

# The distribution of index futures realised volatility under seasonality and microstructure noise

## Abstract

Previous research documents that the distribution of realised volatility appears approximately log-normal. However, formal tests reject normality fairly convincingly, which may indicate intrinsic features in the intraday data series, namely, the presence of seasonal intraday patterns and microstructure noise. Because many models are based on a normality assumption, this must be verified in order to validate the results. We find departures from normality due to the seasonal and noise components of intraday data, such that, after controlling for both features, the volatility estimates follow a log-normal distribution. Our results reveal that failing to account for these market imperfections can have important implications for analyses of volatility transmission and for investment and hedging decisions.

## 1 Introduction

The distributional characteristics of asset returns are pivotal in financial economics for asset pricing, portfolio selection, performance evaluation, and managerial decision-making, among others. Because the second moment structure is probably the most critical feature of the conditional return distribution, it has triggered a growing body of research (see Bollerslev and Ole Mikkelsen, 1999; Bekaert and Wu, 2000; Campbell *et al.*, 2001; and Andersen *et al.*, 2003, among others). The pioneering study of Clark (1973) states that the log-normal distribution is appropriate for daily volatility. Indeed, adjusting the volatility distribution towards a log-normal distribution becomes crucial, because normality is assumed in many models, such as option pricing models (e.g. Scott, 1987; Hull and White, 1987; Heston, 1993). Furthermore, when a statistical model includes a normality assumption, this must be verified in order to validate the results. Thus, we investigate the effects of departures from normality in the volatility distribution.

Traditionally, (G)ARCH models have been popular parametric approaches for modelling financial asset return volatilities, correlations and distributions. However, the increasing

availability of high-frequency data (HFD) has produced an explosive growth in the financial econometrics of volatility dynamics, allowing for the construction of more accurate daily volatility measures: realised volatility (RV).<sup>1</sup> Additionally, as volatility becomes observable, it can be modelled directly, rather than being treated as a latent variable. Therefore, we can model and forecast it using standard time-series techniques (Andersen *et al.* 2001, 2003). In general, under the assumption that RV follows a log-normal distribution, standard linear Gaussian approaches are used. However, if the hypothesis of normality is not supported, inferences derived from these models may be biased.

Previous research documents the RV distribution. Andersen *et al.* (2001c) study the RV of exchange rates, Andersen *et al.* (2001b) examine the RV of individual stocks in the Dow Jones, and Areal and Taylor (2002) analyse the RV of index futures contracts. Although these studies find that the RV distribution appears approximately Gaussian, formal tests reject normality fairly convincingly, which may indicate intrinsic features in the data series. Our main objective is to analyse the extent to which the presence of such biases may affect normality. These biases are widely studied in the literature. For example, they include the presence of seasonal intraday patterns (behavioural characteristics of financial markets) and the presence of a noisy component generated by intrinsic operations (trading characteristics arising from issues such as price discreteness, bid–ask spreads, or non-synchronous trades/quotes).<sup>2</sup> Additionally, we examine whether the results are influenced by the frequency of the data, and analyse the economic effects of such corrections in areas that rely on volatility estimates, such as investment/hedging decisions and spillover analyses.

We depart from previous studies and make several novel contributions. First, we extend the data. As such, the empirical analysis is performed using CAC and DAX index futures contracts over a long time horizon of more than 12 years. The frequency of observations is critical, because it may affect the accuracy of estimates. Thus, we use different frequencies (1, 5, 10, 20, and 30 minutes). As noted below, the early literature focuses on one frequency only.

Second, we remove those features that introduce bias into the variance estimates, namely, intraday seasonality and market microstructure noise. Andersen and Bollerslev (1997a) provide a stylised intraday return model based on the Fourier flexible methodology to uncover intraday seasonality.<sup>3</sup> This enables us to obtain deseasonalised intraday returns and, there-

---

<sup>1</sup>RV, constructed as a sum of squared intraday returns, has theoretical advantages in the construction of interdaily volatility forecast evaluation criteria (Andersen and Bollerslev, 1998a; Andersen *et al.*, 2001b).

<sup>2</sup>The theory of quadratic variation suggests that under certain conditions (basically, frequencies that tend to infinity and markets free of frictions (Merton, 1980; Andersen and Bollerslev, 1998a; Andersen *et al.*, 2001c), RV is an unbiased and efficient estimation of return volatility; however, these circumstances are rarely met in practice.

<sup>3</sup>Intraday volatility patterns induce serial correlations and result in difficult statistical inferences (e.g. standard variance-ratio statistics or comparable tests). Andersen and Bollerslev (1997a) propose the Fourier

fore, the deseasonalised RV. Later, Barndorff-Nielsen and Shephard (2002) derived a model to measure the RV without the microstructure noise caused by the intrinsic operating and trading limitations.<sup>4</sup> In practice, RV is often modelled by ignoring the effects of these features on the results. In this study, we combine these two models to produce a more accurate estimate of actual volatility. First, to account for the presence of intraday patterns in the volatility of financial market returns (see, among others, Wood *et al.* (1985), Harris (1986), and Tse (1999)), we estimate the deseasonalised<sup>5</sup> RV ( $RV_t^s$ ) using the model of Andersen and Bollerslev (1997a). Prior studies that address intraday seasonal patterns using different methods include those of Andersen *et al.* (2001b) and Areal and Taylor (2002). The former remove the serial correlation in high-frequency returns using a de-measured MA(1) filtered return series; the latter follow Taylor and Xu (1997) to determine the optimal weights of intraday returns, from which they calculate the RV. Then, to address the non-negligible component associated with microstructure frictions, we apply the realised kernel estimator (Barndorff-Nielsen and Shephard (2002); Barndorff-Nielsen *et al.*, 2009) and estimate the RV free of market microstructure noise ( $RV_t^k$ ). In contrast, previous studies have relied on certain frequencies to overcome noise frictions. For example, Andersen *et al.* (2001b) and Areal and Taylor (2002) employ five-minute returns to mitigate market microstructure noise problems, whereas Andersen *et al.* (2003) use observations sampled at a frequency of 30 minutes. Finally, we consider both effects, calculating the deseasonalised RV free of microstructure noise ( $RV_t^{sk}$ ) as the closest approximation to the actual volatility.<sup>6</sup>

Third, we study the implications of removing the noise and seasonality for the unconditional distribution of the RV. Given the importance of obtaining log-normal distributions (Clark, 1973; Areal and Taylor, 2002; Andersen *et al.*, 2001b), and because many models are based on normality assumptions, we analyse how volatility estimates evolve to meet a log-normal distribution when considering both features jointly and explicitly. Our results reveal that remarkable reductions in skewness and kurtosis occur in both indices and all frequencies after removing both components. As such, the adjustment to a normal distribution becomes almost perfect, to the extent that the hypothesis of normality cannot be rejected in certain cases.

Fourth, we analyse the effect on the conditional distribution, that is, the autocorrelation

---

flexible form (FFF) regression (Gallant, 1981, 1982) to accommodate this seasonal intraday component, allowing for robust inferences and reliable hypothesis testing procedures in empirically realistic settings (Andersen *et al.* 2001a).

<sup>4</sup>The model of Barndorff-Nielsen and Shephard (2002) neglects intraday periodicities. Their paper states that the model has a repeating intraday component (i.e. diurnal), and that this diurnal effect may not be completely ignorable when the relation between the deterministic periodic component and the noise component is not additive. However, they neglect this deficiency.

<sup>5</sup>Deseasonalised volatility refers to the RV in which the intraday seasonal pattern of volatility has been removed.

<sup>6</sup>Note that  $s$ ,  $k$ , and  $sk$  denote data that are deseasonalised, free of market microstructure noise, and free of both effects, respectively.

function and the asymmetric effect. Series that account for intraday seasonality and noise frictions show lower autocorrelation coefficients that remain high for up to about 100 days. Then, they decay slowly to zero at a hyperbolic rate, suggesting the removal of a deterministic structure and the presence of a common degree of fractionally integrated long-memory process across intraday sampling frequencies. With regards to the asymmetric responses of volatility, we find that positive returns have a smaller effect on future volatility than do negative returns, but this response is more visible if the seasonal and noise components are neglected. In other words, the ‘*true volatility process*’ (no seasonality and noise) has a minor asymmetric response.

Fifth, we extend the analysis to the multivariate case, focusing on the effects of seasonality and microstructure noise on the covariances and correlations. We encounter significant differences in the unconditional distributions of covariances based on raw data, and in data in which the intraday seasonality and market microstructure noise have been removed. Additionally, we explore two features of the conditional distribution: cross-correlation patterns, and the asymmetric behaviour of the correlations. Our results reveal a significant reduction in cross-correlations after accounting for intraday seasonality. In addition, the greater co-movement of international future indices is mainly observed during negative return scenarios; this response is more obvious if the seasonal and noise components are neglected. Furthermore, when we take these features into account, we observe a decrease in the correlations during positive return scenarios, which are not observed when using raw data.

Finally, we analyse the potential economic effects of the observed differences caused by the seasonality and microstructure noise when estimating RV. The results show a significant effect of ignoring these features in investment and hedging decisions, where we find considerable deviations in the optimal weights. These differences are present at every point in time in which an investor makes investment/hedging decisions. The magnitudes of these differences show high values, involving the full wealth of the investor at certain points in time. In addition, the effect on the spillover analysis is substantial, with differences in magnitude and even in the direction of the net transmission of volatility.

The remainder of this paper is organised as follows. Section 2 describes our data and explains the methodology we employed to compute the RV measures. Section 3 presents the empirical results for the univariate volatility distribution. Section 4 extends the results to the multivariate case, and Section 5 analyses the potential economic effects on investment/hedging decisions and spillover analyses. Finally, Section 6 summarises the results and concludes the paper.

## 2 Database and RV measurement

### 2.1 Data

Our empirical data set, obtained from TickData<sup>TM</sup>, comprises high-frequency observations at different frequencies of transaction prices from two index futures contracts (CAC and DAX) for the period 2 January 2003 to 30 September 2015.<sup>7</sup> This yields 3,263 and 3,243 trading days for the CAC and DAX index futures, respectively; see Table 1.

Table 1: **Data set description**

Table 1 presents the number of observations, trading days, number of intervals per day, and range intervals for each frequency and for both indices used in the analysis.

<b>Panel A. CAC</b>					
FREQUENCY	1	5	10	20	30
NUMBER OBSERVATIONS	1,598,870	332,826	166,413	81,575	55,471
TRADING DAYS			3,263		
INTERVALS PER DAY	490	101	50	24	16
RANGE INTERVALS	09:05–17:15	09:05–17:30	09:10–17:30	09:20–17:20	09:30–17:30
<b>Panel B. DAX</b>					
FREQUENCY	1	5	10	20	30
NUMBER OBSERVATIONS	1,592,313	330,786	165,393	81,075	55,131
TRADING DAYS			3,243		
INTERVALS PER DAY	490	101	50	24	16
RANGE INTERVALS	09:05–17:15	09:05–17:30	09:10–17:30	09:20–17:20	09:30–17:30

We rely on the study of Andersen and Bollerslev (1997a) to filter the data and construct the series of intraday returns. At the opening of each trading day, observations incorporate adjustments to the information accumulated overnight and, as a result, display much higher variability. Therefore, we remove the first observation of each trading day to avoid biased results. We also handle missing prices using linear interpolation, which softens the effect of a sharp price change following a trading interruption.

Then, the continuously compounded intraday returns are computed at each interval by taking the logarithms and subtracting the previous value. Thus, the raw intraday returns  $R_{t,n}$  at the  $n$ th interval at day  $t$ , for  $n = 1, 2, \dots, N$  and  $t = 1, 2, \dots, T$ , are computed as follows:

<sup>7</sup>Our data set spans almost 13 years; a similar length is used by Areal and Taylor (2002), who use 13 years of five-minute returns from FTSE100 index futures contracts (from January 1986 to December 1998) to estimate daily volatility.

$$R_{t,n} = 100 \times \log\left(\frac{P_{t,n}}{P_{t,n-1}}\right), \quad (1)$$

where  $P_{t,n}$  represents the future price level on interval  $n$  at day  $t$ .

## 2.2 Construction of realised measures of volatility under seasonality and microstructure noise

Once the raw intraday returns  $R_{t,n}$  have been calculated, the realised variance at day  $t$  ( $RV_t$ ), defined as the sum of all available intraday high-frequency squared returns, is estimated as follows:<sup>8</sup>

$$RV_t = \sum_{n=0}^N R_{t,n}^2. \quad (2)$$

Here, RV ( $RV_t$ ) is an accurate approximation of the integrated latent volatility ( $IV_t$ ) (see Andersen et al., 2001c; Barndorff-Nielsen and Shephard, 2002). However, microstructure effects (noise) and intraday periodicities can introduce a severe bias on the daily volatility estimation. Therefore, it follows that

$$RV_t = IV_t + v_t, \quad (3)$$

where  $v_t = f(c, u_t)$ ,  $c$  is the deterministic periodic component (intraday seasonality), and  $u_t$  is the noise component. As noted by Barndorff-Nielsen and Shephard (2002), when the relationship between  $c$  and  $u_t$  has an additive nature, the seasonal (or diurnal) effect is nonsignificant. However, when this relation is not additive, the diurnal effect may not be completely ignorable. Because this relation is unknown, we empirically analyse whether these components have an effect on the RV distribution. To this end, we also estimate the RV distributions after removing these biases.

To construct the deseasonalised realised variance at day  $t$  ( $RV_t^s$ ), we first check that the intraday pattern is confirmed for both indices and all frequencies. Figure 1 exhibits the mean absolute return for each interval during a trading day and corroborates the presence

---

<sup>8</sup>A detailed explanation about the RV methodology can be found in Andersen *et al.* (2003).

of intraday seasonality.<sup>9</sup> This pronounced pattern is also evident in the autocorrelogram for the absolute intraday returns depicted out of a lag of 10 days in Figure 2.

Consider, for instance, the intraday volatility pattern for a five-minute frequency represented by Figure 1 (CAC 5 MINUTES). A distorted double U-shaped pattern is evident during the trading day. All markets show a decaying pattern in intraday volatility until 14:30 (corresponding to interval 66). At 14:35 (interval 67), the return volatility increases considerably, before decreasing until 15:30 (interval 77). Then, a remarkable increment occurs again, and remains relatively high until 17:30, reaching its maximum peak at 16:05 (interval 84). Note that the last three five-minute returns for the trading day (intervals 99, 100, and 101) also constitute quite unusual intervals. Similar patterns are found by Harju and Hussain (2011) in the major European equity markets (the CAC, FTSE100, SMI, and DAX 30 indices) for the period 1 September 2000 to 31 March 2006. These authors find evidence that US macroeconomic announcements<sup>10</sup> at 14:30 and 16:05 and the NYSE cash market opening time at 15:30 have cross-border effects on European equity volatilities.

In view of this outcome, we apply the FFF, originally proposed by Gallant (1981, 1982), to account for this intraday seasonality exhibited by the data. To capture irregularities in the seasonal pattern, we include time-of-day volatility dummies in the Fourier regression.<sup>11</sup> Thus, after implementing the FFF, we obtain the deseasonalised returns  $R_{t,n}^s$ . These returns are then used to calculate the deseasonalised realised variance at day  $t$  ( $RV_t^s$ ):

$$RV_t^s = \sum_{n=0}^N (R_{t,n}^s)^2. \quad (4)$$

To overcome the microstructure noise problem, we adopt the realised kernel estimator<sup>12</sup> of Barndorff-Nielsen *et al.* (2009). Under this specification, the realised variance at day  $t$ , free of market microstructure noise, is defined as the following weighted sum (by a kernel function) of the intraday returns:

---

<sup>9</sup>A well-known stylised feature of the intraday statistical characteristics of many financial markets is that volatility is higher at the opening and closing of the trading day, and lower in the middle (e.g. see Wood *et al.* (1985), Harris (1986), and Tse (1999)).

<sup>10</sup>For example, the producer price index, retail sales, consumer price index, consumer confidence, and so on.

<sup>11</sup>Dummy variables in which these observations have been removed are incorporated to capture irregularities in the seasonal pattern, except for the last 15 minutes of the trading day, when the frequency used is one minute. Andersen and Bollerslev, 1997a also remove the last 15 minutes returns. Further details on the dummy variables are available upon request. A more detailed explanation of the FFF procedure is given in Appendix A.1.

<sup>12</sup>See the estimation details of the realised kernel estimator in Appendix A.2.

**Figure 1: Mean absolute returns at different frequencies for the CAC and DAX indices**

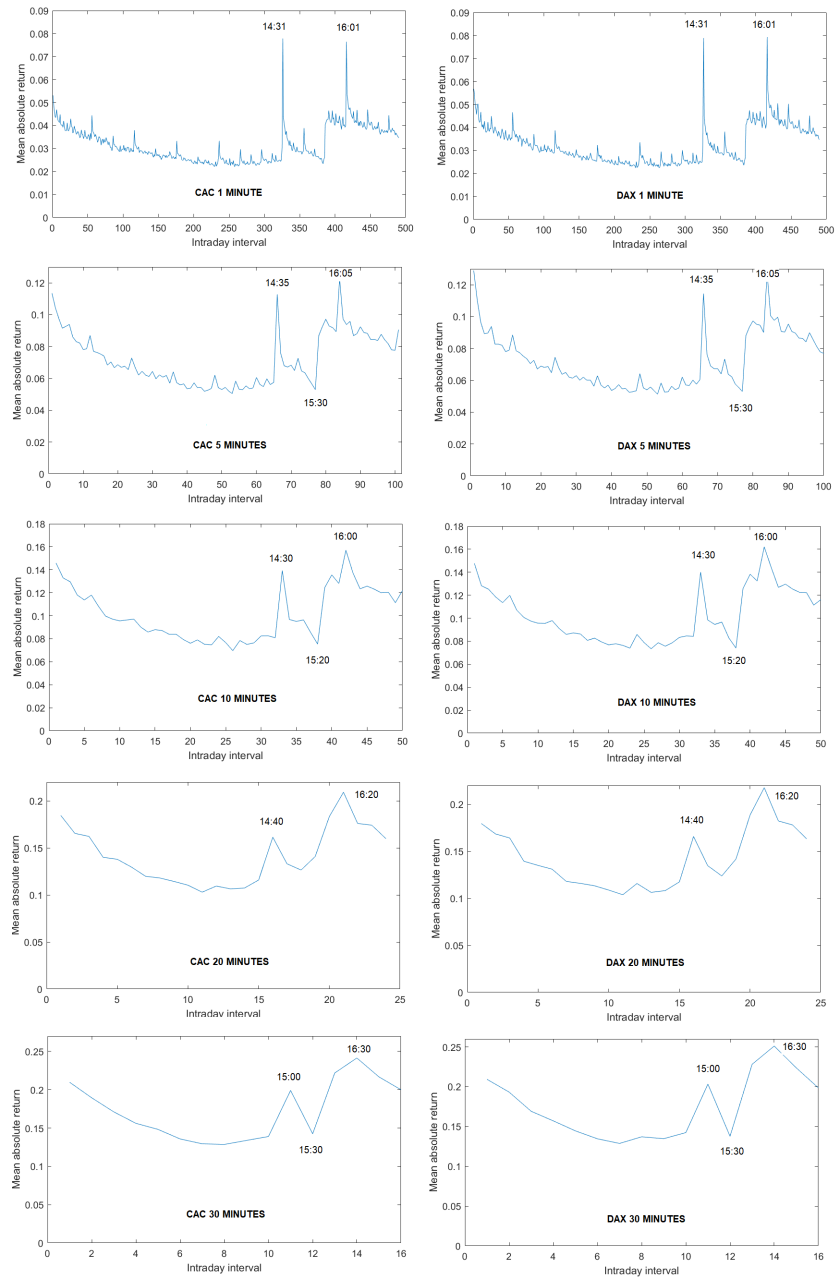


Figure 1 reports the average absolute returns for each interval of the trading day for the CAC (left-hand) and DAX (right-hand) stock index futures using all available observations at different intraday frequencies. The horizontal axis represents the number of intervals in a trading day for each frequency considered. The vertical axis represents the mean absolute returns.



**Figure 2: Autocorrelogram for absolute returns at different frequencies for the CAC and DAX indices**

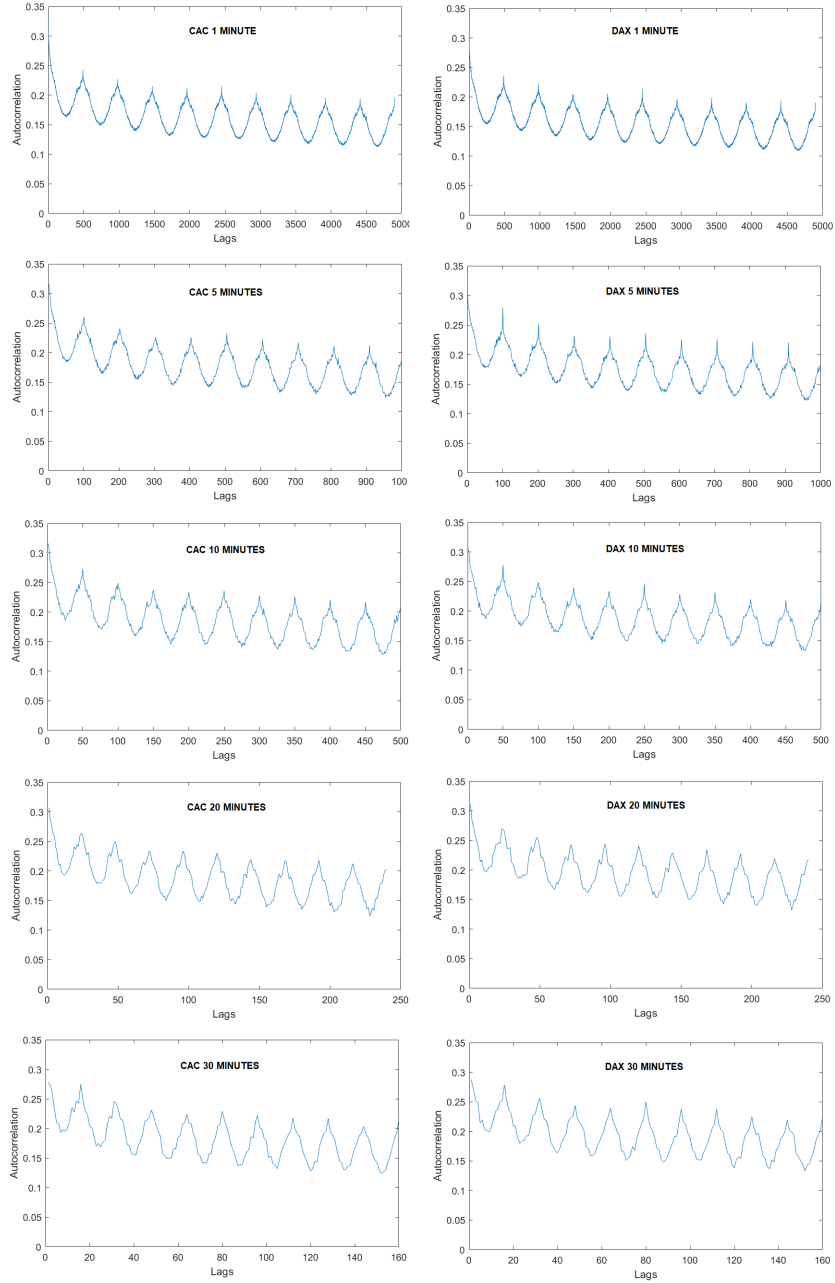


Figure 2 reports the autocorrelogram for the absolute returns for the CAC (left-hand) and DAX (right-hand) stock index futures at different intraday frequencies up to a lag of 10 days.

$$RV_t^k = \sum_{h=-H}^H k\left(\frac{h}{H+1}\right) \sum_{n=|h|+1}^N R_{t,n} R_{t,n-|h|}, \quad (5)$$

where  $k\left(\frac{h}{H+1}\right)$  is the kernel function,  $H$  is the optimal bandwidth, and  $R_{t,n}$  is the raw intraday return.

Finally, to jointly account for intraday seasonality and microstructure noise, we estimate the deseasonalised realised variance free of market microstructure noise ( $RV_t^{sk}$ ). To do so, we use the deseasonalised returns ( $R_{t,n}^s$ ) calculated previously, and then apply the following Barndorff-Nielsen *et al.* (2009) kernel estimator:

$$RV_t^{sk} = \sum_{h=-H}^H k\left(\frac{h}{H+1}\right) \sum_{n=|h|+1}^N R_{t,n}^s R_{t,n-|h|}^s. \quad (6)$$

Thus, for each index and for each frequency, four series of RV (i.e.  $RV_t$ ,  $RV_t^s$ ,  $RV_t^k$ , and  $RV_t^{sk}$ ) are available for analysis.<sup>13</sup>

### 3 Empirical results for univariate volatility distributions

This section examines the effects on the distributional properties, autocorrelation function, and asymmetric responses of the return volatility after controlling for intraday seasonality and noise. For this purpose, we use as input data the four RV series obtained in the previous section.

#### 3.1 Unconditional distribution of volatility

We extend the findings in the extant literature based on HFD to study the distributional characteristics of volatility by jointly considering intraday seasonality and market microstructure noise. To evaluate the extent to which these two factors may affect the distributional characteristics of the RV series, we start by investigating the similarities between the series of realised variances. To this end, we analyse whether the distributions, medians, and variances between the series  $RV_t$  versus  $RV_t^k$ ,  $RV_t$  versus  $RV_t^s$ , and  $RV_t$  versus  $RV_t^{sk}$

---

<sup>13</sup>Consequently, for each index, we have 20 estimations of volatility (five frequencies  $\times$  four RV series).

are equal, where our primary focus is  $RV_t$  versus  $RV_t^{sk}$ . The results are displayed in Table 2.

Table 2: **Equality tests**

Table 2 shows equality tests for the variances, medians, and distributions of the following series for both indices and all frequencies: a)  $RV_t$  versus  $RV_t^k$ ; b)  $RV_t$  versus  $RV_t^s$ ; and c)  $RV_t$  versus  $RV_t^{sk}$ . The associated p-values are shown below each metric, and \*\*\*, \*\*, and \* indicate that the null hypothesis of equality is rejected at the 1%, 5%, and 10% significance levels, respectively.

PANEL A: VARIANCE (BROWN)										
	CAC					DAX				
	1	5	10	20	30	1	5	10	20	30
$RV_t, RV_t^k$	0.48 (0.48)	0.52 (0.46)	0.01 (0.90)	0.01 (0.92)	0.055 (0.455)	0.34 (0.55)	0.73 (0.39)	0.011 (0.91)	0.038 (0.84)	0.71 (0.39)
$RV_t, RV_t^s$	13.92*** (0.00)	1.91 (0.16)	1.17 (0.27)	3.71* (0.05)	2.04 (0.15)	12.49*** (0.00)	10.74*** (0.00)	12.52*** (0.00)	12.71*** (0.00)	8.00*** (0.00)
$RV_t, RV_t^{sk}$	21.29*** (0.00)	5.033** (0.02)	1.98 (0.15)	4.59** (0.03)	7.84*** (0.00)	18.97*** (0.00)	18.72*** (0.00)	13.48*** (0.00)	12.61*** (0.00)	17.52*** (0.00)
PANEL B: MEDIAN (KRUSKAL)										
	CAC					DAX				
	1	5	10	20	30	1	5	10	20	30
$RV_t, RV_t^k$	20.35*** (0.00)	17.10*** (0.00)	24.74*** (0.00)	53.46*** (0.00)	94.29*** (0.00)	25.04*** (0.00)	19.77*** (0.00)	23.83*** (0.00)	51.17*** (0.00)	91.07*** (0.00)
$RV_t, RV_t^s$	10.01*** (0.00)	0.24 (0.61)	1.18 (0.27)	0.99 (0.31)	1.52 (0.21)	1.76 (0.18)	0.23 (0.60)	0.25 (0.61)	0.45 (0.50)	0.29 (0.59)
$RV_t, RV_t^{sk}$	8.34*** (0.00)	25.23*** (0.00)	41.59*** (0.00)	85.11*** (0.00)	170.31*** (0.00)	21.49*** (0.00)	31.21*** (0.00)	39.65*** (0.00)	88.16*** (0.00)	151.51*** (0.00)
PANEL C: DISTRIBUTION (KOLGOMOROV)										
	CAC					DAX				
	1	5	10	20	30	1	5	10	20	30
$RV_t, RV_t^k$	0.06*** (0.00)	0.06*** (0.00)	0.06*** (0.00)	0.09*** (0.00)	0.13*** (0.00)	0.07*** (0.00)	0.06*** (0.00)	0.08*** (0.00)	0.09*** (0.00)	0.12*** (0.00)
$RV_t, RV_t^s$	0.06*** (0.00)	0.015 (0.83)	0.02 (0.49)	0.02 (0.40)	0.02 (0.53)	0.02 (0.15)	0.02 (0.28)	0.02 (0.24)	0.03 (0.17)	0.02 (0.39)
$RV_t, RV_t^{sk}$	0.04*** (0.00)	0.05*** (0.00)	0.08*** (0.00)	0.11*** (0.00)	0.16*** (0.00)	0.06*** (0.00)	0.06*** (0.00)	0.08*** (0.00)	0.11*** (0.00)	0.14*** (0.00)

As shown, the equality of the variances (Panel A in Table 2), medians (Panel B in Table 2), and distributions (Panel C in Table 2) between the raw series ( $RV_t$ ) and those series in which seasonality and market microstructure noise are removed ( $RV_t^{sk}$ ) is rejected for both indices.<sup>14</sup>

We find that removing seasonality mainly affects the second moment of the series (see Panel A of Table 2), whereas addressing noise affects the median and the distribution (see Panel B and C of Table 2). Considering these results, it seems that the choice of the series matters. Thus, we believe different distributions of the daily realised logarithmic standard deviations

<sup>14</sup>With the exception of the equality of the variances for the CAC index and a 10-minute frequency

will be encountered after controlling for both components.<sup>15</sup>

Table 3 summarises the moments for the distribution of the daily realised logarithmic standard deviations ( $lrv_t, lrv_t^s, lrv_t^k$ , and  $lrv_t^{sk}$ ) obtained at frequencies of 1, 5, 10, 20, and 30 minutes for both index futures, where  $lrv_t = \log(\sqrt{RV_t})$ ,  $lrv_t^s = \log(\sqrt{RV_t^s})$ ,  $lrv_t^k = \log(\sqrt{RV_t^k})$ , and  $lrv_t^{sk} = \log(\sqrt{RV_t^{sk}})$ .

Based on our samples of high-frequency returns for the two stock index futures, we find that when we use raw returns, the unconditional distributions of the logarithmic realised standard deviations are close to a normal distribution. These results are consistent with those reported in the literature, which documents that the distribution of logarithmic monthly standard deviations constructed using daily returns is approximately Gaussian (French, *et al.*, 1987; Andersen *et al.*, 2001b; Areal and Taylor, 2002). However, note that the hypothesis of normality is rejected for all frequencies and indices (see the results for the Jarque–Bera test in Table 3). For instance, for a frequency of 30 minutes, the Jarque–Bera statistic is equal to 52.08 and 27.90 for the CAC and DAX stock index futures, respectively.

Notwithstanding, after considering either seasonality or market microstructure noise, the results change considerably. Note the reduction in skewness and kurtosis after removing the intraday periodic component of volatility ( $lrv_t^s$  series), market microstructure noise ( $lrv_t^k$  series), and both features ( $lrv_t^{sk}$  series), leading to a significant improvement in the closeness of fit to the standard normal distribution for both indices and all frequencies. For instance, Table 3 shows that for the CAC index and a frequency of 1 minute, the skewness decreases from 0.394 to 0.160 ( $lrv_t^{sk}$  series), and the kurtosis decreases from 3.180 to 3.059 ( $lrv_t^{sk}$  series); similar results hold for the rest of the data. Furthermore, note that for 10-minute frequencies onwards, the adjustment to a normal distribution is superior. Thus, the hypothesis of normality cannot be rejected for the CAC index when the frequency of observations is 20 or 30 minutes, whereas for the DAX index, it cannot be rejected for frequencies of 10, 20, or 30 minutes (see the values of the Jarque–Bera test in the last column of Table 3: 2.79, 2.38, 0.95, 2.51, and 8.74, respectively). These findings are consistent with the literature that suggests that using data at the highest available frequency to measure volatility is not always the best approach, because the measure might include additional microstructure effects (Meddahi, 2002; Andersen *et al.*, 2003). These results corroborate the improvement towards the normal distribution for all frequencies studied, suggesting the importance of considering both adjustments up to frequencies of 30 minutes. Failing to account for these effects might lead to bias in a model that relies on a normality assumption.

---

<sup>15</sup>The literature shows that the distribution of daily realised logarithmic standard deviations is more similar to a normal distribution. Thus, we focus on this instead of the distribution of the RV. The distributions of the realised variances are highly right-skewed and leptokurtic, and the assumption of normality is clearly rejected in all cases. For the sake of brevity, these results are not reported here, but are available upon request.

Table 3: Distribution of unconditional daily logarithmic standard deviations

Table 3 shows the fourth first moments of the distributions of the daily logarithmic standard deviations ( $lrvt$ ,  $lrvt^{sk}$ ,  $lrvt^s$ , and  $lrvt^k$ ) for the CAC and DAX index futures and the Jarque–Bera test for normality. The rows within each panel display the results for 1-, 5-, 10-, 20-, and 30-minute intraday returns; \*\* represents rejection at the 5% significance level of the null hypothesis that the daily logarithmic standard deviations are normally distributed.

	$lrvt$ (log raw standard deviation)							$lrvt^s$ (log deseasonalised standard deviation)							$lrvt^k$ (log denoised standard deviation)							$lrvt^k$ (log denoised deseasonalised standard deviation)																		
	Mean	Std. Dev.	Skew.	Kurt.	J.Btest	Mean	Std. Dev.	Skew.	Kurt.	J.Btest	Mean	Std. Dev.	Skew.	Kurt.	J.Btest	Mean	Std. Dev.	Skew.	Kurt.	J.Btest	Mean	Std. Dev.	Skew.	Kurt.	J.Btest															
Frequency	-0.134	0.467	0.394	3.180	89.11**	-0.110	0.412	0.219	3.114	28.06**	-0.188	0.490	0.341	3.153	66.74**	-0.178	0.441	0.160	3.059	14.46**	-0.134	0.467	0.394	3.180	89.11**	-0.110	0.412	0.219	3.114	28.06**	-0.188	0.490	0.341	3.153	66.74**	-0.178	0.441	0.160	3.059	14.46**
1 min.	-0.150	0.481	0.376	3.240	85.02**	-0.160	0.471	0.305	3.162	54.44**	-0.203	0.504	0.309	3.151	55.21**	-0.216	0.492	0.263	3.077	38.69**	-0.150	0.481	0.376	3.240	85.02**	-0.160	0.471	0.305	3.162	54.44**	-0.203	0.504	0.309	3.151	55.21**	-0.216	0.492	0.263	3.077	38.69**
5 min.	-0.175	0.487	0.362	3.261	80.58**	-0.191	0.482	0.326	3.188	62.82**	-0.240	0.531	0.258	3.116	38.15**	-0.261	0.525	0.226	3.039	28.12**	-0.175	0.487	0.362	3.261	80.58**	-0.191	0.482	0.326	3.188	62.82**	-0.240	0.531	0.258	3.116	38.15**	-0.261	0.525	0.226	3.039	28.12**
10 min.	-0.231	0.501	0.319	3.186	60.21**	-0.247	0.486	0.256	3.084	36.76**	-0.337	0.518	0.126	3.003	8.66	-0.365	0.569	0.065	2.940	2.79	-0.231	0.501	0.319	3.186	60.21**	-0.247	0.486	0.256	3.084	36.76**	-0.337	0.518	0.126	3.003	8.66	-0.365	0.569	0.065	2.940	2.79
20 min.	-0.247	0.518	0.294	3.188	52.08**	-0.265	0.508	0.264	3.118	39.99**	-0.396	0.622	0.030	2.959	0.73	-0.445	0.616	0.034	2.886	2.38	-0.247	0.518	0.294	3.188	52.08**	-0.265	0.508	0.264	3.118	39.99**	-0.396	0.622	0.030	2.959	0.73	-0.445	0.616	0.034	2.886	2.38
30 min.																																								
Frequency	-0.105	0.477	0.605	3.513	233.91**	-0.102	0.458	0.438	3.28	115.22**	-0.171	0.517	0.381	3.351	95.16**	-0.177	0.477	0.207	3.137	25.84**	-0.105	0.477	0.605	3.513	233.91**	-0.102	0.458	0.438	3.28	115.22**	-0.171	0.517	0.381	3.351	95.16**	-0.177	0.477	0.207	3.137	25.84**
1 min.	-0.129	0.498	0.508	3.481	171.24**	-0.145	0.467	0.355	3.257	77.13**	-0.193	0.536	0.273	3.242	48.29**	-0.215	0.507	0.137	3.038	10.45**	-0.129	0.498	0.508	3.481	171.24**	-0.145	0.467	0.355	3.257	77.13**	-0.193	0.536	0.273	3.242	48.29**	-0.215	0.507	0.137	3.038	10.45**
5 min.	-0.161	0.504	0.398	3.383	105.50**	-0.176	0.471	0.243	3.160	35.40**	-0.233	0.365	0.157	3.081	14.21**	-0.258	0.535	0.035	2.953	0.95	-0.161	0.504	0.398	3.383	105.50**	-0.176	0.471	0.243	3.160	35.40**	-0.233	0.365	0.157	3.081	14.21**	-0.258	0.535	0.035	2.953	0.95
10 min.	-0.219	0.526	0.254	3.179	39.43**	-0.237	0.491	0.143	3.054	11.51**	-0.330	0.615	0.033	3.043	0.86	-0.367	0.590	-0.057	2.927	2.51	-0.219	0.526	0.254	3.179	39.43**	-0.237	0.491	0.143	3.054	11.51**	-0.330	0.615	0.033	3.043	0.86	-0.367	0.590	-0.057	2.927	2.51
20 min.	-0.237	0.547	0.205	3.194	27.90**	-0.251	0.520	0.103	3.084	6.76	-0.395	0.656	-0.040	2.957	1.13	-0.442	0.6349	-0.080	2.803	8.74	-0.237	0.547	0.205	3.194	27.90**	-0.251	0.520	0.103	3.084	6.76	-0.395	0.656	-0.040	2.957	1.13	-0.442	0.6349	-0.080	2.803	8.74
30 min.																																								

Table 4: **Effect of microstructure noise and intraday seasonality on RV estimates**

This table compares the mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE) between the estimates of the raw RV and the estimates of the denoised RV ( $RV^k$ ), deseasonalised RV ( $RV^s$ ), and denoised and deseasonalised RV ( $RV^{sk}$ ). These statistics are shown for the CAC index futures (Panel A) and DAX index futures (Panel B) at frequencies of 1, 5, 10, 20, and 30 minutes (displayed in rows).

PANEL A: CAC									
	$RV^k$			$RV^s$			$RV^{sk}$		
	MAE	MSE	RMSE	MAE	MSE	RMSE	MAE	MSE	RMSE
1 minute	0.2872	0.8456	0.9196	0.2275	0.2404	0.4903	0.3799	1.1111	1.0541
5 minutes	0.1646	0.2067	0.4546	0.2574	0.3287	0.5733	0.3196	0.6140	0.7836
10 minutes	0.1541	0.1644	0.4054	0.3062	0.4101	0.6404	0.3531	0.6613	0.8132
20 minutes	0.1706	0.2688	0.5184	0.3441	0.6555	0.8096	0.4007	1.0412	1.0204
30 minutes	0.1733	0.2283	0.4778	0.4177	1.1626	1.0783	0.4668	1.4647	1.2102
PANEL B: DAX									
	$RV^k$			$RV^s$			$RV^{sk}$		
	MAE	MSE	RMSE	MAE	MSE	RMSE	MAE	MSE	RMSE
1 minute	0.2920	1.0208	1.0103	0.2522	0.2950	0.5432	0.4161	1.3655	1.1685
5 minutes	0.2627	0.9416	0.9704	0.3124	0.5894	0.7677	0.4459	1.9358	1.3913
10 minutes	0.2603	0.9670	0.9834	0.3501	0.7771	0.8815	0.4586	2.0646	1.4369
20 minutes	0.2408	0.5524	0.7432	0.3744	0.6856	0.8280	0.4651	1.2094	1.0997
30 minutes	0.2500	0.5442	0.7377	0.4346	1.1518	1.0732	0.5392	1.7991	1.3413

Finally, in Table 4, we display the mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE), taking as reference the raw RV ( $RV$ ), for each adjustment made in the estimate: noise ( $RV^k$ ), seasonality, ( $RV^s$ ), and both effects ( $RV^{sk}$ ). We analyse how these differences evolve as the data frequency changes for both the CAC index futures (Panel A) and the DAX index futures (Panel B). The results show that microstructure noise is mostly present at the 1-minute frequency, but the effect on the estimate of RV becomes less important as the frequency decreases. However, we still obtain adjustments due to market microstructure noise for the 30-minute frequency (e.g. MSE of 0.2283 in the DAX index, and MSE of 0.5442 in the CAC index). Thus, although the effect of noise at this frequency is lower, it is still advisable to adjust for this bias. The effect of seasonality shows the opposite relationship, where we observe a higher effect at lower frequencies (e.g. MSE of 1.1622 and 1.1518 for the CAC, and DAX index at the 30-minute frequency), and the effect at the 1-minute frequency is much lower (e.g. MSE of 0.2402 and 0.2950 for the CAC and DAX index, respectively). Finding an optimal frequency is not trivial, owing to the trade-off between the opposite effects (Bandi and Russell, 2006, 2008). These results reinforce the importance of considering market microstructure and seasonality when using HFD (at any frequency), and show the potential deviations in estimates of the

RV if we ignore them.

### 3.2 Conditional distribution of volatility

Volatility clustering is one of the most common stylised facts in financial time series. A quantitative way to view this clustering property is to use an autocorrelogram. To this end, we plot the 300-day autocorrelation function for the series  $lrv_t$  and  $lrv_t^{sk}$  for a 30-minute frequency<sup>16</sup> (see Figure 3).

Note that the intraday periodicities in the return volatility in the  $lrv_t$  series with autocorrelations are repeatedly above the corresponding values for the autocorrelation function of the  $lrv_t^{sk}$  series. This is clearly appreciated in Figure 3, where the autocorrelation coefficient of the daily realised logarithmic standard deviations ( $lrv_t$ ) for one lag is approximately 0.7, but is nearly 0.5 for the  $lrv_t^{sk}$  series. In addition, there are significant differences among the autocorrelation coefficients of the series  $lrv_t$  and  $lrv_t^{sk}$  until approximately lag 250. Furthermore, note that after removing the noise and intraday seasonality, the autocorrelation function remains high up to about 100 days, before decaying slowly to zero. This suggests that a deterministic structure and long-memory volatility dependencies have been removed from the returns.<sup>17</sup> Neglecting this deterministic structure could lead to spurious inferences about the dynamics of the return volatility (Andersen and Bollerslev, 1997a).

To test the suggested long-memory effect, we implement fractional integrated ( $I(d)$ ) testing procedures.<sup>18</sup> Table 5 shows the estimates of  $d$  for the raw against the denoised and deseasonalised absolute returns series. As expected, the estimated coefficients are statistically significant, remarkably stable across the high-frequency returns, and similar to the estimates reported in the extant literature with longer periods of daily data (Table 5 shows that the degree of fractional integration  $d$  ranges from 0.2148 to 0.3546). Other studies, such as those of Andersen and Bollerslev (1997b,1998b,1998c) and Andersen *et al.* (2000), reach the same conclusion. Thus, the commonality across intraday sampling frequencies reveals that the long-memory component of volatility is an intrinsic feature of the return-generating process, with a similar degree of fractional integration across different measures of volatilities ( $RV_t$ ,  $RV_t^{sk}$ ) and frequencies (Andersen *et al.*, 2000).

Moreover, previous studies suggest that good and bad news have different predictive abilities for future volatility. In general, negative returns have a greater effect on volatility than

---

<sup>16</sup>The pattern is similar for other frequencies.

<sup>17</sup>All frequencies follow the same pattern. The results are not reported here, but are available upon request.

<sup>18</sup>The results thus far reveal the importance of removing intraday seasonality and noise. Thus, the following discussion focuses on the measure  $RV_t^{sk}$ .

Figure 3: Autocorrelations of the daily logarithmic standard deviations

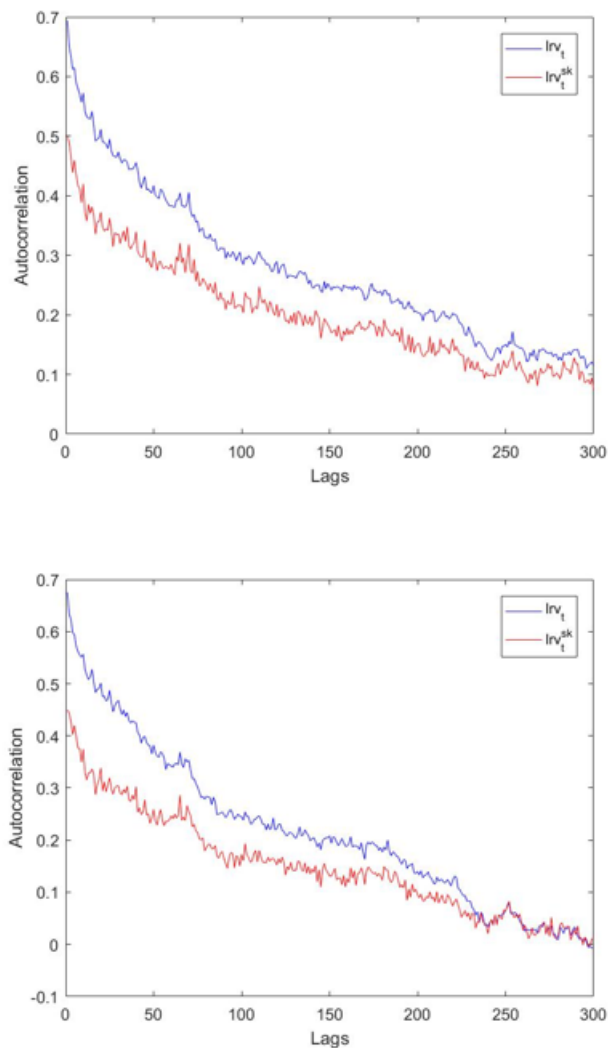


Figure 3 shows the 300-day autocorrelation function for the series of daily logarithmic standard deviations  $lrv_t$  and  $lrv_t^{sk}$  for the CAC (top figure) and DAX (bottom figure) index futures for a 30-minute frequency. The horizontal axis represents the lag length, expressed in days, and the vertical axis denotes the magnitude of the autocorrelation coefficient.



Table 5: **Persistence measures for absolute intraday returns**

Table 5 displays estimates for the log-periodogram estimator of Geweke and Porter-Hudak (1983) ( $d^{GK}$ ) and Robinson (1995) ( $d^{AP}$ ) for the fractional difference parameter  $d$  (standard errors in parentheses) using absolute intraday returns. Rows represent estimates at different data frequencies, and columns are divided into two panels (the CAC and DAX indices, respectively). Each panel displays information for raw and denoised deseasonalised absolute returns; \*\*\*, \*\*, and \* indicate that the null hypothesis of  $d = 0$  (no fractional integration) is rejected at the 1%, 5%, and 10% significance level, respectively.

Frequency	Panel A.- CAC				Panel B.- DAX			
	Raw		Denoised deseasonalised		Raw		Denoised deseasonalised	
	$d_{GK}$	$d_{AP}$	$d_{GK}$	$d_{AP}$	$d_{GK}$	$d_{AP}$	$d_{GK}$	$d_{AP}$
1 minute	0.2812*** (0.0185)	0.3546*** (0.0158)	0.3053*** (0.0192)	0.3400*** (0.0187)	0.2568*** (0.0277)	0.3361*** (0.0141)	0.2887** (0.0180)	0.3234** (0.0158)
5 minutes	0.2754*** (0.0277)	0.3321*** (0.0209)	0.2762*** (0.0259)	0.3015*** (0.0210)	0.2550*** (0.0293)	0.3205*** (0.0209)	0.2787*** (0.0287)	0.2945*** (0.0210)
10 minutes	0.2481*** (0.0325)	0.3078*** (0.0249)	0.2707*** (0.0342)	0.2934*** (0.0252)	0.2370*** (0.0351)	0.2979*** (0.0249)	0.2555*** (0.0252)	0.2760*** (0.0249)
20 minutes	0.2415*** (0.0390)	0.2807*** (0.0299)	0.2702*** (0.0394)	0.2834*** (0.0306)	0.2374*** (0.0395)	0.2760*** (0.0299)	0.2603*** (0.0345)	0.2672*** (0.0306)
30 minutes	0.2410*** (0.0431)	0.2537*** (0.0331)	0.2383*** (0.0456)	0.2499*** (0.0342)	0.2453*** (0.0449)	0.2596*** (0.0331)	0.2148*** (0.0465)	0.2331*** (0.0343)

positive returns of the same magnitude do. This phenomenon, widely known as the leverage effect (Black, 1976; Christie, 1982; Campbell and Hentschel, 1992; Long *et al.*, 2014), has important implications for portfolio selection and asset pricing, and should be considered to better estimate volatilities. To this end, we also analyse these asymmetries under microstructure noise and intraday seasonality, and re-examine the underlying empirical evidence relative to the realised measure  $RV_t$  versus  $RV_t^{sk}$ .

Following Andersen *et al.* (2001b), we use the ordinary least squares method to fit the following regression model:

$$\ln(S_t) = w_t + w_2 r_{t-1} + w_3 r_{t-1} I^- + u_t, \quad (7)$$

where  $\ln(S_t)$  denotes the daily realised logarithmic standard deviations  $lrv_t$  and  $lrv_t^{sk}$ , the indicator  $I^-$  takes the value zero when  $r_{t-1} > 0$ , and one otherwise,  $r_{t-1}$  is the daily return on day  $t - 1$  (calculated using raw returns when the dependent variable is  $lrv_t$ , and using denoised deseasonalised returns when the dependent variable is  $lrv_t^{sk}$ ), and  $u_t$  is an error term (denoted as  $\varepsilon_t$  and  $\varepsilon_t^{sk}$  for the regressions with dependent variables  $lrv_t$  and  $lrv_t^{sk}$ , respectively).

Panel A in Table 6 shows the results of the previous regression for these realised measures

Table 6: **Estimates of the asymmetric behaviour of volatility**

Panel A in Table 6 shows estimates (p-values in parentheses) of equation (7) for the CAC and DAX index futures for a 30-minute frequency, where  $\ln(S_t)$  denotes the daily realised logarithmic standard deviations  $lrv_t$  and  $lrv_t^{sk}$ , the indicator  $I^-$  takes the value zero when  $r_{t-1} > 0$ , and one otherwise,  $r_{t-1}$  is the daily return on day  $t - 1$  (calculated using raw returns when the dependent variable is  $lrv_t$ , and using denoised deseasonalised returns when the dependent variable is  $lrv_t^{sk}$ ), and  $u_t$  is an error term (denoted as  $\varepsilon_t$  and  $\varepsilon_t^{sk}$  for the regressions with dependent variables  $lrv_t$  and  $lrv_t^{sk}$ , respectively). In Panel A, \*\*\* indicates statistical significance at the 1% level. The descriptive statistics for the residuals are displayed in Panel B. Finally, Panel C displays the median, variance, and distribution equality tests for the series  $lrv_t$  versus  $\varepsilon_t$  and  $lrv_t^{sk}$  versus  $\varepsilon_t^{sk}$ ; \*\*\* indicates that the null hypothesis of equality is rejected at the 1% significance level.

<b>Panel A.</b> Estimates for the regression: $\ln(S_t) = w_1 + w_2 r_{t-1} + w_3 r_{t-1} I^- + \varepsilon_t$								
	CAC			DAX				
	$w_1$	$w_2$	$w_3$	$w_1$	$w_2$	$w_3$		
$lrv_t$	-0.495*** (0.00)	0.216*** (0.00)	-0.510*** (0.00)	-0.490*** (0.00)	0.210*** (0.00)	-0.521*** (0.00)		
$lrv_t^{sk}$	-0.641*** (0.00)	2.321*** (0.00)	-5.097*** (0.00)	-0.674*** (0.00)	2.890*** (0.00)	-6.400*** (0.00)		
<b>Panel B.</b> Descriptive statistics for the residuals of the regression: $\ln(S_t) = w_1 + w_2 r_{t-1} + w_3 r_{t-1} I^- + \varepsilon_t$								
	CAC				DAX			
	Mean	Standard Deviation	Skewness	Kurtosis	Mean	Standard Deviation	Skewness	Kurtosis
$\varepsilon_t$	-0.0000	0.4462	0.3477	3.3710	0.0000	0.4718	0.2420	3.4008
$\varepsilon_t^{sk}$	0.0000	0.5830	0.2098	3.0473	-0.0000	0.5927	0.1466	3.0285
<b>Panel C.</b> Equality tests								
	CAC			DAX				
	Variance (Brown)	Median (Kruskal)	Distribution (Kolgomorov)	Variance (Brown)	Median (Kruskal)	Distribution (Kolgomorov)		
$lrv_t$ vs. $\varepsilon_t$	74.81*** (0.00)	409.53*** (0.00)	0.23*** (0.00)	64.32*** (0.00)	341.19*** (0.00)	0.20*** (0.00)		
$lrv_t^{sk}$ vs. $\varepsilon_t^{sk}$	12.23*** (0.00)	759.18*** (0.00)	0.29*** (0.00)	23.39*** (0.00)	697.22*** (0.00)	0.27*** (0.00)		

calculated using a 30-minute frequency.<sup>19</sup> Note that all parameters are significant. Specifically, the parameters  $w_3$  are negative, indicating that negative returns have a greater effect on volatility. Nevertheless, the findings reveal that this asymmetric effect decreases after removing the market microstructure noise and the seasonal component. For instance, the

<sup>19</sup>The results for other frequencies follow a similar pattern, and are available upon request.

coefficient  $w_3$  for the CAC index changes from -0.510 to -5.097.<sup>20</sup> Panel B in Table 6 shows the main statistics indicators relative to the residuals of equation (7). As expected, in view of the results thus far, the skewness and kurtosis in the regression in which the noise and seasonality are considered decrease considerably. Turning to the CAC index, we find that the skewness decreases from 0.3477 to 0.2098, and the kurtosis decreases from 3.3710 to 3.0473. The DAX index follows a similar pattern. Finally, Panel C in Table 6 presents equality tests between the variables  $lrv_t$  and the residuals  $\varepsilon_t$ , and between the variables  $lrv_t^{sk}$  and  $\varepsilon_t^{sk}$ . Note that the equality of distributions does not hold in either case, which means that considering the leverage effect makes a difference. Again, neglecting the noise and seasonal components reveals noticeable differences in the asymmetric effect.

## 4 Empirical results for multivariate volatility distributions

In this section we analyse the differences in the multivariate volatility distribution due to microstructure noise and intraday seasonality. Studies neglecting these characteristics may suffer from model misspecification and provide misleading results. Therefore, understanding the effects of ignoring these two issues when estimating realised measures of volatility in a multivariate setting is important to corroborate previous findings.

### 4.1 Unconditional distribution of volatility for the multivariate case

Although previous studies have analysed the distribution of individual RV (Andersen *et al.*, 2001b; Areal and Taylor, 2002), few works examine the distribution of realised covariances and correlations. Advantages of realised measures of covariances and correlations include that co-movements do not rely on a parametric model and their computation is straightforward. The realised covariance is defined as the cross-product of all available intraday returns in both markets:

$$RCov_{i,j,t} = \sum_{n=0}^N (R_{i,t,n} R_{j,t,n}), \quad (8)$$

where  $RCov_{i,j,t}$  represents the realised covariance between markets  $i$  and  $j$ , and  $R_{i,t,n}, R_{j,t,n}$  are the intraday raw returns at the  $n$ -interval of period  $t$  for markets  $i$  and  $j$ , respectively.

The other covariance specifications considering seasonality and microstructure noise follow

---

<sup>20</sup>Note that  $\exp(-0.510)=0.600, \exp(-5.097)=0.006$ .

similar expressions to that in the univariate case. For the deseasonalised realised covariance, we use the deseasonalised returns ( $R_{t,n}^s$ ) calculated earlier, and then compute the deseasonalised realised covariance as follows:

$$RCov_{i,j,t}^s = \sum_{n=0}^N (R_{i,t,n}^s R_{j,t,n}^s), \quad (9)$$

where  $R_{i,t,n}^s$  and  $R_{j,t,n}^s$  denote the deseasonalised intraday returns for every interval  $n$  of day  $t$  for markets  $i$  and  $j$ , respectively.

To compute the realised covariance free of microstructure noise, we use a variation of the realised kernel in equation (5):

$$RCov_t^k = \sum_{h=-H}^H k\left(\frac{h}{H+1}\right) \sum_{n=|h|+1}^N R_{i,t,n} R_{j,t,n-|h|}, \quad (10)$$

where  $k\left(\frac{h}{H+1}\right)$  is the kernel function,  $H$  is the optimal bandwidth, and  $R_{i,t,n}$ ,  $R_{j,t,n}$  are the intraday raw returns.

Finally, the estimate of the realised covariance considering both effects (seasonality and microstructure noise) is obtained using equation (11) and the deseasonalised returns ( $R_{i,t,n}^s$ ,  $R_{j,t,n}^s$ ) instead of the raw returns ( $R_{i,t,n}$ ,  $R_{j,t,n}$ ):

$$RCov_t^{sk} = \sum_{h=-H}^H k\left(\frac{h}{H+1}\right) \sum_{n=|h|+1}^N R_{i,t,n}^s R_{j,t,n-|h|}^s. \quad (11)$$

Table 7 (Panels A–D) summarises the fourth first moments of the distributions of the daily logarithmic covariances ( $lRCov_t$ ,  $lRCov_t^k$ ,  $lRCov_t^s$ , and  $lRCov_t^{sk}$ ) between the CAC and DAX index futures.<sup>21</sup> Additionally, Table 7 (panel E) shows the equality test for the variances, medians, and distribution between the series  $RCov_t$  versus  $RCov_t^{sk}$ . Similarly to the individual volatilities, we observe significant deviations in the observed levels of skewness and kurtosis among the series (in the mildest case, for denoised deseasonalised realised covariances, the skewness and kurtosis reach values of 0.0532 and 3.0211, respectively; for the raw realised covariances, the skewness and kurtosis reach values of 0.1940 and 3.1225, respectively). The equality tests also reveal significant differences. Indeed, the

---

<sup>21</sup> $lRCov_t = \log(RCov_t)$ ,  $lRCov_t^k = \log(RCov_t^k)$ ,  $lRCov_t^s = \log(RCov_t^s)$ , and  $lRCov_t^{sk} = \log(RCov_t^{sk})$ .

null hypothesis of equality is rejected when comparing  $RCov_t$  versus  $RCov_t^{sk}$ , that is, when removing seasonality and noise (with the exception of a frequency of 5 minutes).<sup>22</sup> Although the results for the distribution of the realised covariance are not conclusive, they extend our previous findings to the multivariate level. Furthermore, they show the perils of ignoring seasonality and microstructure noise when analysing volatility interactions, because different specifications of realised covariances will lead to different conclusions and decisions.

**Table 7: Moments and equality test of the distribution of unconditional daily covariances**

Panels A–D in Table 7 show the fourth first moments of the distributions of the daily logarithmic covariances ( $LRcov_t$ ,  $LRcov_t^k$ ,  $LRcov_t^s$ , and  $LRcov_t^{sk}$ ) between the CAC and DAX index futures. Panel E shows the equality tests for the variances, medians, and distributions between  $RCov_t$  versus  $RCov_t^{sk}$ . The associated p-values are shown below each metric; \*\*\*, \*\*, and \* denote that the null hypothesis of equality is rejected at the 1%, 5%, and 10% significance levels, respectively.

	<b>Panel A.</b> Raw realised covariance( $LRcov_t$ )				<b>Panel B.</b> Denoised realised covariance ( $LRcov_t^k$ )			
	Mean	Std. dev.	Skewness	Kurtosis	Mean	Std. dev.	Skewness	Kurtosis
1 minute	-0.4029	0.9795	0.3463	3.1898	-0.4093	0.9811	0.3416	3.1975
5 minutes	-0.4349	0.9954	0.3637	3.2425	-0.4629	0.9989	0.3551	3.2594
10 minutes	-0.4742	1.0116	0.3340	3.1831	-0.5232	1.0167	0.3160	3.1868
20 minutes	-0.5790	1.0515	0.2461	3.1225	-0.6771	1.0732	0.2078	3.1100
30 minutes	-0.6170	1.0973	0.1940	3.1572	-0.7711	1.1424	0.1067	3.1220
	<b>Panel C.</b> Deseasonalised realised covariance( $LRcov_t^s$ )				<b>Panel D.</b> Denoised deseasonalised realised covariance( $LRcov_t^{sk}$ )			
	Mean	Std. dev.	Skewness	Kurtosis	Mean	Std. dev.	Skewness	Kurtosis
1 minute	-0.2848	0.8592	0.0788	3.3479	-0.2872	0.8704	0.0532	3.2439
5 minutes	-0.4381	0.9863	0.2637	3.1329	-0.4685	0.9882	0.2396	3.0840
10 minutes	-0.4930	1.0196	0.2782	3.1007	-0.5426	0.9752	0.1823	3.0435
20 minutes	-0.6075	1.0133	0.1683	3.0468	-0.7143	1.0658	0.1851	3.0211
30 minutes	-0.6446	1.0634	0.1338	3.0643	-0.8064	1.1276	0.0995	3.0389
	<b>Panel E.</b> Equality tests $RCov_t$ vs. $RCov_t^{sk}$							
		1 minute	5 minutes	10 minutes	20 minutes	30 minutes		
Variance (Brown)		7.62***	2.18	0.03*	5.37**	8.97***		
		(0.00)	(0.14)	(0.07)	(0.02)	(0.00)		
Median (Kruskal)		40.00***	0.80	4.37**	23.59**	40.87***		
		(0.00)	(0.37)	(0.03)	(0.02)	(0.00)		
Distribution (Kolmogorov)		0.09***	0.01	0.03**	0.05***	0.06***		
		(0.00)	(0.83)	(0.07)	(0.02)	(0.00)		

## 4.2 Conditional distribution of volatility for the multivariate case

Similarly to the analysis conducted for the univariate case, we analyse two important features of the conditional distribution for the multivariate case: the serial cross-correlation,

<sup>22</sup>Following the same structure as in Table 2, we apply equality tests over the RCov series (the results hold for the log (RCov) series).

and the asymmetric response to positive and negative shocks.

**Figure 4: Cross-correlations of the daily logarithmic standard deviations for the CAC and DAX indices**

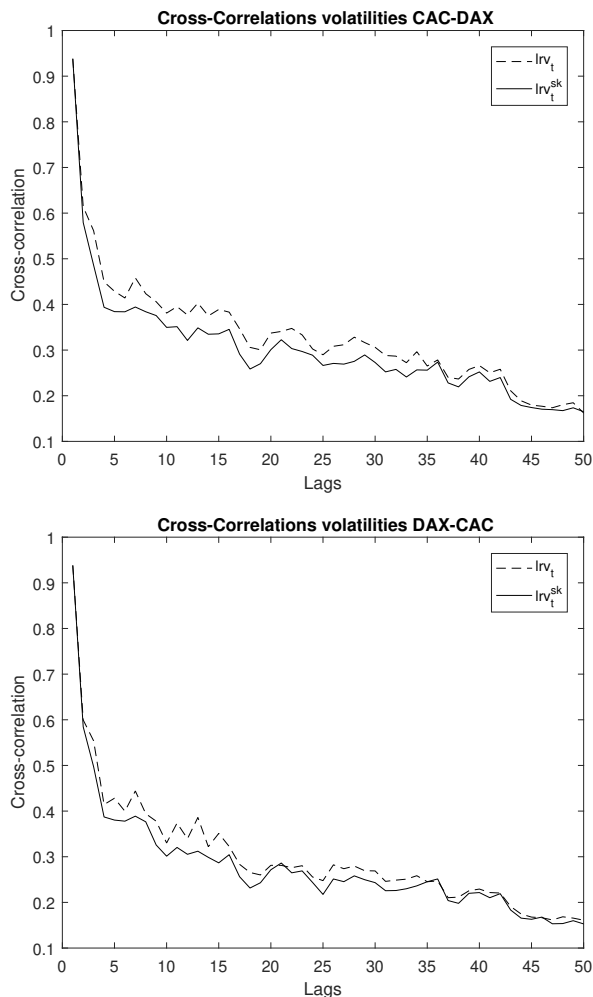


Figure 4 shows the 50-day serial cross-correlation function for the series of daily logarithmic standard deviations using raw data ( $lrv_t$ ) and the series of daily logarithmic standard deviations obtained using denoised and deseasonalised data ( $lrv_t^{sk}$ ) for the CAC and DAX index futures for a 30-minute frequency. The horizontal axis represents the lag length, expressed in days, and the vertical axis shows the magnitude of the cross-correlation coefficient. The top plot displays the cross-correlation of past lags of the DAX index on the contemporaneous volatility of the CAC index. The bottom plot displays the cross-correlation of past lags of the CAC index on the contemporaneous volatility of the DAX index.

Figure 4 shows the 50-day cross-correlation function for the series of daily realised logarithmic standard deviations ( $lrv_t$  and  $lrv_t^{sk}$ ) of the CAC and DAX indices for a 30-minute frequency. This plot shows how the past volatility in one index affects the volatility in the other index. Note that the contemporaneous cross-correlation is very high for both measures of volatility, with values ranging between 0.8 and 0.95. The persistence of the cross-effects is severely reduced, with the decline being more obvious in the series in which we consider the noise and the periodic components ( $lrv_t$  versus  $lrv_t^{sk}$  series). Note that the estimates when using raw data ( $lrv_t$ ) show much higher persistence up to a lag of 35 days (approximately). From this lag onwards, the differences in persistence between the two series are less visible, but still have significant effects on the contemporaneous volatility of the other index. We argue that this higher persistence observed in the cross-correlation for series when neglecting seasonality is due to the repetition of a deterministic structure in the intraday periodic component. This structure may lead to biased results on the transmission of information between markets (Alemany *et al.*, 2019). We could conclude that this transmission exists, when in fact there is a mere repetition of an intraday scheme. Hence, a proper volatility estimator is essential because different choices lead to significant differences.

Previous research suggests that correlations behave differently during good and bad economic scenarios, increasing during crisis periods. However, there is still a debate in the literature about the reasons for this empirical pattern, and even what to call it (some authors refer to it as financial contagion, while others state this is just interdependence; see Forbes and Rigobbon (2002), Corsetti *et al.* (2005) for further discussion). This phenomenon has important implications for portfolio selection and diversification because, during a financial crisis, the possibilities for risk diversification decrease when they are needed the most. Next, we analyse how different estimates of the RV affect this observed phenomenon.

To do so, we start by computing the realised correlations in our framework, as follows:

$$RCorr_{i,j,t}^{k,s,sk} = \frac{RCov_{i,j,t}^{k,s,sk}}{\sqrt{RV_{i,t}^{k,s,sk}} \sqrt{RV_{j,t}^{k,s,sk}}}, \quad (12)$$

where  $RCorr_{i,j,t}^{k,s,sk}$  is the correlation between indices  $i$  and  $j$  at day  $t$ , for  $t = 1, 2, \dots, T$ . For different specifications that consider noise ( $k$ ), intraday seasonalities ( $s$ ), and both effects ( $sk$ ), we have that  $RV_{i,t}^{k,s,sk}$  and  $RV_{j,t}^{k,s,sk}$  are the RVs (obtained from equations 2, 4, 5, and 6), and  $RCov_{i,j,t}^{k,s,sk}$  is defined as the realised covariance (obtained from equations 8 to 11).

Then, we regress the following two models using data from the CAC and DAX indices at a

30-minute frequency:<sup>23</sup>

$$RCorr_{CAC-DAX,t}^{k,s,sk} = w_0 + w_1 I^{++} + w_2 I^{--} + \varepsilon_t, \quad (13)$$

where the variable  $RCorr_{CAC-DAX,t}^{k,s,sk}$  stands for the different correlations at day  $t$  between the CAC and DAX indices, with  $k, s, sk$  representing the adjustment made for noise, seasonality, and both effects, respectively. Furthermore,  $I^{(++)}$  is a dummy variable that takes the value one when the return of the CAC and the return of the DAX are both positive, and zero otherwise, and  $I^{(--)}$  is a dummy variable that takes the value one when the return of the CAC and the return of the DAX are both negative, and zero otherwise. Thus, in this regression, we can see the increase in correlation during bull and bear markets compared with the average correlation.

In addition, to link the estimates of the correlations with the magnitudes of the past returns, we regress the following model:

$$RCorr_{CAC-DAX,t}^{k,s,sk} = w_0 + w_1 r_{t-1} I^{++} + w_2 r_{t-1} I^{--} + \varepsilon_t, \quad (14)$$

where  $r_{t-1}$  is the daily return on day  $t - 1$ , calculated using the intraday returns corresponding to each correlation.<sup>24</sup>

The results for these regressions are displayed in Table 8. Panel A shows that the average estimate of the correlation increases when the indices move downwards together (the coefficients  $w_2$  are positive and significant). This result is robust for all RV estimations, corroborating that financial markets tend to co-move during a crisis. However, the differences in this asymmetric effect of the correlation are more obvious in panel B of Table 8, where we include the magnitude of past returns. Here, we observe differences in the estimates, regardless of whether we consider the intraday seasonality. When we ignore the seasonality, the correlation is positively affected by both positive and negative past returns, where negative past returns lead to a bigger increase in the correlation. For instance, the  $w_1$  coefficient is equal to 0.9951, whereas the  $w_2$  coefficient is equal to -2.0710 when we consider  $RCorr_{CAC-DAX,t}$  as the dependent variable and  $r_{CAC,t-1}$  as the independent variable in the regression (14). However, the results after considering intraday seasonality draw a different picture. With these volatility estimates, positive past returns decrease the correlation among the indices (the  $w_1$  coefficients are negative: -0.0407 and -0.0749 for  $r_{CAC,t-1}$ ;

---

<sup>23</sup>The rest of frequencies lead to similar results. The results are available upon request.

<sup>24</sup>For instance, when the dependent variable is  $RCorr_{CAC-DAX,t}^k$ , we employ denoised returns in equation (22).



Table 8: **Estimates of the asymmetric behaviour of correlation**

This table shows estimates (standard errors in parentheses) of equations (15) (in panel A) and (11) (panel B) for the CAC and DAX indices using data at a 30-minute frequency. The dummy  $I^{(++)}$  takes the value one when the return of the CAC and the return of the DAX are both positive, and zero otherwise. The dummy  $I^{(--)}$  takes the value one when the return of the CAC and the return of the DAX are both negative, and zero otherwise. \*\*\*, \*\*, and \* indicate significance at the 1%, 5%, and 10% levels, respectively.

Panel A. Estimates for the regression			
$RCorr_{CAC-DAX,t} = w_0 + w_1 I^{++} + w_2 I^{--} + \varepsilon_t$			
	$w_0$	$w_1$	$w_2$
$RCorr_t$	0.8673*** (0.0044)	0.0071 (0.0050)	0.0220*** (0.0051)
$RCorr_t^k$	0.8762*** (0.0047)	-0.0035 (0.0054)	0.0132*** (0.0054)
$RCorr_t^s$	0.8676*** (0.0043)	-0.0006 (0.0050)	0.0172*** (0.0050)
$RCorr_t^{sk}$	0.8649*** (0.0047)	0.0018 (0.0053)	0.0182*** (0.0054)
Panel B. Estimates for the regression			
$RCorr_{CAC-DAX,t} = w_0 + w_1 r_{t-1} I^{++} + w_2 r_{t-1} I^{--} + \varepsilon_t$			
	$r_{CAC,t-1}$		
	$w_0$	$w_1$	$w_2$
$RCorr_t$	0.8686*** (0.0022)	0.9951*** (0.2688)	-2.0710*** (0.2451)
$RCorr_t^k$	0.8686*** (0.0047)	1.2146*** (0.3064)	-2.2120*** (0.2824)
$RCorr_t^s$	0.8739*** (0.0023)	-0.0407** (0.0197)	-0.0480*** (0.0173)
$RCorr_t^{sk}$	0.8739*** (0.0047)	-0.0749** (0.0306)	-0.0520*** (0.0263)
	$r_{DAX,t-1}$		
	$w_0$	$w_1$	$w_2$
$RCorr_t$	0.8698*** (0.0022)	0.8211*** (0.2631)	-1.8805*** (0.2417)
$RCorr_t^k$	0.8693*** (0.0047)	1.0569*** (0.2978)	-2.0690*** (0.2757)
$RCorr_t^s$	0.8761*** (0.0022)	-0.0629*** (0.0189)	-0.0284* (0.0166)
$RCorr_t^{sk}$	0.8739*** (0.0024)	-0.0726*** (0.0328)	-0.0481* (0.0272)

-0.0629 and -0.0726 for  $r_{DAX,t-1}$ ), while past negative returns increase the levels of correlation (the  $w_2$  coefficients are negative: -0.0480 and -0.0520 for  $r_{CAC,t-1}$ ; -0.0284 and -0.0481 for  $r_{DAX,t-1}$ ). In addition, note that the coefficients of  $w_1$  and  $w_2$  in absolute terms are lower after controlling for noise and seasonality, indicating a lower effect of past returns in

the correlation. In these estimates, although we still see an increase in the co-movement during crisis periods, the role of positive returns on the correlation is different. Obviously, this has an important effect on all decisions that rely on estimates of correlations among financial markets, such as international asset allocation.

## 5 Effects of seasonality and microstructure noise on economic decisions

Conclusions in different fields of economics rely on the features of multivariate volatility distributions, individual standard deviations, and correlations. The decision-making process in areas such as asset allocation, risk management, or volatility transmission is based directly on estimates of volatility. Therefore, poor estimates might result in suboptimal decisions. Thus, exploring the effects of the corrections caused by seasonality and microstructure noise in such applications will provide a measure of their economic impact.

We analyse two of these applications. First, we examine the effects of seasonality and microstructure noise on the weights of a ‘volatility timing’ strategy (Fleming *et al.*, 2001; Wang *et al.*, 2020) between the CAC and DAX indices. Second, we construct a minimum variance portfolio (suitable for an investor who wants to hedge her investment) using the CAC and DAX indices.

The top plot in Figure 5 displays the differences (using  $RV$  against  $RV^{sk}$ ) in the estimated weights of a ‘volatility timing’ strategy between the CAC and DAX index, where we rebalance the portfolio on a monthly basis. The allocation rule in this strategy is the inverse of the estimated individual volatility; thus, a higher share of wealth will be placed on an asset with lower volatility. This allocation strategy has been shown to provide advantages over more sophisticated allocation strategies (Kirby and Ostdiek, 2012) because it does not require optimization, does not require covariance matrix inversion, and generates positive weights. Specifically, the weight invested in asset  $i$  at time  $t$  is computed as

$$\hat{\omega}_{it}^{(h)} = \frac{1/RV_{it}^{(h)}}{\sum_{i=1}^N 1/RV_{it}^{(h)}}, \quad (15)$$

where  $RV_{it}^{(h)}$  represents the estimate of the RV for the raw ( $RV$ ) or denoised and deseasonalised ( $RV^{sk}$ ) cases of asset  $i$  at period  $t$ , and  $N$  is the number of assets (two, in our case).

Figure 5: Differences in allocated weights due to seasonality and microstructure noise

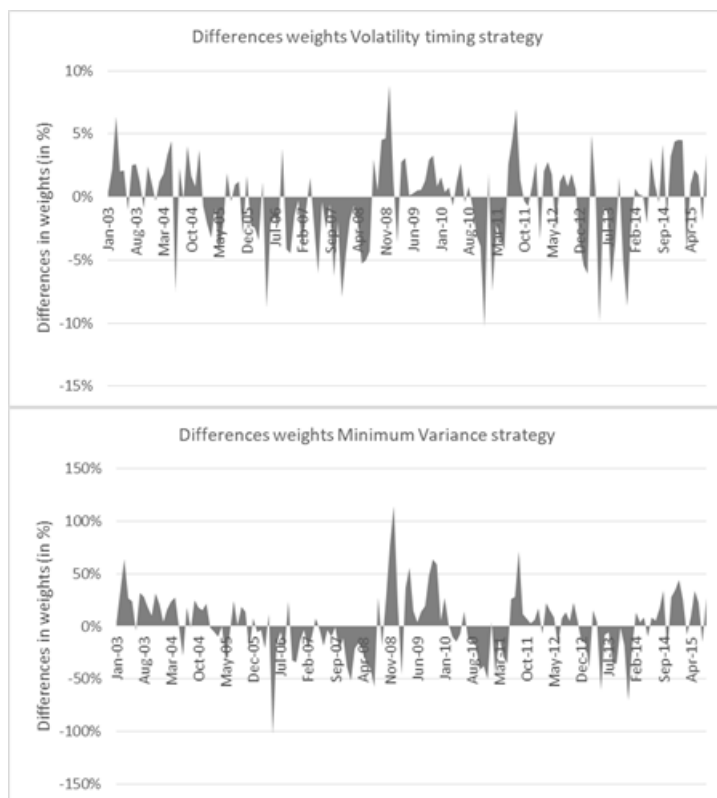


Figure 5 shows the differences in the allocated weights of the CAC40 index futures for a ‘volatility timing’ strategy (top plot) and a minimum-variance strategy (bottom plot). The horizontal axis represents the date (from January 2003 to September 2015). The vertical axis represents the differences in the allocated weights obtained using estimates from raw RVs ( $RV$ ) against those obtained using estimates from denoised deseasonalised RVs ( $RV^{sk}$ ). These differences are expressed as a percentage of total wealth.

Figure 6: Net pairwise volatility spillover between the CAC40 index futures and the DAX index futures.

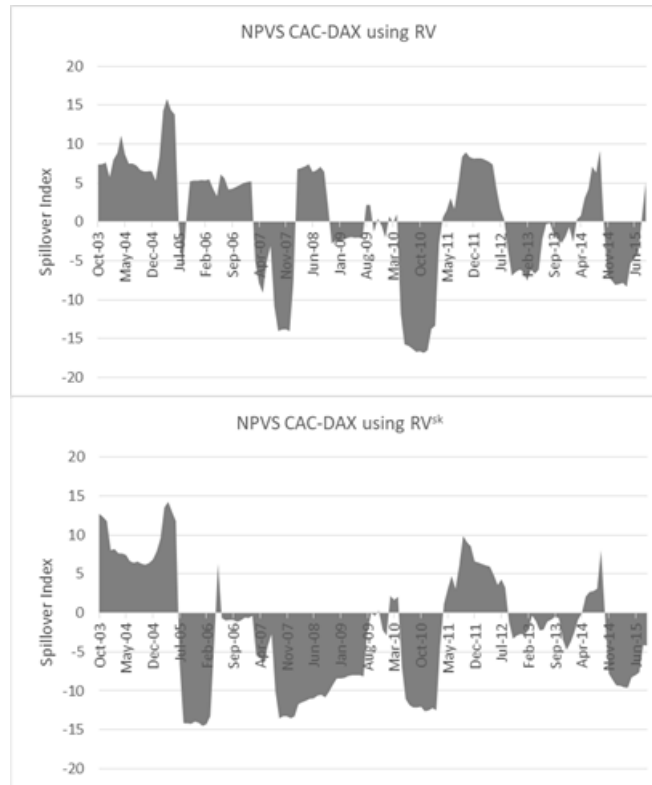


Figure 6 shows the net pairwise volatility spillover between the CAC40 index futures and CAC40 index futures when using raw RVs (top plot) and when using denoised RVs (bottom plot). The horizontal axis represents the date (from October 2003 to September 2015), and the vertical axis represents the net spillover (expressed as a percentage). When the net spillover has a positive value, the CAC index futures market transmits spillovers to the DAX index futures market. In contrast, when the net spillover has a negative value, the CAC index futures market receives spillovers from the DAX index futures market.

We observe differences in the allocated weights for the CAC index<sup>25</sup> across the sample period, depending on the volatility estimate. The magnitude and direction of these deviations do not follow a clear pattern, with periods of overestimates and underestimates in the allocated weights of the CAC index futures. The sizes of these deviations are not marginal, and lead to different allocations, up to approximately 10%, at certain points of time.

The bottom plot in Figure 5 shows the differences (using RV against  $RV^{sk}$ ) in the estimated weights of a hedging strategy using the CAC and DAX indices. In this strategy, the investor is interested in obtaining an allocated portfolio with a minimum variance. The vector of weights at every point of time  $t$  ( $\hat{\omega}_t^{(h)}$ ) can be estimated using the following expression:

$$\hat{\omega}_t^{(h)} = \min_{\omega_t^{(h)}} Var_p = \frac{\Sigma_t^{(h)} \mathbf{1}}{\mathbf{1}' \Sigma_t^{(h)} \mathbf{1}}, \quad (16)$$

where  $\min_{\omega_t^{(h)}} Var_p$  represents an optimization problem that minimises the variance of the allocated portfolio,  $\Sigma_t^{(h)}$  is a  $2 \times 2$  covariance matrix of the CAC and DAX indices in period  $t$  for the raw ( $\Sigma_t$ ) or denoised and deseasonalised ( $\Sigma_t^{(sk)}$ ) cases, and  $\mathbf{1}$  is a vector of ones. The differences in the allocated weights under this strategy are more evident. The average absolute deviation in weights between using the raw estimated RV and the estimate considering noise and seasonality is around 22%. We also observe differences in the weights at certain points higher than 100%, which represent not only a change in the amount invested in the asset, but also a different position in the asset: a short (long) position instead of a long (short) position. In this dynamic hedging strategy, the selection of the RV estimator turns out to be crucial for its proper implementation, because differences lead to very different decisions.

Certainly, asset allocation and risk management are two applications in which the effects of seasonality and microstructure noise on the distribution of the RV are evident. However, there are other fields in which these effects may have a significant effect. In Figure 6, we plot the net pairwise volatility spillover (NPVS) measure of Diebold and Yilmaz (2012) between the CAC and DAX indices. This measure identifies which markets act as net transmitters and net receivers of volatility spillovers, allowing us to disentangle how international markets are interconnected. The figure shows that the magnitude and the direction of the spillovers between the CAC and DAX indices also depends on the choice of the RV estimator.<sup>26</sup> When using the raw RV, we observe that the CAC index acts as a net transmitter of volatility in periods around 2006 and 2008. However, after considering seasonality and microstructure noise the DAX index shows a more dominant role during these periods. In addition, in this

<sup>25</sup>Note that the differences in the weights allocated in the DAX index are the negative values of the weights in the plot

<sup>26</sup>Positive values of the spillover index indicate that the CAC index is acting as a net transmitter of volatility, and negative values indicate that the DAX index is acting as a net transmitter of volatility.

second specification, we observe that the DAX index acts as a net transmitter for most of the sample periods analysed, where the magnitude of the estimated spillover is also higher. These results highlight again the potential economic effects of microstructure noise and intraday seasonality, and the importance of considering these characteristics in economic applications based on HFD.

## 6 Conclusion

This study examines the conditional and unconditional distributions of RV using HFD, while controlling for features that introduce bias in variance estimates, that is, intraday seasonality and market microstructure noise. We evaluate these estimates for the univariate and multivariate case using two index futures contracts (CAC and DAX) and several frequencies of observations (1, 5, 10, 20, and 30 minutes).

The results of the univariate analysis reveal that filtered series ( $RV_t^{sk}$ ) better fit the log-normal distribution for all frequencies (to the extent that the hypothesis of normality cannot be rejected), have lower autocorrelation coefficients, and have a lower asymmetric volatility response. The results from the multivariate case are in line with those of the univariate analysis. Here, we find significant differences in the bivariate distribution of volatility, lower cross-correlation coefficients, and that the asymmetric response of volatility remains visible. However, the effect decreases after accounting for intraday seasonality and noise.

Overall, our findings show that intraday seasonality and market microstructure noise are key factors to meet normality. Because normality is the basis of many models, considering both features becomes crucial. Divergences of the empirical distribution from the normal distribution have important implications for both the conditional and the unconditional distributions of volatility, the degree of autocorrelation, cross-correlations, and asymmetric effects. Failing to account for these features leads to RV estimates that are far from the ideal integrated volatility, resulting in inconsistent conclusions in fields such as information transmission, asset allocation, or risk management.

## References

- [1] Alemany, N., Aragó, V., Salvador, E. (2019). The influence of intraday seasonality on volatility transmission pattern. *Quantitative Finance*, 19(7), 1179-1197.
- [2] Andersen, T. G. and Bollerslev, T. (1997a). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2-3), 115-158.
- [3] Andersen, T. G., Bollerslev, T. (1997b). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Journal of Finance*, 52(3), 975-1005.
- [4] Andersen, T. G. and Bollerslev, T. (1998a). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review*, 39(4), 885-905.
- [5] Andersen, T. G., Bollerslev, T. (1998b). Towards a unified framework for high and low frequency return volatility modeling. *Statistica Neerlandica*, 52(3), 273-302.
- [6] Andersen, T. G., Bollerslev, T. (1998c). Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies. *Journal of Finance*, 53(1), 219-265.
- [7] Andersen, T. G. Bollerslev, T., and . Cai (2000). Intraday and interday volatility in the Japanese stock market. *Journal of International Financial Markets, Institutions Money* 10, 107-130.
- [8] Andersen, T. G. Bollerslev, T., and A. Das (2001a). Variance-ratio Statistics and High-frequency Data: Testing for Changes in Intraday Volatility Patterns. *Journal of Finance* 56( 1), 305-327.
- [9] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001b). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1), 43-76.
- [10] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2001c). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453), 42-55.
- [11] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2), 579-625.
- [12] Areal, N. M., and Taylor, S. J. (2002). The realized volatility of FTSE 100 futures prices. *The Journal of Futures Markets*, 22(7), 627-648.

- [13] Bandi, F. M., and Russell, J. R. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics*, 79(3), 655-692.
- [14] Bandi, F. M., and Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies*, 75(2), 339-369.
- [15] Barndorff-Nielsen, O. E., and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 253-280.
- [16] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realised kernels in practice: trades and quotes. *The Econometrics Journal*, 12(3), C1-C32.
- [17] Black, B. (1976). Studies of Stock Price Volatility Changes. *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics*, 177181
- [18] Bekaert, G., and Wu, G. (2000). Asymmetric Volatility and Risk in Equity Markets. *Review of Financial Studies*, 13(1), 142.
- [19] Bollerslev, T., and Ole Mikkelsen, H. (1999). Long-term equity anticipation securities and stock market volatility dynamics. *Journal of Econometrics*, 92(1), 7599.
- [20] Campbell, J. Y., and Hentschel, L.. (1992). No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns. *Journal of Financial Economics* 31, 281331.
- [21] Campbell, J. Y., Lettau, M., Malkiel, B. G., and Xu, Y. (2001). Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk. *The Journal of Finance*, 56(1), 143.
- [22] Clark, P. K. (1973). A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. Source: *Econometrica*, 41(1), 135155.
- [23] Corsetti, G., Pericoli, M. and Sbracia, M. (2005). Some contagion, some interdependence: More pitfalls in tests of financial contagion. *Journal of International Money and Finance* 24(8), 1177 99.
- [24] Christie, A. C. (1982). The Stochastic Behavior of Common Stock Variances Value, Leverage and Interest Rate Effects. *Journal of Financial Economics* 3, 145166.
- [25] Diebold, F. X., and Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1), 57-66.



- [26] Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, 15(4), 157-168.
- [27] Fleming, J., Kirby, C., and Ostdiek, B., (2001). The economic value of volatility timing. *Journal of Finance*. 56, 329352
- [28] Forbes, K. J. and Rigobon, R. (2002). No Contagion, Only Interdependence: Measuring Stock Market Comovements. *The Journal of Finance* 57(5), 2223-2261.
- [29] French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19(1), 329.
- [30] Gallant, A. R. (1981). On the bias in flexible functional forms and an essentially unbiased form: the Fourier flexible form. *Journal of Econometrics*, 15(2), 211-245.
- [31] Gallant, A. R. (1982). Unbiased determination of production technologies. *Journal of Econometrics*, 20(2), 285-323.
- [32] Geweke, J., and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of time series analysis*, 4(4), 221-238.
- [33] Harju, K., and Hussain, S. M. (2011). Intraday Seasonalities and Macroeconomic News Announcements. *European Financial Management*, 17(2), 367-390.
- [34] Harris, L. (1986). A transaction data study of weekly and intradaily patterns in stock returns. *Journal of Financial Economics*, 16(1), 99-117.
- [35] Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2), 327-343.
- [36] Hull, J., White, A. (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, 42(2), 281-300.
- [37] Kirby, C. Ostdiek, B. (2012). It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis*, 47(2), 437-467.
- [38] Long, L., Tsui, A. K., and Zhang, Z. (2014). Conditional Heteroscedasticity with leverage effect in stock returns: Evidence from the Chinese stock market. *Economic Modelling* 37, 89-102
- [39] Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics*, 17(5), 479-508.

- [40] Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4), 323-361.
- [41] Robinson, P. M. (1995). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 1048-1072.
- [42] Scott, L. O. (1987). Option pricing when the variance changes randomly: Theory, estimation, and an application. *Journal of Financial and Quantitative Analysis*, 22(4), 419-438.
- [43] Taylor, S. J., and Xu, X. (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance*, 4(4), 317-340.
- [44] Tse, Y. (1999). Price discovery and volatility spillovers in the DJIA index and futures markets. *Journal of Futures Markets*, 19(8), 911-930.
- [45] Wang, Y., Liang, F., Tianyi, W. and Huang, Z. (2020). Does measurement error matter in volatility forecasting? Empirical evidence from the Chinese stock market. *Economic Modelling*, 87, 148-157
- [46] Wood, R. A., McInish, T. H., and Ord, J. K. (1985). An Investigation of Transactions Data for NYSE Stocks. *The Journal of Finance*, 40(3), 723-739.

# A Seasonal and market microstructure noise adjustment

## A.1 Modelling intraday seasonality

The HFD literature documents an intraday periodic pattern in volatility that suggests either a U-shaped or a W-shaped form (see, among others, Wood *et al.*, 1985; Harris, 1986; Andersen and Bollerslev, 1997a; and Tse, 1999). Owing to this intraday periodicity, volatility models might lead to spurious inferences about the dynamic of the returns, suggesting the importance of addressing seasonality exhibited by the data. The approach of Gallant (1981,1982), based on the FFF, has proven particularly adept at overcoming this drawback, making it possible to obtain deseasonalised data (Andersen and Bollerslev, 1997a).

We employ the FFF methodology to approximate the intraday periodic component of volatility. This method uses linear polynomial regressions and Fourier methods, which consider sines and cosines. The decomposition for the intraday returns can be expressed as follows:

$$R_{t,n} = E(R_{t,n}) + \frac{\sigma_t S_{t,n} Z_{t,n}}{N^{1/2}}, \quad (17)$$

where  $E(R_{t,n})$  indicates the unconditional mean,  $N$  indicates the number of return intervals per day,  $S_{t,n}$  is the periodic component for the  $n$ th intraday interval,  $\sigma_t$  is the conditional volatility factor for day  $t$ , and  $Z_{t,n}$  is an independent and identically distributed (i.i.d.) error term with mean zero and unit variance that is assumed to be independent of the daily volatility process. By taking the square of both sides and applying logarithmic transformations, we have

$$\log\left((R_{t,n} - E[R_{t,n}])^2\right) = \log\left(\frac{\sigma_t^2 S_{t,n}^2 Z_{t,n}^2}{N}\right). \quad (18)$$

From equation (17), define:

$$X_{t,n} = 2\log[abs(R_{t,n} - E(R_{t,n}))] - \log\sigma_t^2 + \log N \equiv \log S_{t,n}^2 + \log Z_{t,n}^2. \quad (19)$$

The linear FFF regression modelling approach is based on the non-linear regression of  $X_{t,n} \equiv f(\theta; \sigma_t; n) + \varepsilon_{t,n}$ , where the error term  $\varepsilon_{t,n} \equiv \log Z_{t,n}^2 - E(\log Z_{t,n}^2)$  is i.i.d. with mean zero. The non-linear regression  $X_{t,n}$  is proxied by the following parameterised form:

$$f(\theta; \sigma_t; n) = \sum_{j=0}^J \sigma_t^j [\mu_{0j} + \mu_{1j} \frac{n}{N_1} + \mu_{2j} \frac{n^2}{N_2} + \sum_{i=1}^D \lambda_{ij} I_{n=d_i} + \sum_{p=1}^P (\gamma_{pj} \cos \frac{pn2\pi}{N} \delta_{pj} \sin \frac{pn2\pi}{N})], \quad (20)$$

where  $\theta \equiv (\mu_{0j}, \dots, \mu_{0J}, \mu_{1j}, \dots, \mu_{1J}, \mu_{2j}, \dots, \mu_{2J}, \lambda_{ij}, \dots, \lambda_{DJ}, \gamma_{pj}, \dots, \gamma_{PJ}, \delta_{pj}, \dots, \delta_{PJ})$ ,  $N_1$ ,  $N_2$  are normalizing constants defined as  $N_1 = ((N + 1))/2$ , and  $N_2 = ((N + 1)(N + 2))/6$  and  $I_{(n=d_i)}$  are dummy variables that capture irregularities in the seasonal pattern. Additionally, if  $J > 0$ , the whole regression is multiplied by the daily volatility factor  $\sigma_t^j$ .<sup>27</sup> We follow Andersen and Bollerslev (1997) and set  $J = 1$ .

To apply the Fourier approach, a two-step procedure is followed. In the first stage,  $\hat{X}_{t,n}$  is computed from equation (18). Then,  $\hat{X}_{t,n}$  is considered as a dependent variable in the Fourier regression (19), which is estimated using a non-linear regression.<sup>28</sup> Once  $\hat{f}(\hat{\theta}; \sigma_t; n)$  is calculated, the intraday periodic component  $\hat{S}_{t,n}$  for interval  $n$  on day  $t$ , which provides a close approximation to the overall volatility patterns in each market, is retrieved as:<sup>29</sup>

$$\hat{S}_{t,n} = \frac{Te(\hat{f}_{t,n}/2)}{\sum_{t=1}^T \sum_{n=1}^N e(\hat{f}_{t,n}/2)}. \quad (21)$$

Finally, the deseasonalised intraday returns series are defined as follows:

$$\hat{R}_{t,n}^s = \frac{R_{t,n}}{\hat{S}_{t,n}}. \quad (22)$$

## A.2 Microstructure noise consistent estimators

Previous studies have shown that measures of RV can be sensitive to microstructure noise at short intervals. The observed price process at these high frequencies is the result of two components: the true price-generating process and noise. However, we can only observe

<sup>27</sup>We employ the widely used parametric GARCH(1,1) model to capture daily volatility. In most empirical applications, the GARCH(1,1) is enough to reproduce the volatility dynamics of financial series; thus, it is popular with both academics and practitioners (Engle, 2001).

<sup>28</sup>Following Andersen and Bollerslev (1997), the choice of (J,P) is determined by choosing the model that best matches the basic shape of the periodic pattern with the fewest number of parameters. Our selected model for both indices sets  $J = 1$  and  $P = 2$ . Expanding the Fourier beyond this produces nonsignificant estimates for any additional  $\mu_{0j}$ ,  $\mu_{1j}$ ,  $\mu_{2j}$ ,  $\gamma_{Pj}$ , and  $\delta_{Pj}$  coefficients.

<sup>29</sup>For more detail, see Andersen and Bollerslev, 1997a.

both components together. To minimise the effect of microstructure noise in the estimates of the realised variance, studies have proposed techniques that design efficient estimators robust to certain types of frictions.

In our application, we follow Barndorff-Nielsen *et al.* (2009) to compute the realised kernel estimator<sup>30</sup> as the sum of the intraday returns weighted by a kernel function:

$$RV_t^k = \sum_{h=-H}^H k\left(\frac{h}{H+1}\right) \sum_{n=|h|+1}^N R_{t,n} R_{t,n-|h|}, \quad (23)$$

where  $k(x)$  is a kernel weight function. We implement the Parzen kernel applied to the non-flat-top case, which is given by:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3 & 0 \leq x \leq 1/2 \\ 2(1-x)^3 & 1/2 \leq x \leq 1 \\ 0 & x > 1 \end{cases}$$

One of the critical steps when implementing this kernel is the selection of the bandwidth  $H$ . Previous studies have shown that an optimal choice for this bandwidth is:

$$H^* = c^* \varepsilon^{4/5} n^{3/5}, \quad (24)$$

where  $c^* = 3.5134$  for the Parzen kernel,  $n$  is the number of intraday intervals, and the variable  $\varepsilon^2$  is defined as  $\varepsilon^2 = \frac{\omega^2}{\sqrt{T \int_0^T \sigma_u^4 du}}$ .

We estimate  $\varepsilon$  simply by:

$$\hat{\varepsilon}^2 = \frac{\hat{\omega}^2}{\hat{IV}}, \quad (25)$$

where  $\hat{\omega}^2$  is an estimate of  $\omega^2$ , and  $\hat{IV}$  is an estimate of  $IV = \int_0^T \sigma_u^2 du$ . We use this last approximation because  $IV^2 \approx T \int_0^T \sigma_u^4 du$ , and it is simpler to obtain an estimate of  $IV$  than it is for  $\sqrt{T \int_0^T \sigma_u^4 du}$ . In our case, we use the raw RV at the corresponding frequency

---

<sup>30</sup>See Barndorff-Nielsen *et al.* (2009) for a detailed explanation of the procedure.

as the estimate of  $\hat{IV}$ . The estimate of  $\omega^2$  is obtained as follows. By varying the starting point, we obtain  $k$  distinct auxiliary realised variances, defined as  $RV_{aux}^{(1)}, RV_{aux}^{(2)} \dots RV_{aux}^{(k)}$ .

For each of the auxiliary realised variances, we compute  $\hat{\omega}_{(i)}^2 = \frac{RV_{aux}^{(i)}}{2n}$ . Finally, we compute the average of these  $k$  estimates of  $\omega_{(i)}^2$  as our final estimate of  $\hat{\omega}^2$ :

$$\hat{\omega}^2 = \frac{1}{k} \sum_{i=1}^k \hat{\omega}_{(i)}^2. \quad (26)$$

Both techniques (FFF and realised kernel estimator) are combined to obtain the deseasonalised RV free of market microstructure noise. In a first step, the deseasonalised returns ( $R_{t,n}^s$ ) are obtained using equations (17) to (22). Then, a similar procedure to that described for equations (23) to (26) is implemented, except we use the deseasonalised returns ( $R_{t,n}^s$ ) instead of the raw returns ( $R_{t,n}$ ).