^*, γ_D^*, ρ_N^*, and γ_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 ω_D, which determines the unconditional short-run scale, is significantly larger than ω_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, β_j, γ, ρ_j, γ_j^*, ρ_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 et al. (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. We find that the intraday volatility significantly dominates the overnight volatility 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; see Figure A.2 in Linton and Wu (2018). 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 Fig. 1. 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). 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.

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 $\widehat{t}(s)$ 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.

[5] Suppose that hourly stock returns satisfy $r_{ht} \sim \mu_h, \sigma_h^2, \kappa_{3h}, \kappa_{4h}$, which is consistent with French and Roll (1986). Daily (based on a 6-hour trading day) and weekend (66 h from Friday close to Monday open) returns should then satisfy

$$r_{Dt} \sim 6\mu_h, 6\sigma_h^2, \frac{\kappa_{3h}}{\sqrt{6}}, \frac{\kappa_{4h}}{6} \quad ; \; r_{Wt} \sim 66\mu_h, 66\sigma_h^2, \frac{\kappa_{3h}}{\sqrt{66}}, \frac{\kappa_{4h}}{66}.$$

In fact, overnight returns including weekend returns are very leptokurtic.

Please cite this article as: O. Linton and J. Wu, A coupled component DCS-EGARCH model for intraday and overnight volatility. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2019.12.015.

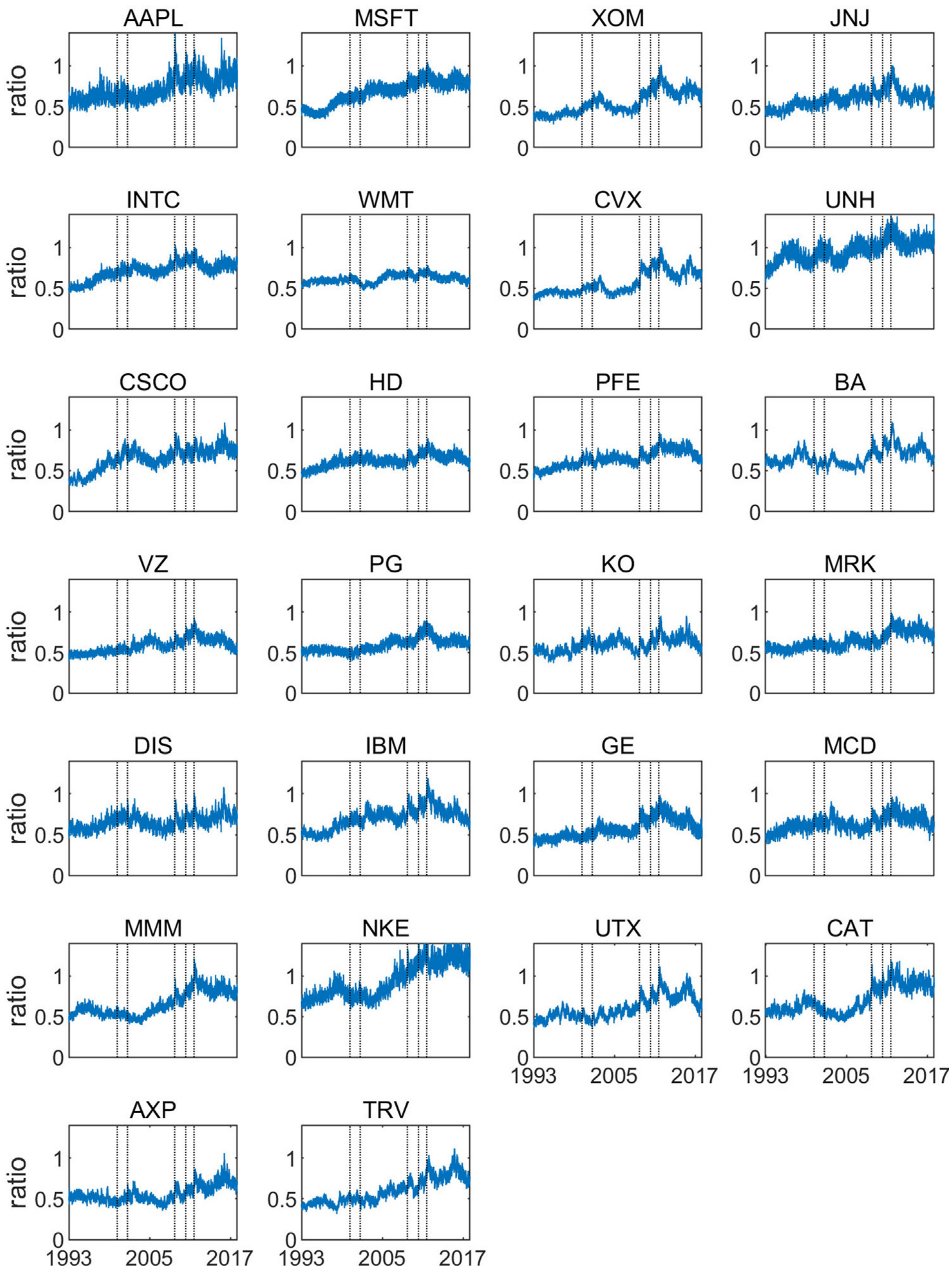

Fig. 1. Ratios of overnight to intraday volatility: univariate model. This figure shows the dynamic ratio of overnight to intraday volatility, based on the univariate coupled-component model, with one subplot for each stock. The five dashed vertical lines from left to right represent the dates: 10 March 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 $\sqrt{\frac{\nu_j}{\nu_j-2}exp(2\lambda_t^j+2\sigma^j(\frac{t}{T}))}$, for $j = D, N$.

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 τ ($s = \tau/T$), we set the two-sided weight function $K_h(s - t/T) = 0$, for $t >= \tau$, 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 et al. (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 β_D and β_N are significantly different from 1, and ρ_D, γ_D, ρ_N and γ_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.

Please cite this article as: O. Linton and J. Wu, A coupled component DCS-EGARCH model for intraday and overnight volatility. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2019.12.015.

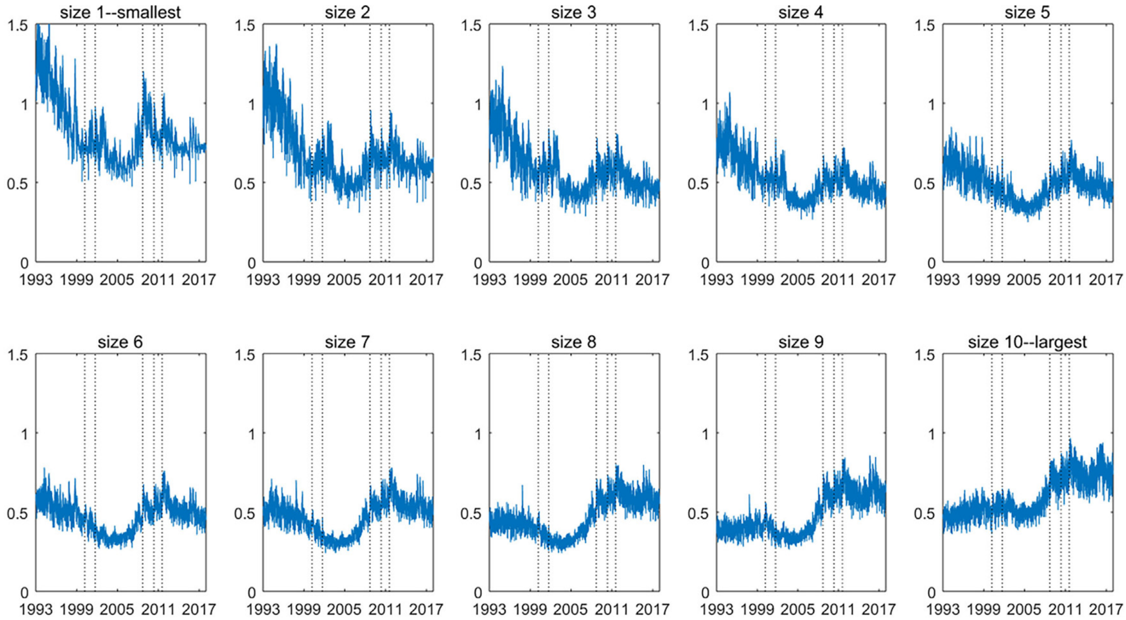

Fig. 2. Ratio of overnight to intraday volatility of size-based portfolios. This figure plots the ratio of overnight to intraday volatility for portfolios formed on size. Decile 1 is the portfolio with the smallest market capitalizations and decile 10 is the portfolio with the largest market capitalizations. The intraday and overnight volatilities are $\sqrt{\frac{v_j}{v_j-2}exp(2\lambda_t^j+2\sigma^j(\frac{t}{T}))}$ 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.

Fig. 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.[6] 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.

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 et al., 1994).

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].

[6] 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.

Appendix A. Lemmas

Lemma 1. *Suppose that Assumptions A1–A8 hold. Then,*

$$\sup_{u\in[0,1]}\left|\widetilde{\sigma}^j(u)-\sigma_0^j(u)\right| = O_p\left(h^2+\sqrt{\frac{\log T}{Th}}\right).$$

$$\int_0^1\left(\widetilde{\sigma}^j(u)-\sigma_0^j(u)\right)^2du = O_p\left(h^2+\sqrt{\frac{1}{Th}}\right).$$

Furthermore $\left\|\tilde{\theta}-\theta\right\|^2 = O_p\left(h^2+\sqrt{\frac{1}{Th}}\right)$.

Proof of Lemma 1. Denote $H^j(s)=\exp(\sigma^j(s))$. We drop the superscript j in what follows and have

$$\begin{aligned}|u_t| &= H(t/T)\,|e_t| = E\,|e_t|\,H(t/T)+H(t/T)\,(|e_t|-E\,|e_t|)\\ \frac{|u_t|}{E\,|e_t|} &= H(t/T)+\frac{H(t/T)}{E\,|e_t|}\,(|e_t|-E\,|e_t|)\\ &=: H(t/T)+\xi_t,\end{aligned}$$

where $E\xi_t = 0$. Suppose we know $E\,|e_t|$. This gives a non-parametric regression function, so we can invoke the Nadaraya–Watson estimator

$$\widetilde{H}(s)^* = \frac{\sum_{t=1}^T K_h(s-t/T)\frac{|u_t|}{E|e_t|}}{\sum_{t=1}^T K_h(s-t/T)}.$$

From Lemma 2, $\{e_t\}$ is a β mixing process with exponential decay, and ξ_t thereby is also a β mixing process with exponential decay. Invoking Theorem 3 in Vogt and Linton (2014), Theorem 4.1 in Vogt (2012) or Kristensen (2009) yields

$$\sup_{s\in[C_1h,1-C_1h]}\left|\widetilde{H}(s)^*-H_0(s)\right| = O_p\left(\sqrt{\frac{\log T}{Th}}+h^2\right).$$

Denote $\widetilde{\sigma}(s)^*=\log\widetilde{H}(s)^*$. Taylor expansion at $H_0(s)$ gives

$$\widetilde{\sigma}(s)^* = \sigma(s)+\left(\widetilde{H}(s)^*-H(s)\right)\frac{1}{H(s)}-\frac{1}{2}\left(\widetilde{H}(s)^*-H(s)\right)^2\frac{1}{\bar{H}(s)^2},$$

where $\bar{H}(s)$ is between $\widetilde{H}(s)^*$ and $H_0(s)$. Therefore,

$$\sup_{s\in[C_1h,1-C_1h]}\left|\widetilde{\sigma}(s)^*-\sigma_0(s)\right| = O_p\left(h^2+\sqrt{\frac{\log T}{Th}}\right).$$

For $s\in[0,h]\cup[1-h,1]$, we use a boundary kernel to ensure the bias property holds through [0, 1].

Until now we have obtained the property for the un-rescaled estimator $\widetilde{\sigma}(s)^*$. Next, we are going to show the convergence rate of the rescaled estimator $\widetilde{\sigma}(s)$. Recall that

$$\widetilde{\sigma}(s)=\widetilde{\sigma}(s)-\frac{1}{T}\sum_{t=1}^T\widetilde{\sigma}(\frac{t}{T}),$$

and we can rewrite $\widetilde{\sigma}(s)$ as:

$$\widetilde{\sigma}(s)=\widetilde{\sigma}(s)^*-\frac{1}{T}\sum_{t=1}^T\widetilde{\sigma}(\frac{t}{T})^*,$$

as $E\,|e_t|$ in $\widetilde{\sigma}(s)^*$ has vanished due to the rescaling. Plugging this into $\sup_{s\in[C_1h,1-C_1h]}|\widetilde{\sigma}(s)-\sigma_0(s)|$ gives

$$\begin{aligned}&\sup_{s\in[0,1]}|\widetilde{\sigma}(s)-\sigma_0(s)|\\ &=\sup_{s\in[0,1]}\left|\widetilde{\sigma}(s)^*-\frac{1}{T}\sum_{t=1}^T\widetilde{\sigma}(\frac{t}{T})^*-\sigma_0(s)\right|\\ &=\sup_{s\in[0,1]}\left|\widetilde{\sigma}(s)^*-\frac{1}{T}\sum_{t=1}^T\widetilde{\sigma}(\frac{t}{T})^*-\sigma_0(s)-\frac{1}{T}\sum_{t=1}^T\sigma_0(\frac{t}{T})+\frac{1}{T}\sum_{t=1}^T\sigma_0(\frac{t}{T})\right|\end{aligned}$$

$$\leq \sup_{s\in[0,1]}\left|\widetilde{\sigma}(s)^{*}-\sigma_0(s)\right|+\left|\frac{1}{T}\sum_{t=1}^{T}\left(\widetilde{\sigma}(\frac{t}{T})^{*}-\sigma_0(\frac{t}{T})\right)\right|+\left|\frac{1}{T}\sum_{t=1}^{T}\sigma_0(\frac{t}{T})\right|$$
$$=O_p\left(h^2+\sqrt{\frac{\log T}{Th}}\right)+O_p\left(h^2+\sqrt{\frac{\log T}{Th}}\right)+\left|\frac{1}{T}\sum_{t=1}^{T}\sigma_0(\frac{t}{T})\right|$$
$$=O_p\left(h^2+\sqrt{\frac{\log T}{Th}}\right)+\left|\frac{1}{T}\sum_{t=1}^{T}\sigma_0(\frac{t}{T})\right|.$$

We only have to work out the second term $\left|\frac{1}{T}\sum_{t=1}^{T}\sigma_0(\frac{t}{T})\right|$. According to Theorem 1.3 in Tasaki (2009),

$$\lim_{T\to\infty}T^2\left(\int_0^1\sigma_0(s)ds-\frac{1}{2T}\sum_{t=1}^{T}\sigma_0(\frac{t}{T})-\frac{1}{2T}\sum_{t=0}^{T-1}\sigma_0(\frac{t}{T})\right)=-\frac{1}{12}\left(\sigma_0'(1)-\sigma_0'(0)\right).$$

Since $\int_0^1\sigma_0(s)ds=0$ and $\sigma_0'(1)-\sigma_0'(0)$ is bounded by Assumption A4, it follows

$$\left|\frac{1}{T}\sum_{t=1}^{T}\sigma_0(\frac{t}{T})\right|\leq\left|\frac{1}{2T}\sum_{t=1}^{T}\sigma_0(\frac{t}{T})+\frac{1}{2T}\sum_{t=0}^{T-1}\sigma_0(\frac{t}{T})\right|+\left|\frac{1}{2T}\sum_{t=1}^{T}\sigma_0(\frac{t}{T})-\frac{1}{2T}\sum_{t=0}^{T-1}\sigma_0(\frac{t}{T})\right|$$
$$=O(T^{-2})+\frac{1}{2T}\left|\sigma_0(1)-\sigma_0(0)\right|$$
$$=O(T^{-1}).$$

Therefore, the uniform convergence rate is

$$\sup_{s\in[0,1]}\left|\widetilde{\sigma}(s)-\sigma_0(s)\right|=O_p\left(h^2+\sqrt{\frac{\log T}{Th}}\right)+O(T^{-1})$$
$$=O_p\left(h^2+\sqrt{\frac{\log T}{Th}}\right).$$

The L_2 rate follows by similar arguments.

Recall that $\sigma(s)=\sum_{j=1}^{\infty}\theta_j\psi_j(s)$ for the orthogonal basis ψ_j. By construction $\tilde{\sigma}(s)$ is a member of the same normed space as $\sigma(s)$, in which case we can write $\tilde{\sigma}(s)=\sum_{j=1}^{\infty}\tilde{\theta}_j\psi_j(s)$ for coefficients $\tilde{\theta}_j, j=1,2,\ldots.$ that satisfy $\sum_{j=1}^{\infty}|\tilde{\theta}_j|<\infty$. In particular, let

$$Q(\theta)=\int_0^1\left(\widetilde{\sigma}(s)-\int_0^1\widetilde{\sigma}(u)du-\sum_{k=1}^{\infty}\theta_k\psi_k(s)\right)^2ds.$$

We have for $k=1,2,\ldots$

$$\frac{\partial Q}{\partial\theta_k}(\theta)=\int_0^1\left(\widetilde{\sigma}(s)-\int_0^1\widetilde{\sigma}(u)du-\sum_{k=1}^{\infty}\theta_k\psi_k(s)\right)\psi_k(s)ds$$

and so

$$\widetilde{\theta}_k=\int_0^1\left(\widetilde{\sigma}(s)-\int_0^1\widetilde{\sigma}(u)du\right)\psi_k(s)ds=\int_0^1\widetilde{\sigma}(s)\psi_k(s)ds,$$

since $\int_0^1\psi_k(s)ds=0$. We have $Q(\widetilde{\theta})=0$. The coefficients satisfy $\widetilde{\theta}_k-\theta_k=\int_0^1\left(\widetilde{\sigma}(s)-\sigma(s)\right)\psi_k(s)ds$.

We have

$$\int\left(\tilde{\sigma}(s)-\sigma(s)\right)^2ds=\int\left(\sum_{j=1}^{\infty}\left(\tilde{\theta}_j-\theta_j\right)\psi_j(s)\right)^2ds$$
$$=\sum_{j=1}^{\infty}\left(\tilde{\theta}_j-\theta_j\right)^2\int\psi_j^2(s)ds$$
$$=\sum_{j=1}^{\infty}\left(\tilde{\theta}_j-\theta_j\right)^2=\left\|\tilde{\theta}-\theta\right\|^2$$

under the assumption that ψ_j are orthonormal. So given the L_2 rate of convergence of $\tilde{\sigma}$ we have the same convergence rate for the implied coefficients. ■

Lemma 2. *If $\left|\beta_j\right| < 1, j = D, N$, then e_t^j and λ_t^j are strictly stationary and β-mixing with exponential decay.*

Proof of Lemma 2. For simplicity, we consider the model without leverage effects

$$\lambda_t^D = \omega_D(1-\beta_D) + \beta_D\lambda_{t-1}^D + \gamma_D m_{t-1}^D + \rho_D m_t^N$$
$$\lambda_t^N = \omega_N(1-\beta_N) + \beta_N\lambda_{t-1}^N + \gamma_N m_{t-1}^N + \rho_N m_{t-1}^D.$$

Let us write it as

$$\begin{pmatrix} \lambda_t^D \\ \lambda_t^N \\ m_t^D \\ m_t^N \end{pmatrix} = \begin{pmatrix} \beta_D & 0 & \gamma_D & 0 \\ 0 & \beta_N & \rho_N & \beta_N \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_{t-1}^D \\ \lambda_{t-1}^N \\ m_{t-1}^D \\ m_{t-1}^N \end{pmatrix} + \begin{pmatrix} \rho_D m_t^N + \omega_D(1-\beta_D) \\ \omega_N(1-\beta_N) \\ m_t^D \\ m_t^N \end{pmatrix}.$$

Since m_t^N and m_t^D are i.i.d random variables and follow a beta distribution, we can easily find an integer $s \geq 1$ to satisfy

$$E\begin{vmatrix} \rho_D m_t^N + \omega_D(1-\beta_D) \\ \omega_N(1-\beta_N) \\ m_t^D \\ m_t^N \end{vmatrix}^s < \infty$$

(Condition A_2 in Carrasco and Chen (2002)). The largest eigenvalue of the matrix

$$\begin{vmatrix} \beta_D & 0 & \gamma_D & 0 \\ 0 & \beta_N & \rho_N & \beta_N \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{vmatrix}$$

is smaller than 1 by assumption. Define $X_t = \begin{pmatrix} \lambda_t^D & \lambda_t^N & m_t^D & m_t^N \end{pmatrix}^\top$. According to Proposition 2 in Carrasco and Chen (2002), the process X_t is Markov geometrically ergodic and $E\left|X_t\right|^s < \infty$. Moreover, if X_t is initialized from the invariant distribution, it is then strictly stationary and β-mixing with exponential decay. The process $\{e_t^j\}$ is a generalized hidden Markov model and stationary β-mixing with a decay rate at least as fast as that of $\{\lambda_t^j\}$ by Proposition 4 in Carrasco and Chen (2002). The extension to the model with leverage effects is straightforward, by defining $X_t = \left(\lambda_t^D \quad \lambda_t^N \quad m_t^D \quad m_t^N \quad \text{sign}\left(e_t^D\right) \quad \text{sign}\left(e_t^N\right)\right)^\top$. ■

Lemma 3. *The score functions of h_t^j with respect to β_D, υ_D and $\sigma^j(t/T)$ are*

$$\begin{pmatrix} \frac{\partial}{\partial\beta_D}h_t^D \\ \frac{\partial}{\partial\beta_D}h_t^N \end{pmatrix} = A_t\begin{pmatrix} \lambda_{t-1}^D - \omega_D \\ 0 \end{pmatrix} + A_tB_{t-1}\begin{pmatrix} \frac{\partial}{\partial\beta_D}h_{t-1}^D \\ \frac{\partial}{\partial\beta_D}h_{t-1}^N \end{pmatrix} \tag{15}$$
$$= \sum_{j=1}^{\infty} A_t \prod_{i=1}^{j-1} B_{t-i}A_{t-i}\begin{pmatrix} \lambda_{t-j}^D - \omega_D \\ 0 \end{pmatrix}.$$

$$\begin{pmatrix} \frac{\partial h_t^D}{\partial\sigma^D(t-k/T)} \\ \frac{\partial h_t^N}{\partial\sigma^D(t-k/T)} \end{pmatrix} = A_tB_{t-1}\begin{pmatrix} \frac{\partial h_{t-1}^D}{\partial\sigma^D(t-k/T)} \\ \frac{\partial h_{t-1}^N}{\partial\sigma^D(t-k/T)} \end{pmatrix} \tag{16}$$
$$= A_t\left(\prod_{i=1}^{k-1} B_{t-i}A_{t-i}\right)\Lambda_{t-k}, k > 1$$

$$\begin{pmatrix} \frac{\partial h_t^D}{\partial\sigma^D(t/T)} \\ \frac{\partial h_t^N}{\partial\sigma^D(t/T)} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}; \text{ and } \begin{pmatrix} \frac{\partial h_t^D}{\partial\sigma^D(t-1/T)} \\ \frac{\partial h_t^N}{\partial\sigma^D(t-1/T)} \end{pmatrix} = A_t\begin{pmatrix} a_{t-1}^{DD} \\ a_{t-1}^{ND} \end{pmatrix},$$

with $\Lambda_t = \begin{pmatrix} a_t^{DD} \\ a_t^{ND} \end{pmatrix}$*. If the top-Lyapunov exponent of the sequence of* $A_t B_{t-1}$ *is strictly negative,* $\begin{pmatrix} \frac{\partial}{\partial \beta_D} h_t^D \\ \frac{\partial}{\partial \beta_D} h_t^N \end{pmatrix}$, $\begin{pmatrix} \frac{\partial h_t^D}{\partial \sigma^D(t-k/T)} \\ \frac{\partial h_t^N}{\partial \sigma^D(t-k/T)} \end{pmatrix}$ *and* $\begin{pmatrix} \frac{\partial h_t^D}{\partial \sigma^D(t-k/T)} \\ \frac{\partial h_t^N}{\partial \sigma^D(t-k/T)} \end{pmatrix} \begin{pmatrix} \frac{\partial h_t^D}{\partial \beta_D} & \frac{\partial h_t^N}{\partial \beta_D} \end{pmatrix}$ *are strictly stationary.*

Proof of Lemma 3. Since $h_t^j = \lambda_t^j + \sigma^j(t/T)$, we can write h_t^j in a recursive formula as

$$\begin{aligned} h_t^D &= \sigma^D(t/T) - \beta_D \sigma^D(\frac{t-1}{T}) + \omega_D(1-\beta_D) + \beta_D h_{t-1}^D + \gamma_D m_{t-1}^D \\ &\quad + \rho_D m_t^N + \gamma_D^*(m_{t-1}^D + 1)\text{sign}(u_{t-1}^D) + \rho_D^*(m_t^N + 1)\text{sign}(u_t^N) \end{aligned} \tag{17}$$

$$\begin{aligned} h_t^N &= \sigma^N(t/T) - \beta_N \sigma^N(\frac{t-1}{T}) + \omega_N(1-\beta_N) + \beta_N h_{t-1}^N + \gamma_N m_{t-1}^N \\ &\quad + \rho_N m_{t-1}^D + \rho_N^*(m_{t-1}^D + 1)\text{sign}(u_{t-1}^D) + \gamma_N^*(m_{t-1}^N + 1)\text{sign}(u_{t-1}^N). \end{aligned} \tag{18}$$

and m_t^D and m_t^N can be expressed as

$$m_t^D = \frac{(1+\upsilon_D)(u_t^D)^2 \exp(-2h_t^D)}{\upsilon_D + (u_t^D)^2 \exp(-2h_t^D)} - 1, \quad \upsilon_D > 0$$
$$m_t^N = \frac{(1+\upsilon_N)(u_t^N)^2 \exp(-2h_t^N)}{\upsilon_N + (u_t^N)^2 \exp(-2h_t^N)} - 1, \quad \upsilon_N > 0.$$

Taking the first order derivative of Eqs. (17) and (18) with respect to β_D gives

$$\begin{aligned} \frac{\partial h_t^D}{\partial \beta_D} &= -\sigma^D(\frac{t-1}{T}) - \omega_D + h_{t-1}^D + \beta_D \frac{\partial}{\partial \beta_D} h_{t-1}^D + \frac{\partial}{\partial \beta_D} \gamma_D m_{t-1}^D + \frac{\partial}{\partial \beta_D} \rho_D m_t^N \\ &\quad + \frac{\partial}{\partial \beta_D} \gamma_D^*(m_{t-1}^D + 1)\text{sign}(u_{t-1}^D) + \frac{\partial}{\partial \beta_D} \rho_D^*(m_t^N + 1)\text{sign}(u_t^N) \end{aligned} \tag{19}$$

$$\begin{aligned} \frac{\partial h_t^N}{\partial \beta_D} &= \beta_N \frac{\partial}{\partial \beta_D} h_{t-1}^N + \frac{\partial}{\partial \beta_D} \gamma_N m_{t-1}^N + \frac{\partial}{\partial \beta_D} \rho_N m_{t-1}^D \\ &\quad + \frac{\partial}{\partial \beta_D} \rho_N^*(m_{t-1}^D + 1)\text{sign}(u_{t-1}^D) + \frac{\partial}{\partial \beta_D} \gamma_N^*(m_{t-1}^N + 1)\text{sign}(u_{t-1}^N) \end{aligned} \tag{20}$$

and the derivatives of m_{t-1}^D and m_{t-1}^N are

$$\frac{\partial}{\partial \beta_D} m_{t-1}^D = \frac{\partial m_{t-1}^D}{\partial h_{t-1}^D} \frac{\partial}{\partial \beta_D} h_{t-1}^D = -2\,(\upsilon_D + 1)\, b_{t-1}^D \left(1 - b_{t-1}^D\right) \frac{\partial}{\partial \beta_D} h_{t-1}^D$$
$$\frac{\partial}{\partial \beta_D} m_{t-1}^N = \frac{\partial m_{t-1}^N}{\partial h_{t-1}^N} \frac{\partial}{\partial \beta_D} h_{t-1}^N = -2\,(\upsilon_N + 1)\, b_{t-1}^N \left(1 - b_{t-1}^N\right) \frac{\partial}{\partial \beta_D} h_{t-1}^N.$$

Substituting them back into (19) and (20) gives

$$\frac{\partial h_t^D}{\partial \beta_D} = \lambda_{t-1}^D - \omega_D + \left(\beta_D + a_{t-1}^{DD}\right) \frac{\partial}{\partial \phi} h_{t-1}^D + a_t^{DN} \frac{\partial}{\partial \phi} h_t^N$$
$$\frac{\partial h_t^N}{\partial \beta_D} = 0 + \left(\beta_N + a_{t-1}^{NN}\right) \frac{\partial}{\partial \phi} h_{t-1}^N + a_{t-1}^{ND} \frac{\partial}{\partial \phi} h_{t-1}^D$$

with the matrix form

$$\begin{pmatrix} \frac{\partial}{\partial \beta_D} h_t^D \\ \frac{\partial}{\partial \beta_D} h_t^N \end{pmatrix} = A_t \begin{pmatrix} \lambda_{t-1}^D - \omega_D \\ 0 \end{pmatrix} + A_t B_{t-1} \begin{pmatrix} \frac{\partial}{\partial \beta_D} h_{t-1}^D \\ \frac{\partial}{\partial \beta_D} h_{t-1}^N \end{pmatrix}.$$

Note that $A_t B_{t-1}$ and $A_t \begin{pmatrix} \lambda_{t-1}^D - \omega_D \\ 0 \end{pmatrix}$ are strictly stationary and ergodic, by Theorem 4.27 in Douc et al. (2014), when the top-Lyapunov exponent of the sequence of $A_t B_{t-1}$ is strictly negative, $\begin{pmatrix} \frac{\partial}{\partial \beta_D} h_t^D \\ \frac{\partial}{\partial \beta_D} h_t^N \end{pmatrix}$ converges and is strictly stationary.

Likewise, taking the first order derivative of h_t^j with respect to $\sigma^D\left(\frac{t-k}{T}\right)$ yields

$$\frac{\partial h_t^D}{\partial \sigma^D((t-k)/T)} = \left(\beta_D + a_{t-1}^{DD}\right)\frac{\partial h_{t-1}^D}{\partial \sigma^D((t-k)/T)} + a_t^{DN}\frac{\partial h_t^N}{\partial \sigma^D((t-k)/T)},\ k>1$$
$$\frac{\partial h_t^D}{\partial \sigma^D(t/T)} = 1,\ \frac{\partial h_t^D}{\partial \sigma^D((t-1)/T)} = a_{t-1}^{DD} + a_t^{DN}a_{t-1}^{ND}$$
$$\frac{\partial h_t^N}{\partial \sigma^D((t-k)/T)} = \left(\beta_N + a_{t-1}^{NN}\right)\frac{\partial h_{t-1}^N}{\partial \sigma^D((t-k)/T)} + a_{t-1}^{ND}\frac{\partial h_{t-1}^D}{\partial \sigma^D((t-k)/T)},\ k>1$$
$$\frac{\partial h_t^N}{\partial \sigma^D(t/T)} = 0,\ \frac{\partial h_t^N}{\partial \sigma^D((t-1)/T)} = a_{t-1}^{ND},$$

and (16) follows. Similarly, $\begin{pmatrix}\frac{\partial h_t^D}{\partial \sigma^D((t-k)/T)}\\ \frac{\partial h_t^N}{\partial \sigma^D((t-k)/T)}\end{pmatrix}$ is strictly stationary across time t.

Finally, we can write

$$\begin{pmatrix}\frac{\partial h_t^D}{\partial \sigma^D((t-k)/T)}\\ \frac{\partial h_t^N}{\partial \sigma^D((t-k)/T)}\\ \frac{\partial h_t^D}{\partial \beta_D}\\ \frac{\partial h_t^N}{\partial \beta_D}\end{pmatrix} = \begin{pmatrix}A_tB_{t-1} & 0\\ 0 & A_tB_{t-1}\end{pmatrix}\begin{pmatrix}\frac{\partial h_{t-1}^D}{\partial \sigma^D((t-k)/T)}\\ \frac{\partial h_{t-1}^N}{\partial \sigma^D((t-k)/T)}\\ \frac{\partial h_{t-1}^D}{\partial \beta_D}\\ \frac{\partial h_{t-1}^N}{\partial \beta_D}\end{pmatrix} + \begin{pmatrix}A_t\begin{pmatrix}\lambda_{t-1}^D - \omega_D\\ 0\end{pmatrix}\\ 0\end{pmatrix}.$$

Both $\begin{pmatrix}\frac{\partial h_t^D}{\partial \sigma^D((t-k)/T)}\\ \frac{\partial h_t^N}{\partial \sigma^D((t-k)/T)}\\ \frac{\partial h_t^D}{\partial \beta_D}\\ \frac{\partial h_t^N}{\partial \beta_D}\end{pmatrix}$ and $\begin{pmatrix}\frac{\partial h_t^D}{\partial \sigma^D((t-k)/T)}\\ \frac{\partial h_t^N}{\partial \sigma^D((t-k)/T)}\end{pmatrix}\begin{pmatrix}\frac{\partial h_t^D}{\partial \beta_D} & \frac{\partial h_t^N}{\partial \beta_D}\end{pmatrix}$ are strictly stationary, since the top-Lyapunov exponent of the sequence $\begin{pmatrix}A_tB_{t-1} & 0\\ 0 & A_tB_{t-1}\end{pmatrix}$, same as that of A_tB_{t-1}, is strictly negative by assumption. ∎

Lemma 4. *Suppose that Assumptions A1–A4 hold. Then,*

$$\sum_k k\left\|E\left(\begin{bmatrix}\frac{\partial h_t^D}{\partial \sigma^D(t-k/T)}\\ \frac{\partial h_t^N}{\partial \sigma^D(t-k/T)}\end{bmatrix}\begin{pmatrix}\frac{\partial}{\partial \beta_D}h_t^D & \frac{\partial}{\partial \beta_D}h_t^N\end{pmatrix}\right)\right\|_\infty < \infty.$$

Proof of Lemma 4. By (15) and (16), we have

$$E\begin{pmatrix}\frac{\partial h_{t+1}^D}{\partial \sigma^D(t+k/T)}\\ \frac{\partial h_{t+1}^N}{\partial \sigma^D(t+k/T)}\end{pmatrix}\begin{pmatrix}\frac{\partial}{\partial \beta_D}h_{t+1}^D & \frac{\partial}{\partial \beta_D}h_{t+1}^N\end{pmatrix} = EA_{t+1}\begin{pmatrix}a_t^{DD}\\ a_t^{ND}\end{pmatrix}\begin{pmatrix}\lambda_t^D - \omega_D & 0\end{pmatrix}A_{t+1}^T;\ k=1$$

$$E\begin{pmatrix}\frac{\partial h_{t+1}^D}{\partial \sigma^D(t+k/T)}\\ \frac{\partial h_{t+1}^N}{\partial \sigma^D(t+k/T)}\end{pmatrix}\begin{pmatrix}\frac{\partial}{\partial \beta_D}h_{t+1}^D & \frac{\partial}{\partial \beta_D}h_{t+1}^N\end{pmatrix} = E\begin{pmatrix}1\\ 0\end{pmatrix}\begin{pmatrix}\lambda_t^D - \omega_D & 0\end{pmatrix}A_{t+1}^T = 0;\ k=0.$$

When $k>1$, it holds

$$\begin{aligned}&E\mathrm{vec}\begin{pmatrix}\frac{\partial h_t^D}{\partial \sigma^D(t-k/T)}\\ \frac{\partial h_t^N}{\partial \sigma^D(t-k/T)}\end{pmatrix}\begin{pmatrix}\frac{\partial}{\partial \beta_D}h_t^D & \frac{\partial}{\partial \beta_D}h_t^N\end{pmatrix}\\ &= E\mathrm{vec}A_t\left(\prod_{i=1}^{k-1}B_{t-i}A_{t-i}\right)\Lambda_{t-k}\begin{pmatrix}\lambda_{t-1}^D - \omega_D & 0\end{pmatrix}A_t^T\end{aligned}$$

$$+ \ E\mathrm{vec}A_t\left(\prod_{i=1}^{k-1}B_{t-i}A_{t-i}\right)\Lambda_{t-k}\left(\lambda_{t-2}^D-\omega_D \quad 0\right)A_{t-1}^TB_{t-1}^TA_t^T$$
$$+\cdots$$
$$+ \ E\mathrm{vec}A_t\left(\prod_{i=1}^{k-1}B_{t-i}A_{t-i}\right)\Lambda_{t-k}\left(\lambda_{t-k+1}^D-\omega_D \quad 0\right)A_{t-k+2}^TB_{t-k+2}^T\ldots A_{t-1}^TB_{t-1}^TA_t^T$$
$$= E\sum_{j=1}^{k-1}(A_t\otimes A_t)\left(\prod_{i=1}^{j-1}(B_{t-i}\otimes B_{t-i})(A_{t-i}\otimes A_{t-i})\right)\mathrm{vec}\left(\left(\prod_{i=j}^{k-1}B_{t-i}A_{t-i}\right)\Lambda_{t-k}\left(\lambda_{t-j}^D-\omega_D \quad 0\right)\right).$$

Since $(B_{t-1}\otimes B_{t-1})(A_{t-1}\otimes A_{t-1})$ and B_tA_t are i.i.d, and $EB_tA_t=EB_tEA_t$, we obtain

$$E\mathrm{vec}\begin{pmatrix}\frac{\partial h_t^D}{\partial\sigma^D(t-k/T)}\\ \frac{\partial h_t^N}{\partial\sigma^D(t-k/T)}\end{pmatrix}\left(\frac{\partial}{\partial\beta_D}h_t^D \quad \frac{\partial}{\partial\beta_D}h_t^N\right) \tag{21}$$
$$=\sum_{j=1}^{k-1}E\left(A_t\otimes A_t\right)\left(E\left(B_t\otimes B_t\right)\left(A_t\otimes A_t\right)\right)^{j-1}E\mathrm{vec}\left(\left(\prod_{i=j}^{k-1}B_{t-i}A_{t-i}\right)\Lambda_{t-k}\left(\lambda_{t-j}^D-\omega_D \quad 0\right)\right).$$

Let us write λ_{t-1}^D as

$$\lambda_t^D-\omega_D=\gamma_D\sum_{k=1}^{\infty}\beta_D^{k-1}m_{t-k}^D+\rho_D\sum_{k=1}^{\infty}\beta_D^{k-1}m_{t-k+1}^N+\gamma_D^*\sum_{k=1}^{\infty}\beta_D^{k-1}(m_{t-k}^D+1)\mathrm{sign}(e_{t-k}^D)$$
$$+ \ \rho_D^*\sum_{k=1}^{\infty}\beta_D^{k-1}(m_{t-k+1}^N+1)\mathrm{sign}(e_{t-k+1}^N)$$

which is a function of $\left\{\left(m_{t-i}^D, m_{t-i+1}^N\right), i>1\right\}$, and note that B_t, A_t, and Λ_t are independent of $\left\{\left(m_s^D, m_s^N\right), s\neq t\right\}$. Therefore, we have

$$E\left(\left(\prod_{i=j}^{k-1}B_{t-i}A_{t-i}\right)\Lambda_{t-k}\left(\lambda_{t-j}^D-\omega_D\right)\right)$$
$$=\gamma_DE\left(\prod_{i=j}^{k-1}B_{t-i}A_{t-i}\right)\Lambda_{t-k}\sum_{i=j+1}^{k}\beta_D^{i-1-j}\left(m_{t-i}^D+\left(m_{t-i}^D+1\right)\mathrm{sign}(e_{t-i}^D)\right)$$
$$+ \ \rho_DE\left(\prod_{i=j}^{k-1}B_{t-i}A_{t-i}\right)\Lambda_{t-k}\sum_{i=j+1}^{k}\beta_D^{i-1-j}\left(m_{t-i+1}^N+(m_{t-i+1}^N+1)\mathrm{sign}(e_{t-i+1}^N)\right),$$

with the first term

$$\left\|E\left(\prod_{i=j}^{k-1}B_{t-i}A_{t-i}\right)\Lambda_{t-k}\sum_{i=j+1}^{k}\beta_D^{i-1-j}\left(m_{t-i}^D+\left(m_{t-i}^D+1\right)\mathrm{sign}(e_{t-i}^D)\right)\right\|_\infty$$
$$\leq\left(\sum_{i=j+1}^{k-1}\beta_D^{i-1-j}\right)\left\|E\left(B_t\left(m_t^D+(m_t^D+1)\mathrm{sign}(e_t^D)\right)A_t\right)\right\|_\infty\|EB_tEA_t\|_\infty^{k-j-1}\|E\Lambda_t\|_\infty$$
$$+ \ \beta_D^{k-j-1}\left\|E\Lambda_{t-k}\left(m_{t-k}^D+(m_{t-k}^D+1)\mathrm{sign}(e_{t-k}^D)\right)\right\|_\infty\|EB_tEA_t\|_\infty^{k-j}$$
$$\leq\frac{1}{1-\beta_D}\left\|E\left(B_t\left(m_t^D+(m_t^D+1)\mathrm{sign}(e_t^D)\right)A_t\right)\right\|_\infty\|E\Lambda_t\|_\infty\|EB_tEA_t\|_\infty^{k-j-1}$$
$$+ \ \beta_D^{k-j-1}\left\|E\Lambda_{t-k}\left(m_{t-k}^D+(m_{t-k}^D+1)\mathrm{sign}(e_{t-k}^D)\right)\right\|_\infty\|EB_tEA_t\|_\infty^{k-j}$$

Please cite this article as: O. Linton and J. Wu, A coupled component DCS-EGARCH model for intraday and overnight volatility. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2019.12.015.

and the second term

$$\left\| E\left(\prod_{i=j}^{k-1} B_{t-i}A_{t-i}\right) \Lambda_{t-k} \sum_{i=j+1}^{k} \beta_D^{i-j-1} \left(m_{t-i+1}^N + (m_{t-i+1}^N + 1)\mathrm{sign}(e_{t-i+1}^N)\right) \right\|_\infty$$
$$\leq \frac{1}{1-\beta_D} \left\| E\left(B_t \left(m_t^N + (m_t^N + 1)\mathrm{sign}(e_t^N)\right) A_t\right)\right\|_\infty \|E\Lambda_t\|_\infty \|EB_tEA_t\|_\infty^{k-j-1}.$$

According to the definition of $\|\|_\infty$,

$$\left\| \mathrm{E}\mathrm{vec}\left(\left(\prod_{i=j}^{k-1} B_{t-i}A_{t-i}\right) \Lambda_{t-k} \begin{pmatrix}\lambda_{t-j}^D - \omega_D & 0\end{pmatrix}\right)\right\|_\infty \leq \left\| E\left(\left(\prod_{i=j}^{k-1} B_{t-i}A_{t-i}\right) \Lambda_{t-k} \begin{pmatrix}\lambda_{t-j}^D - \omega_D & 0\end{pmatrix}\right)\right\|_\infty$$

Therefore,

$$\left\| \mathrm{E}\mathrm{vec}\left(\left(\prod_{i=j}^{k-1} B_{t-i}A_{t-i}\right) \Lambda_{t-k} \begin{pmatrix}\lambda_{t-j}^D - \omega_D & 0\end{pmatrix}\right)\right\|_\infty \leq c_T \|EB_tEA_t\|^{k-j-1} \tag{22}$$

with

$$c_T = \frac{1}{1-\beta_D}|\gamma_D| \left\| E\left(B_t \left(m_t^D + (m_t^D + 1)\mathrm{sign}(e_t^D)\right) A_t\right)\right\|_\infty \|E\Lambda_t\|_\infty$$
$$+ \|EB_tEA_t\|_\infty \left\| E\Lambda_{t-k}\left(m_{t-k}^D + (m_{t-k}^D + 1)\mathrm{sign}(e_{t-k}^D)\right)\right\|_\infty$$
$$+ \frac{1}{1-\beta_D}|\rho_D| \left\| E\left(B_t \left(m_t^N + (m_t^N + 1)\mathrm{sign}(e_t^N)\right) A_t\right)\right\|_\infty \|E\Lambda_t\|_\infty.$$

Substituting (22) into (21) gives

$$\left\| \mathrm{E}\mathrm{vec}\begin{pmatrix}\frac{\partial h_t^D}{\partial\sigma^D(t-k/T)} \\ \frac{\partial h_t^N}{\partial\sigma^D(t-k/T)}\end{pmatrix} \begin{pmatrix}\frac{\partial}{\partial\beta_D}h_t^D & \frac{\partial}{\partial\beta_D}h_t^N\end{pmatrix}\right\|_\infty$$
$$\leq \sum_{j=1}^{k-1} \|E(A_t \otimes A_t)\|_\infty \|E(B_{t-i}\otimes B_{t-i})(A_{t-i}\otimes A_{t-i})\|_\infty^{j-1} c_T \|EB_tEA_t\|_\infty^{k-j-1}$$
$$\leq c_T \|E(A_t \otimes A_t)\|_\infty \sum_{j=1}^{k-1} \|E(B_{t-i}\otimes B_{t-i})(A_{t-i}\otimes A_{t-i})\|_\infty^{j-1} \|EB_tEA_t\|_\infty^{k-j-1}$$
$$\leq c_T \|E(A_t \otimes A_t)\|_\infty \frac{\|EB_tEA_t\|_\infty^{k-2}}{1 - \frac{\|E(B_{t-i}\otimes B_{t-i})(A_{t-i}\otimes A_{t-i})\|_\infty}{\|EB_tEA_t\|_\infty}},$$

provided that $\|EB_tEA_t\|_\infty < 1$ and $\|E(B_{t-1}A_{t-1} \otimes B_{t-1}A_{t-1})\|_\infty < \|EB_tEA_t\|_\infty$. It is then straightforward to show

$$\sum_k k \left\| \mathrm{E}\mathrm{vec}\begin{pmatrix}\frac{\partial h_t^D}{\partial\sigma^D(t-k/T)} \\ \frac{\partial h_t^N}{\partial\sigma^D(t-k/T)}\end{pmatrix} \begin{pmatrix}\frac{\partial}{\partial\beta_D}h_t^D & \frac{\partial}{\partial\beta_D}h_t^N\end{pmatrix}\right\|_\infty < \infty$$

and thereby

$$\sum_k k \left\| E\begin{pmatrix}\frac{\partial h_t^D}{\partial\sigma^D(t-k/T)} \\ \frac{\partial h_t^N}{\partial\sigma^D(t-k/T)}\end{pmatrix} \begin{pmatrix}\frac{\partial}{\partial\beta_D}h_t^D & \frac{\partial}{\partial\beta_D}h_t^N\end{pmatrix}\right\|_\infty < \infty. \quad \blacksquare$$

Lemma 5. *Suppose that Assumptions A1–A4 hold. Then, we have*

$$\frac{1}{T}\sum_{t=1}^{T} E\left(\frac{\partial h_t^D}{\partial\theta}\right) = 0.$$

Proof of Lemma 5. Similar to the proof of Theorem 1, we only need to show $\sum_{t=1}^{T} k \left\| E \begin{pmatrix} \frac{\partial h_t^D}{\partial \sigma^N((t-k)/T)} \\ \frac{\partial h_t^N}{\partial \sigma^N((t-k)T)} \end{pmatrix} \right\|_\infty < \infty$. Note that $E \begin{pmatrix} \frac{\partial h_t^D}{\partial \sigma^N((t-k)/T)} \\ \frac{\partial h_t^N}{\partial \sigma^N((t-k)/T)} \end{pmatrix} = EA_t B_{t-1} A_{t-1} B_{t-2} .. A_{t-k+2} B_{t-k+1} A_{t-k+1} \Lambda_{t-k} = EA_t \left(EB_{t-1} A_{t-1}\right)^{k-1} E\Lambda_{t-k}$, when $k > 1$. Obviously, $\sum_{t=1}^{T} k \left\| E \begin{pmatrix} \frac{\partial h_t^D}{\partial \sigma^N((t-k)/T)} \\ \frac{\partial h_t^N}{\partial \sigma^N((t-k)/T)} \end{pmatrix} \right\|_\infty < \infty$. ■

Appendix B. Proof of main results

B.1. Proof of Theorem 1

Let $\phi_i = \beta_D$ and θ_k be an element in function $\sigma^D(\cdot)$ (for simplicity, the subscript k is omitted in the following explanation). Recall that $h_t^j = \lambda_t^j + \sigma^j(t/T)$, and the log-likelihood function, without unnecessary constant, can be rewritten as a function of h_t^j

$$l_t^j = -h_t^j - \frac{v_j + 1}{2} \ln \left(1 + \frac{(u_t^j)^2}{v_j \exp(2h_t^j)}\right) + \ln \Gamma \left(\frac{v_j + 1}{2}\right) - \frac{1}{2} \ln v_j - \ln \Gamma \left(\frac{v_j}{2}\right)$$

with the score functions

$$\frac{\partial l_t}{\partial \theta} = \frac{\partial l_t^D}{\partial h_t^D} \frac{\partial h_t^D}{\partial \theta} + \frac{\partial l_t^N}{\partial h_t^D} \frac{\partial h_t^D}{\partial \theta} = m_t^D \frac{\partial h_t^D}{\partial \theta} + m_t^N \frac{\partial h_t^D}{\partial \theta}$$

$$\frac{\partial l_t}{\partial \beta_D} = \frac{\partial l_t^D}{\partial h_t^D} \frac{\partial h_t^D}{\partial \beta_D} = m_t^D \frac{\partial h_t^D}{\partial \beta_D} + m_t^N \frac{\partial h_t^D}{\partial \beta_D}.$$

Recall that $m_t^j = (v_j + 1) b_t^j - 1$,with b_t^j independent and identically beta distributed, we have $E\left(m_t^N m_t^D\right) = 0$, $E\left(m_t^j\right)^2$ is time invariant, and $E[(m_t^j)^2] < \infty$. Therefore, we can write

$$\sum_{t=1}^{T} E\left(\frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \beta_D}\right) = E\left(m_t^D\right)^2 \sum_{t=1}^{T} E\left(\frac{\partial h_t^D}{\partial \theta} \frac{\partial h_t^D}{\partial \beta_D}\right) + E\left(m_t^N\right)^2 \sum_{t=1}^{T} E\left(\frac{\partial h_t^N}{\partial \theta} \frac{\partial h_t^N}{\partial \beta_D}\right).$$

To prove the Theorem, it then suffices to show that

$$\left\| \sum_{t=1}^{T} E\left(\begin{pmatrix} \frac{\partial h_t^D}{\partial \theta} \\ \frac{\partial h_t^N}{\partial \theta} \end{pmatrix} \begin{pmatrix} \frac{\partial h_t^D}{\partial \beta_D} & \frac{\partial h_t^N}{\partial \beta_D} \end{pmatrix} \right) \right\|_\infty = O(1).$$

By expressing λ_t^j as a function of ϕ and $\left\{\left(\sigma^j(\frac{t-i}{T}), u_{t-i}^j\right), i \geq 0\right\}$, we can write $\frac{\partial h_t^j}{\partial \theta}$ as

$$\begin{pmatrix} \frac{\partial h_t^D}{\partial \theta} \\ \frac{\partial h_t^N}{\partial \theta} \end{pmatrix} = \sum_{k=0}^{T} \begin{pmatrix} \frac{\partial h_t^D}{\partial \sigma^D(\frac{t-k}{T})} \frac{\partial \sigma^D(\frac{t-k}{T})}{\partial \theta} \\ \frac{\partial h_t^N}{\partial \sigma^D(\frac{t-k}{T})} \frac{\partial \sigma^D(\frac{t-k}{T})}{\partial \theta} \end{pmatrix} = \sum_{k=0}^{T} \begin{pmatrix} \frac{\partial h_t^D}{\partial \sigma^D(\frac{t-k}{T})} \\ \frac{\partial h_t^N}{\partial \sigma^D(\frac{t-k}{T})} \end{pmatrix} \psi_i^D\left(\frac{t-k}{T}\right),$$

when the limit exists. Thus we obtain,

$$\frac{1}{T} \sum_{t=1}^{T} E\left(\begin{pmatrix} \frac{\partial h_t^D}{\partial \theta} \\ \frac{\partial h_t^N}{\partial \theta} \end{pmatrix} \begin{pmatrix} \frac{\partial h_t^D}{\partial \beta_D} & \frac{\partial h_t^N}{\partial \beta_D} \end{pmatrix} \right)$$

$$= \frac{1}{T} \sum_{k=0}^{T} \sum_{t=1}^{T} E\left(\begin{pmatrix} \frac{\partial h_t^D}{\partial \sigma^D(\frac{t-k}{T})} \\ \frac{\partial h_t^N}{\partial \sigma^D(\frac{t-k}{T})} \end{pmatrix} \begin{pmatrix} \frac{\partial h_t^D}{\partial \beta_D} & \frac{\partial h_t^N}{\partial \beta_D} \end{pmatrix} \right) \psi_i^D\left(\frac{t-k}{T}\right)$$

$$= \frac{1}{T} \sum_{k=0}^{T} E\left(\begin{pmatrix} \frac{\partial h_t^D}{\partial \sigma^D(\frac{t-k}{T})} \\ \frac{\partial h_t^N}{\partial \sigma^D(\frac{t-k}{T})} \end{pmatrix} \begin{pmatrix} \frac{\partial h_t^D}{\partial \beta_D} & \frac{\partial h_t^N}{\partial \beta_D} \end{pmatrix} \right) \sum_{t=1}^{T} \psi_i^D\left(\frac{t-k}{T}\right).$$

Please cite this article as: O. Linton and J. Wu, A coupled component DCS-EGARCH model for intraday and overnight volatility. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2019.12.015.

The second equality follows since $E\begin{pmatrix}\frac{\partial h_t^D}{\partial \sigma^j(\frac{t-k}{T})}\\ \frac{\partial h_t^N}{\partial \sigma^j(\frac{t-k}{T})}\end{pmatrix}\begin{pmatrix}\frac{\partial h_t^D}{\partial \beta_D} & \frac{\partial h_t^N}{\partial \beta_D}\end{pmatrix}$ is invariant across time t by Lemma 3. Taylor expansion of $\sum_{t=1}^T \psi_i^D\left(\frac{t-k}{T}\right)$ around $\sum_{t=1}^T \psi_i^D\left(\frac{t}{T}\right)$ gives

$$\begin{aligned}\frac{1}{T}\sum_t \psi_i^D\left(\frac{t-k}{T}\right) &= \frac{1}{T}\sum_t \psi_i^D\left(\frac{t}{T}\right) - \frac{1}{T}\frac{k}{T}\sum_t \psi_i^{D'}\left(\frac{t}{T}\right) + O\left(\frac{k}{T}\right)^2\\ &= O\left(\frac{1}{T}\right) + O\left(\frac{k}{T}\right) + O\left(\frac{k}{T}\right)^2\\ &= O\left(\frac{k}{T}\right).\end{aligned}$$

Hence, it suffices to show

$$\sum_{k=0}^T \left\| kE\left(\begin{pmatrix}\frac{\partial h_t^D}{\partial \sigma^D(\frac{t-k}{T})}\\ \frac{\partial h_t^N}{\partial \sigma^D(\frac{t-k}{T})}\end{pmatrix}\begin{pmatrix}\frac{\partial h_t^D}{\partial \beta_D} & \frac{\partial h_t^N}{\partial \beta_D}\end{pmatrix}\right)\right\|_\infty < \infty,$$

which is obtained by Lemma 4.

The proof with respect to υ_D is similar, but the score function is slightly different. The score functions of l_t^D and l_t^N with respect to υ_D are

$$\begin{aligned}\frac{\partial l_t^D}{\partial \upsilon_D} &= -\frac{1}{2}\ln\left(1+\frac{(u_t^D)^2}{\upsilon_D \exp(2h_t^D)}\right) + \frac{\partial}{\partial \upsilon_D}\left(\ln\Gamma\left(\frac{\upsilon_D+1}{2}\right) - \ln\Gamma\left(\frac{\upsilon_D}{2}\right)\right) - \frac{1}{2\upsilon_D}\\ &\quad + \frac{\upsilon_D+1}{2\left(1+\frac{(u_t^D)^2}{\upsilon_D\exp(2h_t^D)}\right)}\frac{(u_t^D)^2}{\upsilon_D^2\exp(2h_t^D)}\left(1+2\upsilon_D\frac{\partial h_t^D}{\partial \upsilon_D}\right) + \frac{\partial h_t^D}{\partial \upsilon_D}\\ \frac{\partial l_t^N}{\partial \upsilon_D} &= \frac{\upsilon_N+1}{2\left(1+\frac{(u_t^N)^2}{\upsilon_N\exp(2h_t^N)}\right)}\frac{(u_t^N)^2}{\upsilon_N^2\exp(2h_t^N)}\left(1+2\upsilon_N\frac{\partial h_t^N}{\partial \upsilon_D}\right) + \frac{\partial h_t^N}{\partial \upsilon_D}.\end{aligned} \tag{23}$$

Then we have

$$\begin{aligned}&\sum_{t=1}^T E\left(\frac{\partial l_t^D}{\partial \theta}\frac{\partial l_t^D}{\partial \upsilon_D}\right)\\ &= \sum_{t=1}^T E\left(m_t^D\frac{\partial h_t^D}{\partial \theta}\left[\frac{\partial h_t^D}{\partial \upsilon_D} - \frac{1}{2}\ln\left(1+\frac{(u_t^D)^2}{\upsilon_D\exp(2h_t^D)}\right)\right]\right)\\ &+ \sum_{t=1}^T E\left(m_t^D\frac{\partial h_t^D}{\partial \theta}\frac{\partial \ln\Gamma\left(\frac{\upsilon_D+1}{2}\right) - \frac{1}{2}\ln\upsilon_D - \ln\Gamma\left(\frac{\upsilon_D}{2}\right)}{\partial \upsilon_D}\right)\\ &+ \sum_{t=1}^T E\left(m_t^D\frac{\partial h_t^D}{\partial \theta}\left[\frac{\upsilon_D+1}{2\left(1+\frac{(u_t^D)^2}{\upsilon_D\exp(2h_t^D)}\right)}\frac{(u_t^D)^2}{\upsilon_D^2\exp(2h_t^D)}\left(1+2\upsilon_D\frac{\partial h_t^D}{\partial \upsilon_D}\right)\right]\right)\\ &= \frac{1}{2}E\left(m_t^D\left(-\ln\left(1+\frac{(\varepsilon_t^D)^2}{\upsilon_D}\right) + \frac{\upsilon_D+1}{2\upsilon_D+(\varepsilon_t^D)^2}\frac{(\varepsilon_t^D)^2}{\upsilon_D}\right)\right)\frac{1}{T}\sum_{t=1}^T E\left(\frac{\partial h_t^D}{\partial \theta}\right)\\ &+ E\left(m_t^D\left(1+\frac{(\upsilon_D+1)(\varepsilon_t^D)^2}{2\upsilon_D+(\varepsilon_t^D)^2}\right)\right)\frac{1}{T}\sum_{t=1}^T E\left(\frac{\partial h_t^D}{\partial \upsilon_D}\frac{\partial h_t^D}{\partial \theta}\right).\end{aligned}$$

The first term vanishes by Lemma 5. Then we can use the same procedure above to obtain $\sum_{t=1}^T E\left(\frac{\partial h_t^D}{\partial \upsilon_D}\frac{\partial h_t^D}{\partial \theta}\right) = O(1)$, and to finish the proof for υ_D. ■

B.2. Proof of Theorem 2

By the triangle inequality,

$$\sup_{\phi\in\Phi}\left|l_T(\phi;\widetilde{\theta})-l(\phi)\right|\leq\sup_{\phi\in\Phi}\left|l_T(\phi;\widetilde{\theta})-l_T(\phi;\theta_0)\right|+\sup_{\phi\in\Phi}\left|l_T(\phi;\theta_0)-l(\phi)\right|,$$

where $l(\phi)=E(l_T(\phi;\theta_0))$. By the identification condition $l(\phi)$ is uniquely maximized at $\phi=\phi_0$ and standard arguments (Harvey, 2013) show that the second term is $o_p(1)$. The first term is also $o_p(1)$ by the uniform consistency of $\tilde{\sigma}(s)$ in Lemma 1 and the smoothness of the objective function in $\tilde{\sigma}(t/T)$ and equivalently θ_k.

We next turn to asymptotic normality. The general strategy is that we first show the estimators obtained by maximizing $l_T(\phi;\widetilde{\theta})$ and $l_T(\phi;\theta_0)$ have the same asymptotic distribution, provided $\|\tilde{\theta}-\theta_0\|$ converges to 0. As a result, the asymptotic property of $\tilde{\phi}$ follows as in the parametric model with known $\sigma(t/T)$.

Following Severini and Wong (1992), the expansion of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial l_t(\phi_0,\tilde{\theta})}{\partial\phi}$ at θ_0 gives

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial l_t(\phi_0;\tilde{\theta})}{\partial\phi}=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial l_t(\phi_0;\theta_0)}{\partial\phi}+\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\sum_{k}\frac{\partial^2 l_t(\phi_0;\theta_0)}{\partial\phi\partial\theta_k}(\tilde{\theta}_k-\theta_{k,0})\right)+o_p(1). \tag{24}$$

According to Theorem 1, we have $\sum_{t=1}^{T}E(\frac{\partial^2 l_t(\phi;\theta)}{\partial\phi_i\partial\theta_k})=O(1)$, for each k and i, where $k\in\{1,\ldots,\infty\}$ and $i\in\{1,\ldots,14\}$. It follows that

$$\sum_{t=1}^{T}\frac{\partial^2 l_t(\phi_0;\theta_0)}{\partial\phi_i\partial\theta_k}=O_p(\sqrt{T}).$$

Given that the dimension of the sieve space grows slowly and $\tilde{\theta}$ converges to θ, the second term in (24) is of order $o_p(1)$.

Therefore, the asymptotic property of $\tilde{\phi}$ can be obtained with a similar procedure to Harvey (2013). He gives the consistency and asymptotic normality of the estimator for the parametric beta-t-egarch model. The basic idea is that the first three derivatives of l_t with respect to ϕ (except v_j) are linear combinations of $b_t^h(1-b_t)^k$, $h, k=0,1,2,\ldots$, with $b_t=\frac{(1+v)(e_t)^2}{v\exp(2\lambda_t)+(e_t)^2}$. Since b_t is beta distributed, these first three derivatives are all bounded. It is then straightforward to show that the score function satisfies a CLT, and its derivative converges to the information matrix by the ergodic theorem.

Obviously, $\hat{\phi}$ has the same limiting distribution as $\tilde{\phi}$, since $\sum_{t=1}^{T}\frac{\partial l_t(\phi_0,\tilde{\theta})}{\partial\phi}$ and $\sum_{t=1}^{T}\frac{\partial l_t(\phi_0,\theta)}{\partial\phi}$ have the same asymptotic property. ■

B.3. Proof of Theorem 3

Consider the local likelihood function given η_t^j and v_j, i.e., minimize the objective function

$$L_T^j(\sigma^j;s)=\frac{1}{T}\sum_{t=1}^{T}K_h(s-t/T)\left[\sigma^j+\frac{v_j+1}{2}\ln\left(1+\frac{(\eta_t^j\exp(-\sigma^j))^2}{v_j}\right)\right]$$

with respect to ω, for $j=D,N$ separately. The first order and second order derivatives are:

$$\frac{\partial L_T^j(\sigma^j;s)}{\partial\sigma^j}=\frac{1}{T}\sum_{t=1}^{T}K_h(s-t/T)\left[-(v_j+1)b_t^j(\sigma^j)+1\right]$$

$$\frac{\partial^2 L_T^j(\sigma^j;s)}{\partial\sigma^{j2}}=2(v_j+1)\frac{1}{T}\sum_{t=1}^{T}K_h(s-t/T)\left[b_t^j(\sigma^j)\left(1-b_t^j(\sigma^j)\right)\right], \tag{25}$$

where

$$b_t^j(\sigma^j)=\frac{\frac{(\eta_t^j)^2}{v_j}}{\exp(2\sigma^j)+\frac{(\eta_t^j)^2}{v_j}}.$$

We have

$$\sqrt{Th}\left(\widehat{\sigma^j}(s)-\sigma_0^j(s)\right)=\left[\frac{1}{Th}\frac{\partial^2 L_T^j(\sigma_0^j;s)}{\partial\sigma^{j2}}\right]^{-1}\frac{1}{\sqrt{Th}}\frac{\partial L_T^j(\sigma_0^j;s)}{\partial\sigma^j}+o_p(1),$$

Please cite this article as: O. Linton and J. Wu, A coupled component DCS-EGARCH model for intraday and overnight volatility. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2019.12.015.

This is asymptotically normal with mean zero and variance (when the t distribution is correct)

$$\operatorname{var}\left[\frac{1}{\sqrt{Th}}\frac{\partial L_T^j(\sigma_0^j;s)}{\partial\sigma^j}\right]=\|K\|_2^2 E\left[\left(1-(v_j+1)b_t^j(\sigma_0^j(s))\right)^2\right]_{t/T=s}.$$

This follows because

$$E\left[\left(1-(v_j+1)b_t^j(\sigma_0^j(s))\right)^2\right]=f(t/T)$$

for some smooth function f, and recall $\eta_t^j=\exp(\sigma^j(t/T))\varepsilon_t^j$. It follows that

$$\frac{h^2}{Th}\sum_{t=1}^{T}K_h^2(s-t/T)f(t/T)\to\|K\|_2^2 f(s),$$

Therefore,

$$\sqrt{Th}\left(\widehat{\sigma^j}(s)-\sigma_0^j(s)\right)\Longrightarrow N\left(0,\frac{\|K\|_2^2}{E\left[\left(1-(v_j+1)b_t^j\right)^2\right]_{t/T=s}}\right)$$

Further, since b_t^j is distributed as $beta(\frac{1}{2},\frac{v_j}{2})$, with

$$E\left[\left(1-(v_j+1)b_t^j\right)^2\right]_{t/T=s}=\frac{2v_j}{(v_j+3)}.$$

It thus follows that

$$\sqrt{Th}\left(\widehat{\sigma^j}(s)-\sigma_0^j(s)\right)\Longrightarrow N\left(0,\sqrt{\frac{(v_j+3)}{2v_j}}\|K\|_2^2\right).$$

when the t distribution is correct. ■

References

Aboody, D., Even-Tov, O., Lehavy, R., Trueman, B., 2018. Overnight returns and firm-specific investor sentiment. J. Financ. Quant. Anal. 53 (2), 485–505.
Andersen, T.G., Bollerslev, T., Huang, X., 2011. A reduced form framework for modeling volatility of speculative prices based on realized variation measures. J. Econometrics 160 (1), 176–189.
Aretz, K., Bartram, S.M., 2015. Making money while you sleep? anomalies in international day and night returns. Available at SSRN: https://ssrn.com/abstract=2670841.
Battey, H., Linton, O.B., 2014. Nonparametric estimation of multivariate elliptic densities via finite mixture sieves. J. Multivariate Anal. 123, 43–67.
Berkman, H., Koch, P.D., Tuttle, L., Zhang, Y.J., 2012. Paying attention: overnight returns and the hidden cost of buying at the open. J. Financ. Quant. Anal. 47 (04), 715–741.
Blanc, P., Chicheportiche, R., Bouchaud, J.-P., 2014. The fine structure of volatility feedback II: Overnight and intra-day effects. Physica A 402, 58–75.
Boehmer, E., Fong, K.Y., Wu, J.J., 2015. International Evidence on Algorithmic Trading. EDHEC Business School Working Paper.
Bougerol, P., 1993. Kalman filtering with random coefficients and contractions. SIAM J. Control Optim. 31 (4), 942–959.
Brogaard, J., 2011. High Frequency Trading and Volatility. Working paper.
Carrasco, M., Chen, X., 2002. Mixing and moment properties of various GARCH and stochastic volatility models. Econometric Theory 18 (01), 17–39.
Cho, Y.-H., Linton, O.B., Whang, Y.-J., 2007. Are there monday effects in stock returns: A stochastic dominance approach. J. Empir. Financ. 14 (5), 736–755.
Cooper, M.J., Cliff, M.T., Gulen, H., 2008. Return differences between trading and non-trading hours: Like night and day. Available at SSRN: https://ssrn.com/abstract=1004081.
Creal, D., Koopman, S.J., Lucas, A., 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. J. Bus. Econom. Statist. 29, 552–563.
Creal, D., Koopman, S.J., Lucas, A., 2013. Generalized autoregressive score models with applications. J. Appl. Econometrics 28 (5), 777–795.
Douc, R., Moulines, E., Stoffer, D., 2014. Nonlinear Time Series: Theory, Methods and Applications with R Examples. CRC Press.
Engle, R.F., Lee, G., 1999. A long-run and short-run component model of stock return volatility. In: Engle, R.F., White, H. (Eds.), Causality, Cointegration, and Forecasting: A Festschrift in Honour of Clive W.J. Granger. Oxford University Press, pp. 475–497.
Engle, R.F., Rangel, J.G., 2008. The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. Rev. Financ. Stud. 21 (3), 1187–1222.
Fan, Y., Li, Q., 1996. Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64 (4), 865–890.
Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using "realized" volatility. J. Financ. Econ. 67 (3), 473–509.
French, K.R., 1980. Stock returns and the weekend effect. J. Financ. Econ. 8 (1), 55–69.
French, K.R., Roll, R., 1986. Stock return variances: The arrival of information and the reaction of traders. J. Financ. Econ. 17 (1), 5–26.
Giacomini, R., White, H., 2006. Tests of conditional predictive ability. Econometrica 74 (6), 1545–1578.
Glosten, L.R., Jagannathan, R., Runkle, D.E., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. J. Finance 48 (5), 1779–1801.
Hafner, C.M., Linton, O.B., 2010. Efficient estimation of a multivariate multiplicative volatility model. J. Econometrics 159 (1), 55–73.

Han, H., Kristensen, D., 2015. Semiparametric Multiplicative GARCH-X Model: Adopting Economic Variables to Explain Volatility. Working Paper.
Hansen, P.R., Lunde, A., 2005. A realized variance for the whole day based on intermittent high-frequency data. J. Financ. Econ. 3 (4), 525–554.
Härdle, W., Linton, O.B., 1994. Applied nonparametric methods. Handb. Econom. 4, 2295–2339.
Harvey, A.C., 2013. Dynamic Models for Volatility and Heavy Tails: with Applications to Financial and Economic Time Series, Vol. 52. Cambridge University Press.
Harvey, A.C., Chakravarty, T., 2008. Beta-t-(e) Garch. University of Cambridge, Faculty of Economics.
Harvey, A., Lange, R.-J., 2018. Modeling the interactions between volatility and returns using EGARCH-M. J. Time Series Anal. 39 (6), 909–919.
Harvey, A., Luati, A., 2014. Filtering with heavy tails. J. Amer. Statist. Assoc. 109 (507), 1112–1122.
He, C., Silvennoinen, A., Teräsvirta, T., 2008. Parameterizing unconditional skewness in models for financial time series. J. Financ. Econ. 6 (2), 208–230.
He, C., Teräsvirta, T., Malmsten, H., 2002. Moment structure of a family of first-order exponential GARCH models. Econometric Theory 18 (4), 868–885.
Imbens, G., Wooldridge, J., 2007. Difference-in-differences estimation. In: NBER Lecture Notes in Econometrics. In: Lecture Notes, vol. 10.
Jones, M., Linton, O., Nielsen, J., 1995. A simple bias reduction method for density estimation. Biometrika 82 (2), 327–338. http://dx.doi.org/10.1093/biomet/82.2.327.
Kelly, M.A., Clark, S.P., 2011. Returns in trading versus non-trading hours: The difference is day and night. J. Asset Manag. 12 (2), 132–145.
Kristensen, D., 2009. Uniform convergence rates of kernel estimators with heterogeneous dependent data. Econometric Theory 25 (05), 1433–1445.
Li, J., Viktor, T., George, T., 2017. Jump Factor Models in Large Cross-Sections. Duke University Discussion Paper.
Lin, W.-L., Engle, R.F., Ito, T., 1994. Do bulls and bears move across borders? International transmission of stock returns and volatility. Rev. Financ. Stud. 7 (3), 507–538.
Lindskog, F., Mcneil, A., Schmock, U., 2003. Kendall's tau for elliptical distributions. In: Georg, B., Gholamreza, N., Svetlozarm, T.R., Thomas, R., Karl-Heinz, V. (Eds.), Credit Risk. Springer, pp. 149–156.
Linton, O.B., 1993. Adaptive estimation in ARCH models. Econometric Theory 9 (04), 539–569.
Linton, O.B., O'Hara, M., Zigrand, J.-P., 2013. The regulatory challenge of high-frequency markets. In: Easley, D., Lopez de Prado, M., O'Hara, M. (Eds.), High Frequency Trading: New Realities for Trades, Markets, and Regulators. Risk Books, London, pp. 207–230.
Linton, O.B., Wu, J., 2018. A Coupled Component GARCH Model for Intraday and Overnight Volatility. Cambridge Working Papers in Economics no. 1879.
Lockwood, L.J., Linn, S.C., 1990. An examination of stock market return volatility during overnight and intraday periods, 1964–1989. J. Finance 45 (2), 591–601.
Martens, M., 2002. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. J. Futures Mark. 22 (6), 497–518.
Mehdian, S., Perry, M.J., 2001. The reversal of the monday effect: new evidence from US equity markets. J. Bus. Financ. Account. 28 (7–8), 1043–1065.
Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: A new approach. Econometrica 59 (2), 347–370.
Ng, V., Masulis, R.W., 1995. Overnight and daytime stock return dynamics on the london stock exchange. J. Bus. Econom. Statist. 13 (4), 365–378.
Polk, C., Lou, D., Skouras, S., 2019. A tug of war: overnight versus intraday expected returns. J. Financ. Econ. 134 (1), 192–213.
Rangel, J.G., Engle, R.F., 2012. The Factor–Spline–GARCH model for high and low frequency correlations. J. Bus. Econom. Statist. 30 (1), 109–124.
Rogalski, R.J., 1984. New findings regarding day-of-the-week returns over trading and non-trading periods: a note. J. Finance 39 (5), 1603–1614.
Rogers, L.C.G., Satchell, S.E., 1991. Estimating variance from high, low and closing prices. Ann. Appl. Probab. 1, 504–512.
Scholes, M., Williams, J., 1977. Estimating betas from nonsynchronous data. J. Financ. Econ. 5 (3), 309–327.
Severini, T.A., Wong, W.H., 1992. Profile likelihood and conditionally parametric models. Ann. Statist. 20, 1768–1802.
Steeley, J.M., 2001. A note on information seasonality and the disappearance of the weekend effect in the UK stock market. J. Bank. Financ. 25 (10), 1941–1956.
Straumann, D., Mikosch, T., 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. Ann. Statist. 34 (5), 2449–2495.
Sullivan, R., Timmermann, A., White, H., 2001. Dangers of data mining: The case of calendar effects in stock returns. J. Econometrics 105 (1), 249–286.
Tasaki, H., 2009. Convergence rates of approximate sums of Riemann integrals. J. Approx. Theory 161 (2), 477–490.
Thiele, S., 2019. Modeling the conditional distribution of financial returns with asymmetric tails. J. Appl. Econometrics.
Tibshirani, R., 1984. Local Likelihood Estimation (Ph.D. thesis). Stanford University.
Tsiakas, I., 2008. Overnight information and stochastic volatility: A study of European and US stock exchanges. J. Bank. Financ. 32 (2), 251–268.
Vogt, M., 2012. Nonparametric regression for locally stationary time series. Ann. Statist. 40 (5), 2601–2633.
Vogt, M., Linton, O.B., 2014. Nonparametric estimation of a periodic sequence in the presence of a smooth trend. Biometrika 101 (1), 121–140.
Wintenberger, O., 2013. Continuous invertibility and stable QML estimation of the EGARCH (1, 1) model. Scand. J. Stat. 40 (4), 846–867.

Please cite this article as: O. Linton and J. Wu, A coupled component DCS-EGARCH model for intraday and overnight volatility. Journal of Econometrics (2020), https://doi.org/10.1016/j.jeconom.2019.12.015.