17 96

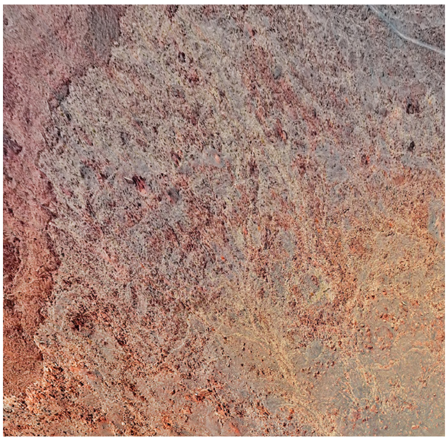
LOMBARD ODIER
INVESTMENT MANAGERS

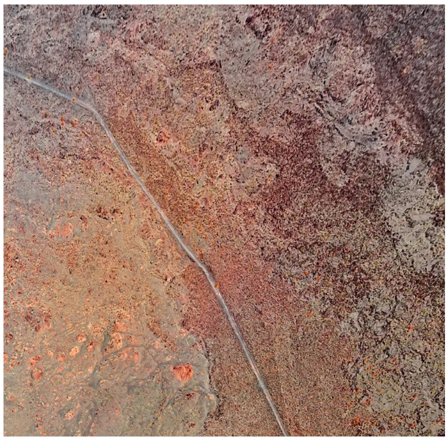
MARS: A primer on range-based volatility estimators

– For professional investor use only –

02/23

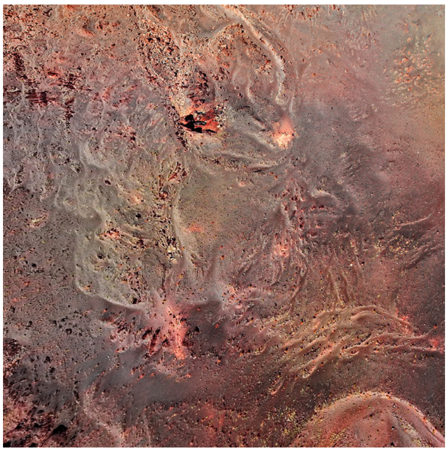
February 2023

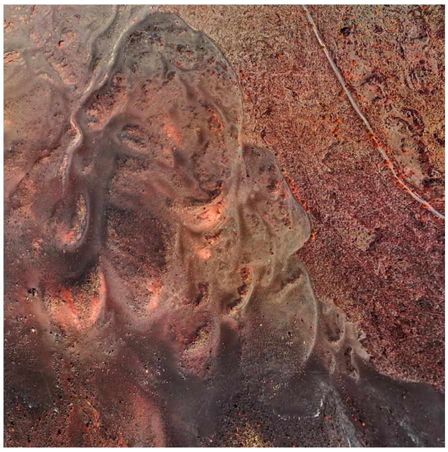




Using range-based estimators for daily volatility can help obtain well-behaved measures.

p.07





MARS, our **M**ulti-**A**sset **R**esearch **S**eries that brings quantitative research back to Earth.

For professional investor use only. Please read important information at the end of this document.

This document has been prepared by:

François Chareyron, ARPM
Portfolio Manager

Julien Royer, PhD
Quantitative Analyst

For further information please visit www.loim.com

Introduction

Modelling the volatility of asset returns has arguably been one of the most prolific subjects in financial literature, from both theoretical statistical and more empirical investment-based points of view. Estimating volatility is crucial to risk management and portfolio construction, especially to risk-based investors. Therefore, since the 1980s multiple studies have been undertaken of this unobservable, time-varying process that underlies financial time series.

Two very influential solutions were developed in parallel. The first was the autoregressive conditional heteroskedastic model of Engle (1982) and its suite of extensions such as the prominent GARCH models. The second was based on the study of derivatives markets and the observation of the implied volatility surface priced by market participants. Both solutions suffer, however, from a misspecification risk: the first solution is a parametric model that may fit inadequately with the underlying data and the second is derived from pricing models.

More recently, the availability of intraday price data of financial assets buttressed the emergence of a third area of research: *realised* volatility where the measure is a function of high-frequency returns. Although such measures feature good theoretical properties and mitigate misspecification risk as they are model-free, they are difficult to use due to limited availability, the high financial and technical costs of intraday data as well the pitfalls of microstructure noise and its impact on such measures.

In this white paper, we investigate how range-based volatility estimators can help proxy intraday realised volatilities. The first section presents the realised volatility framework and explores the advantages and disadvantages of such measures. The second section introduces popular range-based estimators and discusses their benefits in overcoming the shortcomings of high-frequency data-based measures. The third section illustrates the favourable statistical properties of the proposed range-based estimators. The final section concludes and surveys potential future research perspectives.

Realised volatility – the good, the bad and the implementation toll

Realised volatilities are ex-post measures constructed at a low frequency (for example a day, or a month) using the squared returns sampled at a higher frequency over the relevant horizon (for example 5-minute time increments over a day, or daily returns over a month). Let us denote $\sigma^2_{\mathrm{RV},t}$ the daily realised variance defined as

$$\sigma^2_{\mathrm{RV},t} = \sum_{w=1}^{W} r^2_{t-1+w\Delta}$$

where Δ is the sampling step (for example 5-minute increments), W is the number of intervals in a trading day and $r_{t-1+w\Delta}$ denotes the return of the assets over the w-th interval. Although seemingly extremely simple to build, this estimator benefits from an important theoretical result that helped popularise this framework in the academic literature: the estimator is theoretically free of measurement errors as the sampling step Δ tends to zero (Andersen et al. (2001)). Having access to high-frequency returns should thus yield an unbiased estimator of the daily variance process driving the price time series.

Additionally, Andersen et al (2001) and Barndorff-Nielsen and Shephard (2002) show that the realised volatility estimator has favourable empirical properties. In particular, it is highly persistent, making it attractive for volatility forecasting applications, as in Corsi (2009). Furthermore, the estimated volatility process is approximately log-normal and the standardised returns (the observed returns divided by the daily volatility) are empirically Gaussian, which greatly simplifies the derivation of risk measures.

Although appealing, the relative simplicity of the daily realised volatility estimator features three problems:

- First, while theoretically sound, the estimator relies on the availability of well-measured intraday returns. This measure is thus only applicable to highly liquid assets such as major stocks, futures and foreign exchange, because either data do not exist for peripheral assets, or the low liquidity results in a large number of zero returns (see for example Francq and Sucarrat (2023))
- Second, the presence of microstructure noise and jumps in asset prices (e.g. on macro announcements or earnings releases) can make the selection of the intraday sampling frequency as much a science as an art. Jumps and seasonality can have a tremendous impact on the quality of the volatility estimates
- Finally, even if data are available, the sampling of intraday returns from tick data are highly computationally challenging and require sophisticated statistical attention as noted by Barndorff-Nielsen et al. (2009) and De Nard et al. (2022). This translates into a high implementation cost that is impacted by the development of an IT infrastructure able to handle large datasets efficiently, and the high fees attached to the intraday database, especially if both temporal and cross-sectional depth is sought

To circumvent these issues, we focus on range-based volatility estimators that enrich standard close-to-close data with intraday information on the path of the price series. The idea of range-based models is to harvest intraday patterns without having to consider the continuous sample price path, by studying *ranges* formed by the maximum and minimum prices on a trading day (the so-called *high* and *low*). Range-based measures of volatility offer an appealing middle point between close-to-close data used in GARCH models and high-frequency intraday data, as illustrated recently by Baltas and Kosowski (2020).

Range-based volatility measures – an intraday measure with only three additional data points

A first attempt to leverage ranges data was proposed by Parkinson (1980) who estimated the volatility from daily highs and lows only, rather than the standard close-to-close returns. This estimator formula is

$$\sigma^2_{\mathrm{HL},t} = \frac{1}{4 \log 2}\left[\log\left(\frac{\mathrm{H}_t}{\mathrm{L}_t}\right)\right]^2$$

where H_t and L_t denote the high and low of the trading day.[1] Garman and Klass (1980) corrected the squared log-ratio of the range with an additional term accounting for the close-to-open price ratio. In particular, they note that the close-to-close log return is given by

$$\log(1 + r_t) = \log\left(\frac{\mathrm{O}_t}{C_{t-1}}\right) + \log\left(\frac{\mathrm{C}_t}{\mathrm{O}_t}\right)$$

where O_t and C_t denote open and close price of the trading day while C_{t-1} denotes the previous day's closing price. This decomposition allows the authors to derive an estimator where the volatility is decomposed between an overnight jump component and a continuous part component over the trading day. The obtained estimator formula is

$$\sigma^2_{\mathrm{OHLC},t} = \frac{a}{f}\left[\log\left(\frac{\mathrm{O}_t}{C_{t-1}}\right)\right]^2$$

$$+ \frac{1-a}{1-f}\left\{\left[\log\left(\frac{\mathrm{H}_t}{\mathrm{L}_t}\right)\right]^2 - (2\log 2 - 1)\left[\log\left(\frac{\mathrm{C}_t}{\mathrm{O}_t}\right)\right]^2\right\}$$

where a and f are scaling parameters capturing the weights of the overnight jump in the daily volatility.[2] More recently, Yang and Zhang (2000) extended the estimator by allowing for non-zero drift.

1 The $\frac{1}{4 \log 2}$ term is a normalisation coming from the distribution of the high-low range second moment.

2 Focusing on US stocks, Garman and Klass (1980) set $f = 6.5/24$ as the proportion of trading hours over a full day, and calibrate $a = 0.17$.

Some empirical properties of range-based volatility estimators

The success of realised volatility measures computed on intraday data arises from the ability of the estimator to capture important stylised facts about financial returns while being model-free. In particular, it efficiently filters out the noise in the squared returns series, captures the persistence in the volatility, and adequately captures the tails of the returns distribution. Interestingly, range-based estimators empirically share most of the same features, without the implementation costs and issues described earlier.

Figure 1 compares the dispersion of daily volatility measured using close-to-close returns and the dispersion of daily volatility computed using the Garman and Klass estimator, both against the realised volatility computed using 5-minute time increments on the S&P 500 series. The proximity between the range-based and the realised volatilities is striking (with a coefficient of determination, or R^2, around 70%), while the estimator based on close-to-close returns appears much more dispersed. This illustrates the ability of the range-based measures to incorporate intraday patterns efficiently with only two additional data points.

Another interesting feature of realised volatility estimators is their ability to capture the tail behaviour of the underlying returns process. It is well-known that the empirical distribution of financial returns is often far from Gaussian, as it presents heavy tails. This stylised fact is often well mimicked by dynamic volatility models where the return process is given by

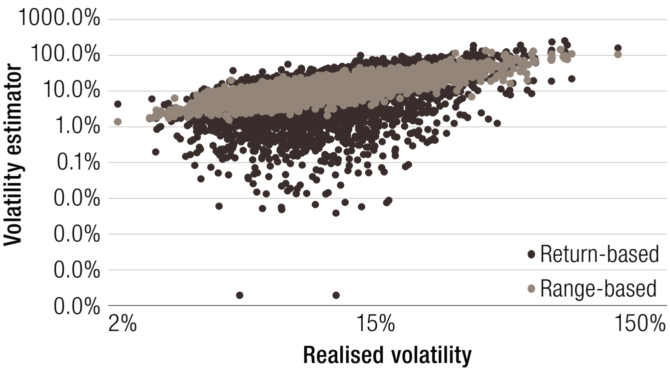
$$r_t = \dot{\sigma}_t \varepsilon_t$$

with $\dot{\sigma}_t$ the unobserved true volatility process and ε_t the innovations that are centered with unit variance. One should note that the distribution of the returns might exhibit a heavy tail, even if the innovations distribution is Gaussian. Actually, the intraday-based realised variance is able to capture the tail event so that, empirically, the estimated ε_t resembles a Gaussian white noise (see e.g. Andersen et al. (2001)). This presents a remarkable advantage when computing conditional risk measures as the statistics are often a function of the innovations cumulated density function – for example, a VaR is obtained by multiplying a quantile of the distribution of ε_t with $\dot{\sigma}_t$.

Alternatively, when considering a GARCH-type model, the estimated residuals are often found to present heavy tails, which renders the derivation of risk measures more involved (see Francq and Zakoïan (2015)). Figure 2 presents the empirical distribution of the innovations when the true volatility process of the S&P 500 returns is estimated with a range-based volatility measure and a realised volatility model computed using 5-minute returns. The range-based volatility residuals are shown to present similar properties to the ones obtained from intraday data.

FIG 1. SCATTER PLOTS OF THE STANDARD AND RANGE-BASED VOLATILITY MEASURES (IN LOG) AGAINST THE REALISED INTRADAY MEASURES (IN LOG)

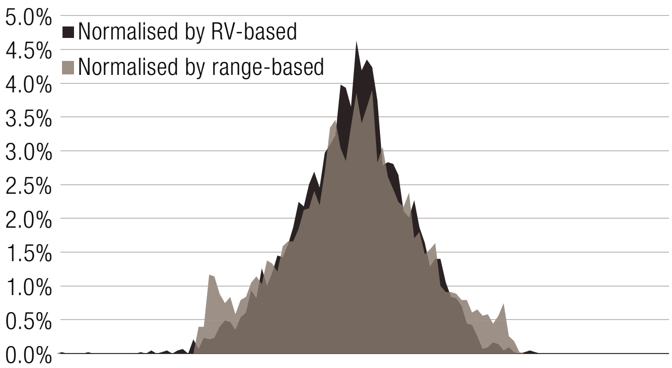
VOLATILITY ESTIMATORS VERSUS REALISED VOLATILITY



Source: LOIM. For illustrative purposes only.

FIG 2. EMPIRICAL DISTRIBUTION OF THE ESTIMATED RESIDUALS USING A REALISED VOLATILITY MEASURE BASED ON INTRADAY DATA AND A RANGE-BASED ESTIMATOR

EMPIRICAL DISTRIBUTION FUNCTION OF NORMALISED RETURNS



Source: LOIM. For illustrative purposes only.

Finally, an important feature of a model is its ability to forecast volatility. This relates to the persistence of the estimator as the higher the persistence, the more informative past values are to the future. As noted by Engle and Patton (2001), a good volatility model should thus present high persistence. Figure 3 presents the auto-correlogram of the estimated volatility processes based on close-to-close returns, range-based estimation, and 5-minute returns for the S&P 500. Again, we see that the range-based estimator behaves in a similar manner to the high-frequency based measure. On the contrary, the measure based on close-to-close returns appears to show moderate persistence, as it decays fast towards zero.

Importantly, it is noteworthy that the persistence of the estimator does not translate into a lack of reactivity – a common critique of volatility measures based on close-to-close returns. A standard benchmark volatility model for close-to-close returns is the exponentially weighted moving average (EWMA) proposed in the famous RiskMetrics framework where the volatility updating formula is

$$\tilde{\sigma}_t = \sqrt{\lambda \tilde{\sigma}_{t-1}^2 + (1-\lambda) r_t^2}$$

with $\lambda = 0.94$. This persistence parameter implies a half-life coefficient of approximately 11.2 days: the mean-reversion of the close-to-close volatility estimator is slow. In comparison, if one defines the EWMA volatility estimators based on range-based volatility and high frequency-based realised volatility as

$$\tilde{\sigma}_{\mathrm{OHLC},t} = \sqrt{\lambda \tilde{\sigma}_{\mathrm{OHLC},t-1}^2 + (1-\lambda) \hat{\sigma}_{\mathrm{OHLC},t}^2}$$

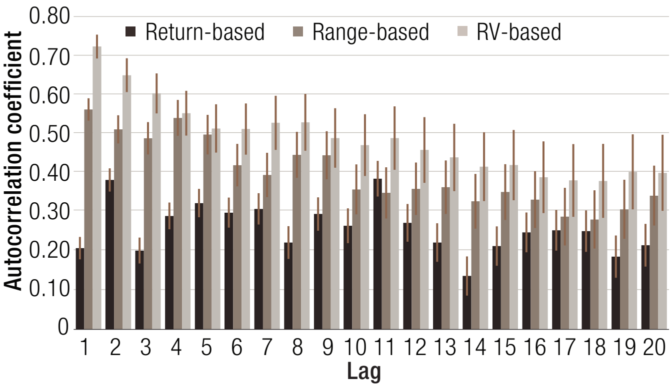
and

$$\tilde{\sigma}_{\mathrm{RV},t} = \sqrt{\lambda \tilde{\sigma}_{\mathrm{RV},t-1}^2 + (1-\lambda) \hat{\sigma}_{\mathrm{RV},t}^2},$$

their respective half-life coefficients for a similar level of dispersion are approximately 5.7 and 3.3 days. Figure 4 illustrates the superior reactivity of the estimators based on intraday data when compared to the measure using close-to-close returns. Focusing on the volatility shock of February 2018 on S&P 500 returns, we see that the range-based estimator has a faster volatility decay than the standard RiskMetrics model, while still being very reactive in case of new market events.

FIG 3. AUTOCORRELATION OF DIFFERENT VOLATILITY MEASURES SHOWING THEIR PERSISTENCE BASED ON CLOSE-TO-CLOSE RETURNS, RANGE-BASED ESTIMATORS, AND HIGH-FREQUENCY RETURNS

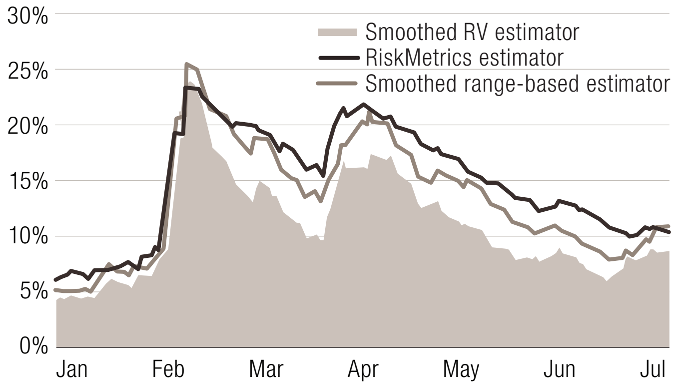
AUTOCORRELATION FUNCTION



Source: LOIM. For illustrative purposes only. Vertical lines indicate the errors bars at the 95% confidence level.

FIG 4. EXPONENTIALLY WEIGHTED MOVING AVERAGES OF VOLATILITY MEASURES SHOWING THEIR REACTIVITY BASED ON CLOSE-TO-CLOSE RETURNS, RANGE-BASED ESTIMATORS, AND HIGH-FREQUENCY RETURNS

JANUARY – JUNE 2018



Source: LOIM. For illustrative purposes only.

Conclusion

In short, using range-based estimators can help obtain well-behaved volatility measures. These measures can replicate most of the empirical properties of realised volatility measures based on high-frequency data, without the computational burden of handling large datasets.

Prior to implementing these in an investment framework, two more areas of investigation are necessary, namely: can such measures help us to better forecast risk? And can such measures improve (risk-adjusted) performance? These topics will be covered in the next two editions of MARS.

References

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

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association, 91(453), 42-55.*

Baltas, N., & Kosowski, R. (2020). Demystifying time-series momentum strategies: volatility estimators, trading rules and pairwise correlations. In S. Satchell, & A. Grant, *Market Momentum: Theory and Practice (S. 30-67).*

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal, 12(3), 1-32.*

Brandt, M. W., & Jones, C. S. (2006). Volatility forecasting with range-based EGARCH models. *Journal of Business & Economic Statistics, 24(4), 470-486.*

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

De Nard, G., Engle, R. F., Ledoit, O., & Wolf, M. (2022). Large dynamic covariance matrices: Enhancements based on intraday data. *Journal of Banking & Finance, 138, 106426.*

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

Engle, R. F., & Patton, A. J. (2001). What good is a volatility model? *Quantitative Finance, 1(2), 237-245.*

Francq, C., & Sucarrat, G. (2023). Volatility estimation when the zero-process is nonstationary. *Journal of Business & Economic Statistics, 41(1), 53-66.*

Francq, C., & Thieu, L. (2019). QML inference for volatility models with covariates. *Econometric Theory, 35(1), 37-72.*

Francq, C., & Zakoïan, J.-M. (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics, 184(1), 158-173.*

Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *The Journal of Business, 53(1), 67-78.*

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *The Journal of Business, 53(1), 61-65.*

Yang, D., & Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices. *The Journal of Business, 73(3), 477-492.*

Important information