Inference from the futures: ranking the noise cancelling accuracy of realized measures

Giorgio Mirone

CREATES Research Paper 2017-24
Inference from the futures:
Ranking the noise cancelling accuracy of realized measures*

Giorgio Mirone†

CREATES, Department of Economics and Business Economics, Aarhus University,
Fuglesangs Allé 4, DK-8210 Aarhus V, Denmark.

December 5, 2017

ABSTRACT
We consider the log-linear relationship between futures contracts and their underlying assets and show that in the classical Brownian semi-martingale (BSM) framework the two series must, by no-arbitrage, have the same integrated variance. We discuss the negligibility of stochastic interest rates using empirical evidence and in simulations. We then introduce the concept of noise cancellation and propose a generally applicable methodology to assess the performance of realized measures when the variable of interest is latent, overcoming the problem posed by the lack of a true value for the integrated variance. We carry out formal testing of several realized measures in the presence of noise and conduct a thorough simulation analysis to evaluate the estimators’ sensitivity to different price and noise processes, sampling frequencies and stochastic components.

JEL classification: C10, C52, C58.

Keywords: realized variance, estimation comparison, noise cancellation, futures, ranking.

*I am grateful to Kim Christensen, Eduardo J. Vera-Valdés, Kristoffer Pons Bertelsen and the participants at the CREATES lunch seminars for the helpful comments. The research project leading to these results received funding from the European Union Research and Innovation programme Horizon 2020 under grant agreement No. 675044 (BigDataFinance).
†E-mail address: gmirone@econ.au.dk
1 Introduction

As of today, high-frequency econometrics is a fast growing field whose mechanisms are yet to be completely unfolded. Given the wide range of practitioners and fields requiring precise volatility estimates, particular interest has been given to the development of an unbiased, noise robust estimator for the integrated variance.

A natural way to estimate the integrated variance over a time period $T$, i.e. $\int_0^T \sigma_t^2 dt$, would be to use the realized variance (i.e. the sum of squared returns). However, due to the presence of the so-called market microstructure noise, using the realized variance as an estimator of the integrated variance would lead to highly biased, inconsistent results. Articles showing this finding are abundant in the literature and extensive details on the topic can be found in: Barndorff-Nielsen and Shephard (2007); Andersen, Bollerslev, and Diebold (2010); Andersen, Bollerslev, Diebold, and Labys (1999); Mykland and Zhang (2009). Therefore, in recent years an ever growing interest in obtaining a bias-free estimate of the integrated variance has been registered and several noise robust estimators have been developed. Yet, an empirical assessment of the quality of the obtained estimates is complicated by the fact that the variable of interest (the integrated variance) is latent and as such remains unobservable even ex-post. A further difficulty is posed by the properties of microstructure noise which varies from series to series and can heavily impact the behaviour of noise robust estimators.

The main technical contribution of this paper is to set out the no-arbitrage relationship between the integrated variance of a futures and its underlying in a continuous Itô semimartingale model for asset pricing. Additionally, we extensively discuss the negligibility of stochastic interest rates and provide simulation results and empirical evidence of the equivalence, both at a local and aggregate level, of the continuous quadratic variation of stocks and futures. To this extent, we base our empirical investigation on Podolskij and Vetter (2009) and Li, Todorov, and Tauchen (2017). Relying on the shown equality, we propose a methodology to evaluate and compare estimation errors of several realized measures.

Related works on the argument are mainly focused on the comparison of forecasts accuracy (see, Aït-Sahalia and Mancini, 2008; Andersen, Bollerslev, and Meddahi, 2011; Li and Patton, 2015). However, a niche literature with a focus on estimation performances has developed methods to obtain consistent rankings using a proxy for the latent variable (see, Patton, 2011). We complement the current literature on estimation accuracy of latent variables by providing a new method with focus on the evaluation of the noise cancelling performances (defined in Section 3) which relies on the results presented in Section 2.3. The noise cancellation ranking doesn’t rely on specific assumptions about the microstructure noise and is therefore generally applicable.

The article is structured as follows: Section 2 sets out the basic framework and the theory on which the paper is based and briefly introduces the topic and the issues that must be assessed. In Section 3 we set out the details of the methodology to be used in the subsequent analysis. Section 4 provides a comprehensive simulation study designed to identify the impact that different noise structures,
noise intensities, stochastic processes and trading intensities can have on the performance of the selected estimators. Section 5 presents an empirical application for the developed methodology using E-mini futures data and ETFs as proxies for the indices underlying the futures contracts. Section 6 concludes.

2 Theory

2.1 Framework

According to the fundamental theory of asset prices, under the assumption of no arbitrage, the log price \( Y_t \) must follow a semimartingale process \((Y \in SM)\) on a filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})\). A discussion is provided by Barndorff-Nielsen and Shephard (2007) and details can be found in Delbaen and Schachermayer (1994). The basic framework of the paper will follow a one-dimensional continuous Itô semimartingale \((Y \in BSM)\) of the form:

\[
Y_t = Y_0 + \int_0^t a_u \, du + \int_0^t \sigma_u \, dW_u,
\]

where \(Y_0\) is \(F_0\)-measurable, \((a_t)_{t\geq 0}\) is a locally bounded and predictable drift process, \((\sigma_t)_{t\geq 0}\) is a càdlàg volatility process and \((W_t)_{t\geq 0}\) is a Brownian motion. For simplicity of exposition, the considered framework rules out the presence of jumps. However, the findings presented can straightforwardly be extended to allow for jumps in the price or even in the volatility process as the introduction of a jump component will not have any impact on the no-arbitrage equivalence.\(^1\)

It is well known (see, Andersen et al. 2010) that if \(Y \in BSM\) then, the integrated variance (our object of interest) is well defined:

\[
IV = \int_0^t \sigma_u^2 \, du.
\]

Further, consider a unit time interval \(t = [0, 1]\). We define \([Y]_t\) the Quadratic Variation process of \(Y\) \(([Y]_t)_{t\geq 0}\) defined for any deterministic sequence of partitions \(0 = t_0 < t_1 < \cdots < t_n = 1\) with

\[
\sup_i \{t_{i+1} - t_i\} \to 0 \quad \text{for} \quad n \to \infty: \quad \text{(check, Protter 2004, pp. 66-77)} \text{ as:}^2
\]

\[
[Y]_t = \lim_{n \to \infty} \sum_{i=0}^{t_i \leq t-1} (Y_{t_{i+1}} - Y_{t_i})^2.
\]

Intuitively, equation (3) shows that the quadratic variation process can be estimated as the sum of squared returns over the interval \([0, t]\) to obtain the in sample variability of the process \(Y_t\). We generally call this estimator the “Realized Variance” (RV). It has been proven that this probability limit exists for all semimartingales and the convergence is also locally uniform in time. (see, for

\(^1\)We can generally say that, in the presence of jumps, a "jump robust" estimator would still estimate the true value of IV instead of QV, while the use of non jump robust estimators would simply provide an estimate of \(QV = IV + JV\), with \(JV\) being the discontinuous component due to jumps (see, Andersen, Bollerslev, and Diebold, 2007).

\(^2\)As discussed in Jacod and Shiryaev 2003, pp. 51-52 the quadratic variation process is still well defined for any Riemann sequence of adapted subdivisions, allowing to relax the assumption of deterministic times.

Hence, as \( Y \in BSM \):

\[
[Y]_t = \int_0^t \sigma_u^2 \, du. \tag{4}
\]

Consequently, according to equation (3), computing the quadratic variation using tick data should deliver an extremely precise estimate of the integrated variance over a single time interval (say from 0 to \( T \)). Indeed, the literature teems with examples of the use of realized variance as estimator of the integrated variance (e.g., Andersen et al. 2001; Barndorff-Nielsen and Shephard 2007; Zhang, Mykland, and Aït-Sahalia 2005 or Andersen et al. 2010 for a survey). However, it has been widely demonstrated that such estimator diverges when the sampling interval becomes too thin, providing inconsistent results (see, Andersen et al. 1999; Bandi and Russell 2008; Aït-Sahalia, Mykland, and Zhang 2011). A visual evidence of the potential bias can be obtained by examining the so-called volatility signature, obtained by plotting the sample average realized variance as a function of the sampling frequencies.

**Figure 1.** Signature plot

Note: Volatility signature plot for the E-mini Nasdaq 100 with September maturity: as initially shown by (Andersen et al., 1999), the values estimated using the RV estimator explodes as the sampling frequency increases.

Figure 1 presents the average daily realized volatility for the E-mini Nasdaq 100 with maturity...
September 2013. It can easily be seen that the average realized variance computed at the highest frequencies becomes unreliable, while at lower frequencies it provides an unbiased estimate of the average integrated variance.

2.2 Noise

The problem lies in the peculiar statistical properties of high-frequency data and, more generally, in a wide variety of phenomenon generally defined “market microstructure” which include, but are not limited to, liquidity effects, market imbalances, discreteness of prices and bid-ask spreads. Since these frictions affect the true efficient price to a large extent it becomes realistic to assert that we are not able to observe the true transaction price $Y_t$ but only a noisy process, expressed through the additive model:

$$X_t = Y_t + U_t.$$  \hspace{1cm} (5)

Where $(U_t)_{t \geq 0}$ is an error term that captures the effect of market frictions and $(X_t)_{t \geq 0}$ is the observable log-price process. 

Hansen and Lunde (2006a) present detailed insights on the structure of the noise process, showing that signs of dependence with the price process and autocorrelation should be modelled when sampling at ultra high frequencies. However, as the structure of the noise process has no impact on the results presented at the end of this section, we will introduce stricter restrictions on the noise component. This will ease the illustration in our theoretical framework and simultaneously grant consistency of the realized measures we are going to study in sections 4 and 5. Specifically, we will assume that the error terms $U_t$ follow a white noise process ($U \in \mathcal{WN}$). Under this assumption the error term will have:

$$\mathbb{E}(U_t) = 0, \quad \mathbb{E}(U_t^2) = \omega^2, \quad U_t \perp U_s, \; t \neq s.$$  \hspace{1cm} (6)

While the white noise assumption is not realistic, it is useful for illustration purposes when considering market frictions as operating in tick time (Barndorff-Nielsen, Hansen, Lunde, and Shephard, 2008). Further, the independent noise assumption may still be valid when sampling prices at lower sampling frequencies, from one minute onward (see, Hansen and Lunde, 2006a). Additionally, we will assume the error term to be exogenous, i.e. independent from the true underlying process $Y_t$ ($U \perp Y$). This is again a strict assumption as it might be expected the noise process $U$ to be correlated with increments in the price process $Y$ (see, Kalnina and Linton, 2008; Hansen and Lunde, 2006a). However, Hansen and Lunde (2006a) suggests that the assumption does not significantly impact the estimates when analysing densely traded stocks.

Although we decided to rely on strict assumptions for the noise process in the theoretical framework, for the simulation study we will allow for a more general structure of the noise component taking into account the findings presented in Hansen and Lunde (2006a). This will allow us to examine the estimators behaviour under different noise processes. Particularly we will assume dependency
between the noise and the underlying price process and relax the white noise assumption. Further, the empirical noise structure will also be analysed in section 5 of the paper.

As previously mentioned, with the introduction of a noise process in the framework our findings on the quadratic variation process are in jeopardy. In fact, we can now see that at the highest frequencies market frictions greatly affect the underlying price process and subsequently the quadratic variation process which will then provide inconsistent estimates of the integrated variance.

Technically, defining $RV_t^{(m)}$ the realized variance estimated upon $m$ intraday returns for day $t$ and letting $m \to \infty$ (i.e., when we work in the in-fill asymptotics) we have:

$$RV_t^{(m)} \overset{p}{\to} \infty, \text{ as } m \to \infty.$$  \hspace{1cm} (7)

Particularly, given (6), we will have:

$$RV_t^{(m)} \approx 2m\omega^2.$$  \hspace{1cm} (8)

Where $\overset{p}{\to}$ indicates limit in probability. Details can be found in Hansen and Lunde (2006a); Andersen et al. (2010). The main insight we obtain from equation (8) is that in the presence of contaminating noise as modelled in (5), RV will not deliver consistent estimates of the integrated variance but rather, through appropriate scaling, the variation of the noise itself.

2.3 Different series - same volatility

This section presents the primary theoretical contribution of this paper, showing that, in the Brownian semimartingale framework, futures of any maturity and their underlying asset must, by no arbitrage, share the same volatility.

The key ingredient required to evaluate an estimation error is the availability of a true, known value to measure the distance between the two. Unfortunately, even ex-post, our variable of interest (the integrated variance) remains unobservable, not allowing the accuracy of the obtained estimate to be directly verified. Therefore, while the recent ground-breaking advancements in the field have made available a different number of bias-correcting estimators, their relative empirical performance has not yet been assessed as the true value of the underlying integrated variance remains unknown. In the current section we will present a way to make up for this shortcoming using futures contracts to overcome the problem and identify the presence of bias in the obtained estimates even if the variable of interest remains indiscernible.

Consider again the simple case without noise. We define $F_{t,T} = \log(P_{t,T})$ to be the futures log-price. We know that (see, Björk, 2009) the futures price for final settlement at time $T$ is given by:

$$P_{t,T} = \mathbb{E}_t^Q [S_T].$$

Where $S_T = \exp(Y_T)$ is the level price of the underlying asset at time $T$. Hence, under risk neutral probability, given the no-arbitrage assumption, we can express the relation between futures and
spot prices in continuous time:
\[ P_{t,T} = e^{Y_t + \int_t^T r_u du}, \]  
(9)

where \( r_t \) is a deterministic, continuously compounded interest rate.\(^3\)

Taking logarithms, we obtain:
\[ F_{t,T} = Y_0 + \int_0^t a_u du + \int_0^t \sigma_u dW_u + \int_t^T r_u du. \]  
(10)

Evidently, from (10), we can see that the integrated variance of a futures contract is given by:
\[ IV_{F_{t,T}} = \int_0^t \sigma_u^2 du. \]

Hence, it directly follows:\(^4\)
\[ IV_{F_{t,T}} = IV_{Y_t}. \]  
(11)

Clearly, this result will hold true for every futures written on the same underlying asset. Therefore, considering two different futures \( F_{t,T_1} \) and \( F_{t,T_2} \), with \( T_1 \neq T_2 \) and both written on asset \( Y_t \), we will have that:
\[ IV_{F_{t,T_1}} = IV_{F_{t,T_2}} = IV_{Y_t}. \]

Additionally, these findings will not change when introducing stochastic interest rates into the model as the impact of the new stochastic factor on the integrated variance would be negligible. The intuition is that the interest rate would simply add up to the drift term, whose order of magnitude is smaller than the dominating diffusion term.\(^5\)

Unfortunately, being in a high-frequency, noisy, framework we need to take into account the impact of market microstructure frictions.

The futures price process is itself affected by market microstructure frictions. Expressing the previously introduced noise process \( U_t \) as \( U_t = \omega_1 Z_{1t} \), with \( Z_1 \sim \mathcal{N}(0,1) \) we can define \( V_t \) to be an \( iid \) error term affecting the futures price process of the form:
\[ V_t = \omega_2 \rho Z_{1t} + \omega_2 \sqrt{1 - \rho^2} Z_{2t}, \]  
(12)

where, \( Z_{2t} \sim \mathcal{N}(0,1) \) is independent from \( Z_{1t} \) and \( \rho = \text{corr}(U_t, V_t) \).

The rationale behind this definition of \( V_t \) comes from the idea that the noise process affecting \( S_t \) influences, to a certain extent also the futures price process. In turn, \( F_t \) is affected by a second

\(^3\)We investigate the effect of stochastic interest rates in section 4 and show that the impact on the quadratic variation of the futures is of four orders of magnitude smaller than the integrated variance; therefore, negligible.

\(^4\)Interestingly, a closely related finding with a focus on range based volatility estimation is discussed in Rossi and Santucci de Magistris (2013).

\(^5\)This is further validated by several empirical and simulations evidence present in the literature, which shows how the use of alternative stochastic interest rate models to price index futures do not generally provide significantly different results from the constant interest rate model presented in equation (9) (see, e.g., Cornell and French, 1983; Cakici and Chatterjee, 1991).
source of noise proper to the market on which the instrument trades. Therefore, equation (12) allows us to define the noise term which encompass the effect of both the common $Z_1$, and idiosyncratic component $Z_2$ of the noise term affecting $F_t$.

We then define:

$$K_{t,T} = F_{t,T} + V_t.$$  \hspace{1cm} (13)

Consequently, we will have again:

$$\lim_{m \to \infty} RV_{K,t}^{(m)} \overset{p}{\to} \infty,$$

as $m \to \infty$.

However, the magnitude of divergence will not necessarily be the same as the two price processes are subject to different microstructure noise, which will not affect equally the two processes. A clear evidence can be obtained by jointly plot the volatility signature of the two series as in figure (2).

**Figure 2.** Signature plot for ES-SPY

![Volatility Signature Plot](image)

Note: The figure shows the signature plot for the Emini S&P 500 with December expiration together with the underlying asset proxied by the SPY. It can be easily seen how the two series are differently affected by the microstructure noise.

Fortunately, the use of noise robust estimators allows us to retrieve the true integrated variance by isolating and removing the bias component from the estimates; we should therefore obtain the same result when estimating the IV of the two series using the same estimator.
Empirically, we can visually verify this finding analysing a single trading day as an example. The different price exhibited by the two contracts (the so-called futures basis) is *approximately*\(^6\) constant and entirely imputable to the different discount rate used to price the futures (Chen, Cuny, and Haugen, 1995).

**Figure 3.** Sample trading activity

![Sample trading activity graph](image)

Note: The day trade occurred on September 18 2013 of the QQQ, the NQ Sept. and the NQ Dec. are plotted. It can easily be seen that the price of the underlying and the two futures move identically. Hence, when correctly eliminating the bias caused by microstructure noise, the two estimated volatilities should not be significantly different from each other. The price of the QQQ has been scaled up to allow the comparison

### 2.3.1 The case of stochastic interest rates

Although the assumption of deterministic interest rates is widely employed in the literature on futures contracts readers might consider it as excessively restrictive. For this reason, in Section 5.2 and Section 5.3 we aim to investigate to which extent such an assumption is legitimate by empirical evidence. Further, in Section 4.2.1 we investigate the special case where the interest rate is assumed to follow a CIR process Cox, Ingersoll, and Ross (1985) and provide results showing the negligibility of the bias if the interest rate is assumed non-stochastic.

\(^6\)Indeed, market microstructure differences, e.g. different tick sizes, and more generally the microstructure noise might lead to small deviations from the equality.
3 Methodology

In this section we introduce the concept of noise cancellation and formalize a method to consistently evaluate the noise cancelling accuracy of a given estimator and compare its relative performance with respect to the competing estimators. We base our methodology on the previously shown relationship between a realized measure computed on a futures and the same measure computed on the underlying asset and use a proper measure of distance to evaluate the noise cancelling performance of each estimator.

To ease the notation we introduce the concepts considering only one futures contract $F$ and its corresponding underlying asset $S$; we then define:

$$\Theta_{t,i} = \widehat{IV}_{t,i}(S) \text{ and } \Gamma_{t,i} = \widehat{IV}_{t,i}(F)$$

with $t = 1, \ldots, N$; and $i = 1, \ldots, M$ and where $N$ is the number of sample days and $M$ the number of competing estimators considered; $\widehat{IV}$ refers to the obtained estimate of the integrated variance.

3.1 The noise cancellation

We define noise cancellation as the finite sample ability of an estimator to eliminate the bias caused by the presence of noise and we consider it as a factor of primary interest in the estimation of noisy variables. Indeed, while it has been already assessed that many realized measures are able to provide an estimate of the integrated variance with a good degree of accuracy, it is a fact that none is able to obtain exact results (due to the nature of the estimator, the particular characteristic of the price process, etc.). Further, if a realized measure is not able to completely remove the effect of microstructure noise, an additional source of inaccuracy will be present, affecting the overall performance of the estimator.\(^7\) In this situation, the estimator will further be influenced by the characteristics of the noise process and, if applied to two different time series of the same efficient price process but affected by different noises (as in the case of a futures and its underlying asset) it will produce different estimates of the same latent variable.

Recall the two noise processes $U_t$ and $V_t$ as previously defined

$$U_t = \omega_1 Z_{1,t}$$
$$V_t = \omega_2 \rho Z_{1,t} + \omega_2 \sqrt{1 - \rho^2} Z_{2,t}.$$
with $Z_{1t}, Z_{2t} \sim \mathcal{N}(0, 1)$.

Then, given any unbiased realized measure $RM_{i}$, we can write

$$RM_{t,i}(S) = \hat{IV}_{t,i}(S) = IV_{t} + \varepsilon_{t,i}$$
$$RM_{t,i}(F) = \hat{IV}_{t,i}(F) = IV_{t} + \eta_{t,i}$$

and $\varepsilon$ and $\eta$ respectively being the residual components of $U$ and $V$. Hence, knowing that $\mathbb{E}[\varepsilon_{t,i}] = 0 \ \forall \ i$, it follows that any increasing function$^{9}$ will maintain the order between the noise components and their residual terms; that is to say:

$$U_{i} \lessgtr V_{t} \iff \varepsilon_{t,i} \lessgtr \eta_{t,i} \ \forall \ i.$$  \hspace{1cm} (14)

It must be further noticed that we can obtain (14) if we apply a biased but linear and increasing function.$^{10}$ However, in this case $\varepsilon$ and $\eta$ will incorporate the bias term and their expected value will differ from zero. Nonetheless, the linearity of the function will still entail the order preservation. Consequently, given any appropriate pseudo-distance measure $L_{t,i}(\Theta_{t,i}, \Gamma_{t,i})$ we will have

$$L_{t,i}(\Theta_{t,i}, \Gamma_{t,i}) = 0$$

if and only if

$$IV_{t} + \varepsilon_{t,i} = IV_{t} + \eta_{t,i}$$

which implies

$$\varepsilon_{t,i} = \eta_{t,i} = 0.$$ 

Therefore, the existence of two uncorrelated noise components grants that the loss function will be exactly equal to zero only if no residual noise component is present. It follows that $L_{t,i}(\Theta_{t,i}, \Gamma_{t,i})$ can be consistently used to assess and compare the finite sample ability of each estimator to obtain a noise free estimator.

Unfortunately, the noise cancellation ranking is not guaranteed to provide a consistent ranking for the IV estimates. That is to say:

$$\mathbb{E}[L(\Theta_{t,i}, \Gamma_{t,i})] \lessgtr \mathbb{E}[L(\Omega_{t,j}, \Gamma_{t,j})] \iff \mathbb{E}[L(IV_{t}, \Gamma_{t,i})] \lessgtr \mathbb{E}[L(IV_{t}, \Gamma_{t,j})], \text{ with } i \neq j.$$ 

In fact, following Patton (2011),$^{11}$ we can easily see that the log-linear relationship between underlying and futures implies a perfect correlation of the estimation error of $\Theta$ and $\Gamma$ with the variable of interest which breaks the equality $\mathbb{E}[\Delta L(\Gamma_{t}, \Theta_{t})] = \mathbb{E}[\Delta L(IV_{t}, \Theta_{t})]$.

$^{8}$The unbiasedness of the realized measure grants us the possibility of doing inference on the residual term as we know that the expected value of both $\varepsilon_{t,i}$ and $\eta_{t,i}$ is equal to zero and no further, unknown bias term will contaminate the estimate.

$^{9}$This is the case for any realized measure.

$^{10}$This is the case of the realized variance for example.

$^{11}$For details refer to: Patton 2011, Appendix, Proof of Proposition 1.
 Fortunately, this is unimportant in our empirical application as it will not hinder the assessment of the noise cancelling accuracy.

To obtain a robust ranking we implement a Model Confidence Set (MCS) procedure (Hansen, Lunde, and Nason, 2011) to identify the set of superior estimators, discarding the ones identified as significantly inferior. Additionally, we apply a Diebold-Mariano (DM) type test (Diebold and Mariano, 1995) to obtain a pairwise comparison of the analysed estimators’ accuracy.

To compare the competing estimators we need to define a proper measure of their distance. Our choice fell on the popular QLIKE loss function as several empirical and simulation results presented in the literature suggests more power to reject inferior estimators (see, Hansen and Lunde, 2005; Patton and Sheppard, 2009; Patton, 2011).

$$L_{t,i}(\Theta_{t,i}, \Gamma_{t,i}) = \frac{\Theta_{t,i}}{\Gamma_{t,i}} - \log \frac{\Theta_{t,i}}{\Gamma_{t,i}} - 1.$$  

Hence, we will obtain a distance of zero when $\Theta_{t,i} = \Gamma_{t,i}$. The relative accuracy of any two competing estimators $i$ and $j$ is then appraised through the loss differential:

$$d_{ij,t} = L_{t,i} - L_{t,j}. \tag{15}$$

Additionally,

$$\hat{d}_{ij} = \frac{1}{N} \sum_{t=1}^{N} d_{ij,t}$$

will provide a consistent estimate of the difference in accuracy between any two realized measures (as $N \to \infty$) and under standard regularity conditions\footnote{Following the literature on the argument, throughout the paper we assume the loss differential series to be strictly stationary and short memory (see, Diebold and Mariano, 1995; Hansen and Lunde, 2006b) and Patton (2011).} we can implement bootstrap methods to test for differences in the estimators’ performance.

### 3.2 Pair-wise comparison

Evidently, the two estimators will have equal accuracy if and only if $\mathbb{E}[d_{ij,t}] = 0 \forall t$ (i.e. the loss differential function has expectation equal to zero for all $t$). Hence, we will test the null hypothesis $H_0 : \mathbb{E}[d_{ij,t}] = 0$, against the alternative $H_1 : \mathbb{E}[d_{ij,t}] \neq 0$.

Straightforwardly, we can apply a Diebold-Mariano type test to obtain pair-wise comparisons of the competing estimators.

$$t_{ij} = \frac{\hat{d}_{ij}}{\sqrt{\hat{\text{var}}(\hat{d}_{ij})}}, \tag{16}$$

where $\hat{\text{var}}(\hat{d}_{ij})$ is the Newey-West estimate of the loss differential series variance. Then, the null hypothesis will be rejected in favor of the alternative when $|t_{ij}|$ exceeds the critical
value of a standard normal distribution.

3.3 Joint comparison

We rely on the Model Confidence Set (MCS) approach of Hansen et al. (2011) to jointly test the performance of our realized measures and identify and discard inferior estimators. Starting from an initial set of objects $\mathcal{M}^0$, the MCS procedure allows us to obtain a set of superior objects $\mathcal{M}^* = \{i \in \mathcal{M}^0 | \mathbb{E}(d_{ij,t}) \leq 0 \ \forall \ j \in \mathcal{M}^0\}$, which will contain the best object with a given level of confidence ($\alpha$). In this sense, the procedure is well suited to compare the estimation accuracy of a large set of competing measures.

The MCS follows a multiple step procedure; it initially sets $\mathcal{M} = \mathcal{M}^0$ and then checks the null hypothesis $H_{0,\mathcal{M}}$ at a given level $\alpha$ through the use of an equivalence test $\delta_{\mathcal{M}}$. Then, in the event that $H_{0,\mathcal{M}}$ is rejected ($\delta_{\mathcal{M}} = 1$), the algorithm identifies the object to be removed using an elimination rule $e_{\mathcal{M}}$ and tests again the new set of elements. The procedure stops when $H_{0,\mathcal{M}}$ cannot be rejected, in which case the set of superiors objects (at the given confidence level) is identified: $\mathcal{M}^*_1 = \mathcal{M}$.

Following Hansen et al. (2011) we will adopt the range equivalence test and its coherent elimination rule. This approach grants us that as the null hypothesis $H_{0,\mathcal{M}}$ is rejected for any given set $\mathcal{M}$, the elimination rule will identify and remove the worst performing estimator still included in the set. Hence, we will now test the null hypothesis of equal loss differentials jointly for all the models in the set.

$$H_{0,\mathcal{M}}: \mathbb{E}[d_{ij}] = 0 \ \forall \ i, j \in \mathcal{M}.$$ 

The null is tested by using the following statistic:

$$T_{R,\mathcal{M}} = \max_{i,j \in \mathcal{M}} |t_{i,j}|,$$

where $t_{i,j}$ has been introduced in equation (16) and tests the null $H_{0,ij}: \mathbb{E}[d_{ij}] = 0$.

While the asymptotic distribution of this statistic is non-standard, the use of appropriate bootstrap methods allows to retrieve the relevant distributions.

Finally, the associated elimination rules follow naturally; for $T_{R,\mathcal{M}}$ we have:

$$e_{R,\mathcal{M}} = \arg \max_{i \in \mathcal{M}} \sup_{j \in \mathcal{M}} t_{ij},$$

as the statistic is built such that $t_{k,e_{R,\mathcal{M}}} = T_{R,\mathcal{M}}$ for some element $k \in \mathcal{M}$.

4 Simulation

In the simulation exercise we analyse the performance of the chosen competing estimators for six different models under different characteristics and combinations of the noise process and different trading intensities ($\lambda = [23400; 7800; 1560; 390; 78; 39]$) in order to check the sensitivity of the
analysed estimators to a wide variety of scenarios. Two different structures (IID and MA(1)) of the noise process and three different levels of noise will be considered in our study: $(\gamma_L \approx 0.5, \gamma_M \approx 2.5$ and $\gamma_H \approx 5)$ corresponding respectively to a low, medium and high level of noise (Christensen, Oomen, and Podolskij, 2010). Following equation (12) the two noise processes will be correlated with each other with a fixed correlation parameter $\rho_{V,U} = -0.7$.

The simulation is structured as follows for $R = 10000$ replications:

- Generate a series of $N = 23400$ observations using six different DGPs for both the underlying and the futures.
- Generate two IID and two non-IID (MA(1)) noise processes ($U_t$ and $V_t$) correlated following equation (12) and create the noisy price series.
- Obtain the estimates of the integrated variance from the noisy series using the competing estimators. Repeat for different levels of trading intensity.
- Compute the QLIKE loss from the obtained estimates and construct a sequence of $R$ losses.
- Determine the MCS and DM pairwise performances.
- Repeat the previous steps for each level of noise.

### 4.1 The estimators

In both the simulation exercise and the empirical study presented in section 5 we consider six different estimators of the integrated variance, which are good representatives of different classes of realized measures. All the estimates are obtained using tick time sampling.

The first estimator considered is the pre-averaged realized variance (RPA), initially introduced by Podolskij and Vetter (2009) and Jacod, Li, Mykland, Podolskij, and Vetter (2009). In our application we closely follow the work of Christensen, Oomen, and Podolskij (2014) and therefore select a kernel bandwidth $K = \Theta \sqrt{n}$, with the pre-averaging horizon $\Theta = 1$ and where $n$ refers to the number of intraday returns.

The second estimator included in our study is the realized kernel (RK) of Barndorff-Nielsen et al. (2008). Following Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) we opted for the non-flat-top Parzen kernel with convergence rate $n^{-1/5}$, which grants always positive estimates and robustness to autocorrelated noise.

The third and fourth estimators fall both in the same class of realized measures; they are the two-scale realized variance (RVTS) of Zhang et al. (2005) and the subsequent multi-scale realized variance (RVMS) by Zhang (2006). The main difference between the two is their rate of convergence, being $n^{-1/6}$ for the RVTS and improving to $n^{-1/4}$ for the RVMS.

The fifth estimator selected for the study is the simple realized variance using data sampled at a 5 minutes frequency (RV5). We chose to include this estimator as it is widely recognized in the literature as a consistent estimator of the integrated variance. In fact, lowering the frequency

---

13 with MA parameter: $\phi_{U,1} = \phi_{V,1} = 0.8$

14 We define the Noise Ratio $\gamma$ introduced by Oomen (2006) in section 5.4.
allows to benefit the improved accuracy from higher frequency data while granting a reduction of
the negative effect of market microstructure noise. Although there are ways to obtain an optimal
sampling frequency (the second best estimator of Zhang et al. (2005) is an example), we opted for
the RV5 as it represents a good compromise between precision and simplicity and has widely been
used in the literature (e.g., see: Andersen, 2000; Andersen et al., 2011; Bianco, Corsi, and Renò,
The last estimator analysed is the standard realized variance (RV) computed at the highest possible
frequency. The idea behind the inclusion of an estimator acknowledged as biased is to have a
“robustness check” for our methodology as it allows to verify if the proposed procedure is able
to identify and discard the known lower performing estimator. Additionally, the presence of the
RV becomes pertinent when analysing the performance of the pool of realized measures for lower
trading intensities, where the effect of market microstructure becomes negligible.

4.2 The models

This section introduces the DGPs used in the simulation exercise together with the parameters
values used for each. The models have been chosen to assess how different stochastic components
and their combinations affect the performance of the considered realized measures.
We opted to generate a main “unrestricted” model encompassing four different stochastic com-
ponents and further include in our study five restricted versions of the former where one or more
stochastic components have been removed. This allows us to identify both the effect of each isolated
component and the impact that different stochastic terms have jointly on the realized measures
investigated.
Particularly, in generating the main model we allowed for realistic features such as stochastic volatility
and interest rates and jumps in the volatility process and in the price process. Table 1 presents
a detailed summary of the unrestricted model, while Table 2 reports the parameters used in the
simulation and clarifies the relationship between the different DGPs considered. For the sake of
brevity we decided to report only results for the unrestricted model, DGP(6) and salient findings
for the five restricted versions are briefly discussed.\footnote{The complete results for the restricted models are available upon request.}

\textsuperscript{15}
Table 1 The unrestricted model

\[ dS_t = \mu dt + \tilde{\sigma}_{S,t} dW_{1,t} + dJ_{S,t} \]
\[ d\log \tilde{\sigma}^2_{S,t} = d\log \sigma^2_{S,t} + dJ_{v,t} \]
\[ d\sigma^2_{S,t} = \kappa_S \left( \theta_S - \sigma^2_{S,t} \right) dt + \sigma_v \sigma_{S,t} dW_{2,t} \]
\[ dJ_{v,t} = \sum_{i=1}^{N_t} Y_{v,i} \]
\[ dJ_{S,t} = \sum_{i=1}^{N_t} Y_{S,i} \]
\[ dr_t = \kappa_r (\theta_r - r_t) dt + \sigma_r \sqrt{r} dW_{3,t} \]

With \( S_t \) the log price process of the underlying asset.

Table 2 Parameters values

<table>
<thead>
<tr>
<th>Including DGP:</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>( 2.4e^{-3} )</td>
</tr>
<tr>
<td>( \sigma^2_S )</td>
<td>( 3.6e^{-4} )</td>
</tr>
<tr>
<td>( r )</td>
<td>( 3.6e^{-4} )</td>
</tr>
<tr>
<td>( \kappa_S )</td>
<td>( 0.5 )</td>
</tr>
<tr>
<td>( \theta_S )</td>
<td>( 3.6e^{-4} )</td>
</tr>
<tr>
<td>( \sigma_v )</td>
<td>( 3.0e^{-3} )</td>
</tr>
<tr>
<td>( \sigma_{J_S} )</td>
<td>( 9.0e^{-7} )</td>
</tr>
<tr>
<td>( \lambda_{J_S} )</td>
<td>( 3 )</td>
</tr>
<tr>
<td>( \sigma_{J_v} )</td>
<td>( 9.0e^{-4} )</td>
</tr>
<tr>
<td>( \lambda_{J_v} )</td>
<td>( 3 )</td>
</tr>
<tr>
<td>( \kappa_r )</td>
<td>( 2 )</td>
</tr>
<tr>
<td>( \theta_r )</td>
<td>( 4.0e^{-4} )</td>
</tr>
<tr>
<td>( \sigma_r )</td>
<td>( 0.02 )</td>
</tr>
</tbody>
</table>

We mark with ✓ a parameter included by the model and with X when the parameter is implemented as initial value of the stochastic process. Following Christensen et al. (2014), we choose the size of \( \sigma_{J_v} \) such that \( JV = \sum_{i=1}^{N_t} \sigma^2_{J_v} \approx 0.2 \) where \( [X]_1 = \int_0^1 \sigma^2_{u} du + \sum J^2 \).

4.2.1 The case of stochastic interest rates - Simulation analysis

Given our assumption in section 2.1, of particular interest is DGP(5). From this model we can isolate and evaluate the impact that stochastic interest rates have on our results.

It is well known that the choice of an affine model for the short rate allows to obtain a closed form solution for the price process of a futures on a zero-coupon bond (see, e.g. Munk, 2011). Additionally, under the assumption of independence between the stochastic process driving the interest rate and the one driving the underlying price process, we can follow Ramaswamy and Sundaresan (1985)
to expand this finding and obtain a closed-form solution for a futures written on any asset. Thus, to generate DGP(5) and DGP(6) we consider a parametric setting and model the interest rate dynamics as a CIR process (Cox et al., 1985). Consequently, instead of (9) we have:

$$F = Sa(\tau)e^{b(\tau)r}$$

(17)

where:

$$a(\tau) = \left\{ \frac{2\gamma \exp\{\gamma + \kappa_r\tau\}/2}{2\gamma + (\gamma + \kappa_r)(\exp\{\gamma\tau\} - 1)} \right\}^{2\kappa_r \theta_r / \sigma_r^2}$$

$$b(\tau) = \frac{2(\exp\{\gamma \tau\} - 1)}{2\gamma + (\gamma + \kappa_r)(\exp\{\gamma\tau\} - 1)}$$

and $\gamma = \sqrt{\kappa_r^2 - 2\sigma_r^2} > 0$ (by assumption).

Indeed, the introduction of a stochastic interest rate will affect the QV of the futures, which might be problematic in view of equation (11). Straightforwardly, we evaluate the impact as:

$$[\Delta]_t = [F]_t - [Y]_t$$

where $F$ and $Y$ are the efficient log-prices of the futures and the underlying as defined in (17) and (1) respectively; we further have

$$[\Delta]_t = \sigma_r^2 \int_0^t b_r^2 r_u du$$

(18)

The simulation results are reassuring and show that $[\Delta]_t$ is about four orders of magnitude smaller than the integrated variance. This implies that the bias arising from the assumption of non-stochastic interest rate has an impact on the quadratic variation of $F$ of about 0.01% and is therefore negligible.17

4.3 Simulation Results

An analysis of the results obtained highlights an overall invariance across the different models analysed, which suggests robustness of the estimators to the effect of different stochastic components. The only noticeable difference being the inclusion of RK in the confidence set for high levels of trading intensity ($\lambda = 23400$, $\lambda = \text{All}$) for all the DGPs including a stochastic interest rate component. Otherwise, only minor variations in the order of elimination of the sub-optimal estimators can be evidenced. For ease of exposition only the $T_R$ test results for the unrestricted model will be presented.

---

16 As $a$ and $b$ are continuously differentiable deterministic functions, we obtain the results in Equation 18 by simply applying Itô’s lemma to $b(\tau)r$.

17 We tested our results with different combinations of parameters for the interest rate process and obtained values ranging from 0.01% to approx 0.04%.
Table 3 MCS Ranking - DGP(6) - \( T_R \)

<table>
<thead>
<tr>
<th></th>
<th>All λ = 23400</th>
<th>λ = 7800</th>
<th>λ = 1560</th>
<th>λ = 390</th>
<th>λ = 78</th>
<th>λ = 39</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IID MA(1)</td>
<td>IID MA(1)</td>
<td>IID MA(1)</td>
<td>IID MA(1)</td>
<td>IID MA(1)</td>
<td>IID MA(1)</td>
</tr>
<tr>
<td>RPA:</td>
<td>γ_{θ}</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>M_{4}</td>
</tr>
<tr>
<td>RK:</td>
<td>γ_{θ}</td>
<td>M_{4}</td>
<td>M_{4}</td>
<td>M_{4}</td>
<td>M_{4}</td>
<td>M_{4}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{4}</td>
<td>M_{4}</td>
<td>M_{4}</td>
<td>M_{4}</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{4}</td>
<td>M_{4}</td>
<td>M_{4}</td>
<td>M_{4}</td>
<td>M_{4}</td>
</tr>
<tr>
<td>RVTS:</td>
<td>γ_{θ}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
<td>M_{2}</td>
</tr>
<tr>
<td>RVM:</td>
<td>γ_{θ}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
<td>M_{3}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
</tr>
<tr>
<td></td>
<td>γ_{θ}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
<td>M_{0}</td>
</tr>
</tbody>
</table>

Included estimators (X) and elimination order for the \( T_R \) test with 10,000 bootstrap replications at a 95% confidence level. \( \gamma_{θ} = 5.22 \), \( \gamma_{θ} = 2.43 \) and \( \gamma_{θ} = 0.57 \). DGP(6) is our unrestricted model which allows for four different stochastic components: Stochastic volatility, stochastic interest rates, jumps in the price process and jumps in the volatility process. The different colors help comparing the results for each level of noise intensity.

Table 3, reports the MCS results for the unrestricted model; \( M_0 \) identifies an element included in the initial set of elements and eliminated at the first iteration, (i.e. the one identified as worst performing), \( M_1 \) identifies an element included in \( M_1 \) and so forth.

The Model Confidence Set procedure shows a constant over performance of the pre-averaged realized variance estimator over the competing realized measures for high and medium levels of trading intensity (\( \lambda \geq 1560 \)) irrespective of the noise level considered. However, for high values of \( \gamma \) with a coloured noise process, the procedure tends to include simultaneously the RK and the RPA in the set of superior elements. Additionally, we can infer that the introduction of a stochastic component for the interest rate has a negative impact on the pre-averaged realized variance performance as for both DGP(5) and DGP(6) the MCS fails to identify any statistical difference between RK and RPA, independently of the noise structure considered.

Yet, when we set the trading intensity to low values, the performance of RPA drops sharply and the estimator often gets eliminated during the third or fourth iteration of the MCS procedure. At the same time, the results obtained by the Realized Kernel show a stable, remarkable performance across all trading intensities and level of noise. In fact, while it is true that for high levels of \( \lambda \) the estimator is often marked as under performing compared to the pre-averaged realized variance, it only gets eliminated from the set of optimal elements during the last iteration of the procedure. On the other hand, RK gets constantly selected, together with the simple realized variance, as optimal estimator for very low levels of trading intensity (\( \lambda \leq 390 \)). This suggests that RK is not sensitive
to the level of $\lambda$ and its noise cancelling accuracy can be relied upon even for low liquid stocks.

Finally, according to expectations, we can observe that RV gets always immediately discarded for high and medium trading intensities across all the models considered and for each level of noise analysed but tends to over perform as the trading intensity gets low, when the effect of market microstructure noise becomes negligible.

However, when analysing the deviation from the true IV, the results are not as clear. While for the naïve DGP(1) and DGP(2) the MCS selects the RVMS as over-performing, when different stochastic sources are included the estimation accuracy tends to reflect the results obtained with the noise cancellation ranking. Nonetheless, the RVMS seems to show better performances overall and in some instances it gets included in the set of superior elements together with the RK and the RPA for ($\lambda \geq 1560$). A clear explanation for these results can be obtained inspecting the average deviation from the true IV. Table 4 shows the standardized values of the estimated IV for the unrestricted model. A quick look at the table reveal the values of the non-trivial estimators (RK, RPA, RVTS and RVMS) to be very close to the true latent variable and to each other which explain why the model confidence set procedure does not find any statistical difference among them when the loss function is computed with respect to the true integrated variance. Thus, we can infer that while the noise cancelling accuracy consistently shows statistical differences among the considered estimators and under the different scenarios analysed, the estimation accuracy of the non-trivial realized measures do not vary enough to evidence a general over-performance of one of the estimators over the others. The latter finding is unsurprising given the asymptotic equivalence of realized kernel, two and multi-scale realized variance and realized pre-averaged variance.

\footnote{The results are broadly the same for all the DGP analysed with minor variation which in few cases lead to differences in the MCS results. Results for the other models are available upon request.}
Table 4 Estimated average standardized volatility over 10,000 Simulations

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>$\lambda = 23400$</th>
<th>$\lambda = 7800$</th>
<th>$\lambda = 1560$</th>
<th>$\lambda = 390$</th>
<th>$\lambda = 78$</th>
<th>$\lambda = 39$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ID</td>
<td>MA(1)</td>
<td>ID</td>
<td>MA(1)</td>
<td>ID</td>
<td>MA(1)</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>1.012</td>
<td>1.027</td>
<td>1.012</td>
<td>1.017</td>
<td>1.01</td>
<td>1.011</td>
<td>1.008</td>
</tr>
<tr>
<td>RPA:</td>
<td>$\gamma_2$</td>
<td>0.999</td>
<td>1.002</td>
<td>0.999</td>
<td>1.002</td>
<td>0.999</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>$\gamma_3$</td>
<td>0.996</td>
<td>0.998</td>
<td>0.996</td>
<td>0.998</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.007</td>
<td>1.009</td>
<td>1.007</td>
<td>1.009</td>
<td>1.003</td>
<td>1.004</td>
</tr>
<tr>
<td>RK:</td>
<td>$\gamma_4$</td>
<td>0.999</td>
<td>3.224</td>
<td>0.998</td>
<td>3.214</td>
<td>0.996</td>
<td>3.169</td>
</tr>
<tr>
<td></td>
<td>$\gamma_5$</td>
<td>0.997</td>
<td>1.24</td>
<td>0.998</td>
<td>1.238</td>
<td>0.997</td>
<td>1.028</td>
</tr>
<tr>
<td>RVTS:</td>
<td>$\gamma_6$</td>
<td>0.998</td>
<td>2.163</td>
<td>0.997</td>
<td>2.156</td>
<td>0.997</td>
<td>1.186</td>
</tr>
<tr>
<td></td>
<td>$\gamma_7$</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>$\gamma_8$</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td>RVMS:</td>
<td>$\gamma_9$</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{10}$</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{11}$</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td>RV5:</td>
<td>$\gamma_{12}$</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.997</td>
<td>0.996</td>
<td>0.996</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{13}$</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.998</td>
<td>0.997</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{14}$</td>
<td>1.19</td>
<td>1.19</td>
<td>1.19</td>
<td>1.19</td>
<td>1.19</td>
<td>1.19</td>
</tr>
<tr>
<td>RV:</td>
<td>$\gamma_{15}$</td>
<td>63.352</td>
<td>33.011</td>
<td>63.186</td>
<td>33.006</td>
<td>21.777</td>
<td>18.411</td>
</tr>
<tr>
<td></td>
<td>$\gamma_{16}$</td>
<td>12.486</td>
<td>6.955</td>
<td>12.455</td>
<td>6.954</td>
<td>4.826</td>
<td>4.213</td>
</tr>
</tbody>
</table>

The table reports the average daily standardized volatility ($\hat{IV}$) for DGP(6) computed for different trading intensities and different level of noises. The different colors help comparing the results for each level of noise intensity.

5 Empirical Application

The dataset of futures used consists of transaction data recorded over a 4 months period (from July 1st - October 31st 2013), corresponding to 87 trading days and about 2,000 trading hours for four different futures contracts traded on the Chicago Mercantile Exchange’s Globex platform: the E-mini S&P 500 (ES), the E-mini NASDAQ 100 (NQ), the E-mini Dow Jones (YM) and the E-mini S&P MidCap 400 (EMD). Additionally, for each instrument two different expirations have been taken into account, September and December.

E-minis futures contracts represent a portion of a standard futures contract, allowing to buy and sell a fraction of the cash value of the underlying index. The small size and low margin rates makes the E-minis contracts extremely liquid and optimal for portfolio hedging. Additionally, the round-the-clock trading hours minimize the loss of information associated with periodic closure. Hence, the high liquidity registered and high daily average trading volume represent appealing characteristics for an econometric analysis, providing an ideal situation to gain valuable insights from high-frequency data. However, the high amount of uninformed traders makes these contracts more prone to fluctuations due to a relatively high volume of noise trades (SEC & CFTC, 2010).

To proxy the underlying indices of the E-minis four highly traded ETFs exchanged on the NYSE have been used. The choice has mainly been based on the high liquidity of the proxies and the absence of leverage. Based on these requirements we have identified the SPY as proxy for the S&P 500, the QQQ as proxy for the NASDAQ, the DIA as proxy for the Dow Jones and finally the MDY as proxy for the S&P MidCap 400. While using ultra-high frequency data it comes natural to
opt for transaction time sampling, we decided to aggregate the data to avoid considering multiple observations recorded at the same time; in which case an average of the recorded prices weighted by the traded volumes is taken. One peculiar characteristic of futures contract is that they expire at a specific date and at each point in time multiple futures contract written on the same underlying but with different maturity dates are available. It is common practice to create a continuous time series by “chaining” multiple contracts together following the criterion of selecting at each point in time the contract with the highest liquidity. Such contract is usually the one closest to maturity (call it $F_1$) which then gets replaced by the following in order of expiration ($F_2$) when the maturity date approaches. Specifically, a roll-over window of one week prior to the maturity date is generally considered and, within such window, as soon as the number of observations of $F_2$ is higher than the number of observations for $F_1$, the series gets rolled over. For the core part of our empirical analysis we will not roll-over our futures as it is of interest to investigate the equality between the integrated variance of futures contracts with different expiration. We will, however, apply the described roll-over scheme to create a continuous time series in Section 5.3 as a comparison between futures contracts would not provide any additional insight.

5.1 Filtering

It is important to remark that a cautious approach is required when working with the dataset considered as the properties of the data vary widely both across different contracts and, within each series, across the sample period. One important difference is given by the different states of liquidity characterizing each futures during the examined time frame. The main reason is that these futures contracts are primarily used for hedging purposes and investors tend to roll-over their hedging from the contract closest to maturity to the next one when the former is close to expiration. Hence, the number of daily transactions (and consequently the time between transactions) is subject to a great deal of variability.
Figure 4. Histogram of the average traded volume and number of transactions

Note: The figure shows a trading day for the E-mini Dow Jones with December expiration. Each bar represents a 30 minutes interval with the top graph showing the number of transactions occurred and the bottom graph showing the cumulative traded volume per interval.

To avoid spurious results caused by liquidity asymmetries the analysed data have been divided into different subsamples according to the average transaction per minute; this will help us discern the diverse empirical performance of each estimator under distinct conditions. Further, in order to remove the effects of intraday seasonality, the considered time frame has been shrunk to consider only the most liquid hours (from 08.30 to 15.00) and allow for a perfect comparison with the trading hours considered for the underlying assets.\(^{19}\)

Figure 4 shows how the trading activity is mainly concentrated during daytime. In fact, while the night period still counts thousands of transactions, its relative amount is extremely lower (especially in terms of traded volume) than the one registered during daily hours.

Among the considered futures ES and NQ represent the most liquid stocks with, respectively, an average of about 400 and 136 trades per minute for the contracts with September expiration and 250 and 231 trades per minute for the December maturity. In terms of liquidity YM follows with an average of 95 and 143 trades per minute for the September and December maturities. While EMD is the less liquid futures with an average of 33 and 36 observations per minute, respectively

\(^{19}\)The standard trading session on the NYSE starts at 9.30 and ends at 16.00 but we need to account for the one hour time-shift between Chicago and New York.
for the September and December contracts. Table 5 and 6 report few key summary statistics of the considered dataset. Together with the exhibited difference in liquidity, the most striking feature is the marked variation in the estimated average daily noise variance, computed as a modified version of the noise variance estimator introduced by Bandi and Russell (2008),\(^{20}\) which tends to be higher for the contracts with December expiration. A possible explanation might be given by the wide liquidity difference across the period of interest. In fact, the series with December maturity tend to be less traded until an earlier maturity contract is available on the market and they experience a rapid increase in the number of transactions when the contract with September maturity is about to expire. Hence, even though only the days with at least 4680 observations per day have been selected, a residual effect from the different liquidity might still be present.

### Table 5 Futures Data: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>SP</th>
<th>NQ</th>
<th>YM</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. time between trades (sec.)</td>
<td>0.15</td>
<td>0.24</td>
<td>0.44</td>
<td>0.26</td>
</tr>
<tr>
<td>Avg. number of trades</td>
<td>268,352</td>
<td>303,584</td>
<td>86,193</td>
<td>109,665</td>
</tr>
<tr>
<td>Max number of trades</td>
<td>510,497</td>
<td>531,992</td>
<td>163,069</td>
<td>212,007</td>
</tr>
<tr>
<td>Min number of trades</td>
<td>8,520</td>
<td>7,174</td>
<td>7,043</td>
<td>15,357</td>
</tr>
<tr>
<td>Avg. Noise Variance ($e^{-7}$)</td>
<td>0.482</td>
<td>0.547</td>
<td>0.523</td>
<td>3.851</td>
</tr>
<tr>
<td>Number of observations</td>
<td>14,759,380</td>
<td>11,153,203</td>
<td>4,568,248</td>
<td>3,947,955</td>
</tr>
</tbody>
</table>

### Table 6 ETF Data: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>SPY</th>
<th>QQQ</th>
<th>DIA</th>
<th>MDY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. time between trades (sec.)</td>
<td>0.21</td>
<td>0.55</td>
<td>2.12</td>
<td>3.24</td>
</tr>
<tr>
<td>Avg. number of trades</td>
<td>118,875</td>
<td>46,638</td>
<td>12,230</td>
<td>9,397</td>
</tr>
<tr>
<td>Max number of trades</td>
<td>224,258</td>
<td>134,503</td>
<td>22,955</td>
<td>83,226</td>
</tr>
<tr>
<td>Min number of trades</td>
<td>59,300</td>
<td>25,855</td>
<td>5,821</td>
<td>4,683</td>
</tr>
<tr>
<td>Avg. Noise Variance ($e^{-7}$)</td>
<td>0.419</td>
<td>0.620</td>
<td>0.380</td>
<td>0.677</td>
</tr>
<tr>
<td>Number of observations</td>
<td>10,104,370</td>
<td>3,964,233</td>
<td>1,039,575</td>
<td>686,045</td>
</tr>
</tbody>
</table>

Before proceeding with any estimation the raw data have been filtered for outliers. The first step consisted in removing all the transactions recorded between 3:15:01pm and 3:29:59pm or between 4:15:01 and 4:59:59pm as they fall outside the official trading time. Then, obviously misrecorded transactions are removed before applying a final, more elaborated filtering algorithm.\(^{21}\) As final step of the cleaning procedure we apply a modified version of the methodology proposed by Brownlees and Gallo (2006) which determines the validity of each transaction according to its likelihood with respect to the statistical properties of the series. Specifically, for each element the

---


\(^{21}\)These are, for example, observations with a bid, ask or transaction price equal to zero or entries marked as corrected trades ($\text{CORR} \neq 0$). A detailed procedure can be found in Barndorff-Nielsen et al. (2009).
trimmed mean of the neighbouring observations is computed and if the considered observations is more than 3 standard deviations away from the mean it gets marked as outlier and discarded. While being heuristic, this procedure shows desirable characteristics as statistical consistency and being accurately adaptable to every financial instrument according both to its daily volume of transactions and tick size. Further, the procedure has been modified to take into account the number of observations over fixed intervals (each hour) and adapts the length (k) of the filtering window accordingly. The reason is that less liquid series tend to present cluster of transactions which dramatically increase the frequency of observations and can be preceded and followed by long gaps without any transactions. The applied procedure takes into account this feature and adds the possibility to apply intra-daily corrections to the filtering window based on the trading intensity.

5.2 Difference in QV

As discussed in Section 2.3.1, we aim to verify the equality presented in Equation 11. To this end we provide day to day confidence intervals of the difference between the realized volatility of futures and their corresponding underlying as well as a simple test on the average difference across days. In this section we rely closely on the results derived in Kinnebrock and Podolskij (2008), which show that difference of realized volatilities consistently estimates difference in quadratic variations.\(^{22}\) Hence, defined the difference in quadratic variation as \([\Delta]_t = [F]_t - [S]_t\), we know that \([\Delta]_t\) can be consistently estimated as the difference of the underlying and futures realized variances \([\Delta]_t = RV(F)_t - RV(S)_t\). To avoid spurious results due to microstructure noise, we work with realized variances estimated on data sampled every 5 minutes. Additionally, averaging each return across days we can obtain a new measure ([\(\bar{\Delta}\)]_t) with the same distribution. This allows us to compute the average difference over the dates of interest and construct a simple test to assess the equality of the two measures for the whole period

\[
\frac{[\Delta]_t - [\Delta_0]_t}{SE([\Delta]_t)} \sim \mathcal{N}(0,1).
\]

Particularly we test \(H_0 : [\Delta]_t = 0\) against the alternative \(H_1 : [\Delta]_t \neq 0\).

Further, we can trivially construct a two-sided \((1-\alpha)\) confidence interval by plugging an estimator for the asymptotic variance of the difference. For each stock-futures pair we compute the daily and average differences, and 95% confidence interval of the form

\[
(RV(F)_t - RV(S)_t) \pm z_{1-\alpha} \sqrt{\frac{n}{n} \sum_{i=1}^{n} (r(F)_{i,t} - r(S)_{i,t})^2 - \frac{1}{n} (RV(F)_t - RV(S)_t)^2}
\]

where \(r\) refers to log-returns and \(n\) is the number of daily returns.

\(^{22}\)See, Theorem 2 in Kinnebrock and Podolskij (2008)
Table 7 Summary statistics: Difference in QV

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>min</th>
<th>median</th>
<th>max</th>
<th>0 $\not\in$ CI&lt;sup&gt;95%&lt;/sup&gt;</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES–SPY</td>
<td>-0.0549</td>
<td>-0.0746</td>
<td>-0.0449</td>
<td>0.0006</td>
<td>3.66%</td>
<td>1.0000</td>
</tr>
<tr>
<td>NQ–QQQ</td>
<td>-0.1068</td>
<td>-0.4101</td>
<td>-0.1010</td>
<td>0.0654</td>
<td>20.73%</td>
<td>1.0000</td>
</tr>
<tr>
<td>YM–DIA</td>
<td>-0.0631</td>
<td>-0.4137</td>
<td>-0.0503</td>
<td>0.0287</td>
<td>9.76%</td>
<td>1.0000</td>
</tr>
<tr>
<td>EMD–MDY</td>
<td>-0.1389</td>
<td>-1.4331</td>
<td>-0.1008</td>
<td>0.0139</td>
<td>12.86%</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

The table reports summary statistics of $RV(F)_t - RV(S)_t$. The last column reports the percentage of days in which the 95% confidence interval does not contain 0. The results are expressed as percentages. The reported P-val refers to the test conducted on $|\bar{\Delta}_t|$.

From Table 7 we see that the differences tend to be close to zero and, although some of the reported deviations have an important size, they are generally not statistically significant as shown by the extremely wide confidence intervals reported in Figure 5. The p-values reported in Table 7, refers to the test conducted on the average daily difference measure $|\bar{\Delta}_t|$; and show that our investigation strongly favour the equality theorized in Equation 11. However, it is extremely interesting to notice that, while daily deviations from the equality seems to occur, our data do not hint to $[F]_t > [S]_t$, which could imply a non-negligible impact of the stochastic interest rate. Conversely, looking both at the mean and max values, we can infer that the realized variation estimated for the underlying seems, generally, to be bigger than the corresponding estimate for the futures. Overall, the estimated difference in QV is not significantly different from zero and only occasionally presents small but significant deviations. Such days clearly do not represent the norm and, due to the unstable nature of the estimator, might be produced by residual microstructure noise or relevant differences in the liquidity of the two series.
5.3 Local Equivalence

While Section 5.2 explores the relationship between the two series from an aggregated perspective, here we employ the non-parametric adaptive estimation procedure introduced by Li et al. (2017) to obtain local estimates of the slope coefficient ($\beta$) for the following linear regression:

$$S^c_t = \beta F^c_t + \varepsilon_t, \quad t \in [0, 1].$$

(20)

Where $F^c_t$ and $S^c_t$ are two continuous local martingales and $\varepsilon_t$ identifies the residual component, with $[S, \varepsilon] = 0$.\(^{23}\)

This is of particular interest given the short-rate process introduced in subsubsection 4.2.1 (whose effect turned out to be negligible in our simulation exercise) as a non-significant effect would entail $S^c_t = F^c_t$ and hence a $\beta = 1$ in Equation 20. Hence, we can now empirically verify the impact of

---

\(^{23}\)In the simplified framework presented in Section 2 we ruled out the presence of jumps from our price processes. Such assumption is here abandoned and a finite activity jump process is permitted.
stochastic interest rates on the quadratic variation of the futures.\footnote{More generally, we investigate the presence of additional stochastic components that might lead to deviations from Equation 11.}

\textbf{Figure 6.} Scatter plot of ES-SPY 5 - min. returns

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{scatter_plot}
\caption{Scatter plot of ES-SPY 5 - min. returns (\textbullet{}) and “problematic” returns (\texttimes{}). The blue line shows the linear fit with the average estimated beta.}
\end{figure}

Unfortunately the pronounced illiquidity of some of the series might have a serious detrimental effect on our analysis due to the significant number of zero returns.\footnote{The absence of observations for periods spanning more than the chosen sampling frequency will lead to zero returns and to reduced covariance (and consequently beta) estimates between the two series.} As the issue is not severe for the E-mini S\&P 500 series we report results using the 5 minutes frequency for the ES-SPY pair while we reduce our sampling frequency to 25 minutes in order to attenuate the problem for the other futures-underlying pairs. \textbf{Figure 6} displays the result for the ES-SPY pair at the 5 minutes sampling frequency. The average estimated beta is slightly below unity ($\hat{\beta} = 0.8901$) however the confidence interval includes one, suggesting that there is no statistical evidence of differences between the continuous quadratic variations of the two time series and, consequently, supporting the negligibility argument of the bias caused by stochastic interest rates in favour of Equation 11. The figure provides a clear perception of the linear correlation between the two series and further evidences the distortions caused by lack of liquidity in one of the two series with the green crosses.
marking the intervals upon which only one of the two series registered a zero return.

**Table 8** Adaptive beta summary results

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}$</th>
<th>95% CI</th>
<th>$1 \in \text{CI}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPY-ES</td>
<td>0.7897</td>
<td>[0.5059, 1.0734]</td>
<td>82.22%</td>
</tr>
<tr>
<td>QQQ-NQ</td>
<td>0.9590</td>
<td>[0.8208, 1.0972]</td>
<td>98.90%</td>
</tr>
<tr>
<td>DIA-YM</td>
<td>0.9134</td>
<td>[0.7511, 1.0757]</td>
<td>78.26%</td>
</tr>
<tr>
<td>MDY-EMD</td>
<td>0.8843</td>
<td>[0.6968, 1.0718]</td>
<td>73.26%</td>
</tr>
</tbody>
</table>

The table reports summary statistics for $S_t^c = \beta F_t^c + \epsilon_t$ computed using returns constructed at the 25 minutes frequency. The last column reports the percentage of times in which the confidence interval contains 1. The results are expressed as percentages.

Similar results are reported in Table 8 which shows the estimates at the 25 minutes frequency. All the series have average estimated betas below unity but all the confidence intervals include one. Finally, we can see that the percentage of times 1 is included in the confidence interval is quite high for all the series, indicating only minor deviations of the local beta estimates. Overall, the local level analysis seems to support Equation 11 and our conclusions in subsubsection 4.2.1. Further, while we are aware that liquidity issues might affect the results, we can confidently expect improved results in case such returns were to be excluded as all the provided scatter plot hint to a strong linear relation between the returns of each pair with slope very close to one.
Figure 7. Scatter plot of continuous returns for each futures – underlying pair

Note: Scatter plot of 5-minutes returns (∗) and “problematic” returns (×). The blue line shows the linear fit with the average estimated beta.

5.4 The Noise structure

We assess the empirical characteristics of the noise affecting the data through two different methodologies. The analysis of the ACF and PACF functions for intraday returns, which provides us information on the autocorrelation of the noise and the average noise-to-signal ratio ($\lambda$) and average Noise Ratio ($\gamma$) to assess the magnitude of the noise variance relatively to the true integrated variance.

To get an idea of the size of the noise compared to the actual integrated variance we proceed to estimate the noise-to-signal ratio. Following (Hansen and Lunde, 2006a), we compute: $\lambda = \frac{\bar{\omega}^2}{\bar{IV}}$, where $\bar{\omega}^2 = n^{-1} \sum_{t=1}^{n} \tilde{\omega}_t^2$ and $\bar{IV} = n^{-1} \sum_{t=1}^{n} \tilde{IV}_t$. Clearly, as neither the noise variance nor the IV can reasonably be assumed as constant across days, $\hat{\lambda}$ can only be interpreted as a proxy for the noise-to-signal ratio. Interestingly, from Table 9 we can observe that $\hat{\lambda}$ tends to be very similar across series even though Table 5 presented very different values for $\hat{\omega}^2$. This implies that the average estimated integrated variance is higher for the contracts with December maturity, possibly due to a period of higher volatility pushing down the ratio which occurred after the contracts with September maturity reached expiration.
Table 9 Noise-to-Signal \( (e^{-2}) \) and Noise Ratio

<table>
<thead>
<tr>
<th>SPY</th>
<th>SP</th>
<th>QQQ</th>
<th>NQ</th>
<th>DIA</th>
<th>YM</th>
<th>MDY</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. noise-to-signal</td>
<td>0.18</td>
<td>0.22</td>
<td>0.18</td>
<td>0.17</td>
<td>0.18</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td>Max noise-to-signal</td>
<td>0.28</td>
<td>0.71</td>
<td>0.36</td>
<td>0.24</td>
<td>0.24</td>
<td>0.30</td>
<td>0.26</td>
</tr>
<tr>
<td>Min noise-to-signal</td>
<td>0.09</td>
<td>0.08</td>
<td>0.11</td>
<td>0.11</td>
<td>0.13</td>
<td>0.09</td>
<td>0.13</td>
</tr>
<tr>
<td>Avg. Noise Ratio</td>
<td>0.38</td>
<td>1.41</td>
<td>1.15</td>
<td>0.48</td>
<td>0.42</td>
<td>0.39</td>
<td>0.31</td>
</tr>
<tr>
<td>Max Noise Ratio</td>
<td>0.8</td>
<td>2.36</td>
<td>2.45</td>
<td>0.99</td>
<td>0.77</td>
<td>0.62</td>
<td>0.82</td>
</tr>
<tr>
<td>Min Noise Ratio</td>
<td>0.03</td>
<td>0.62</td>
<td>0.19</td>
<td>0.04</td>
<td>0.03</td>
<td>0.10</td>
<td>0.03</td>
</tr>
</tbody>
</table>

The table presents the daily average, min and max values of the estimated Noise-to-signal ratio \( \hat{\lambda} \) and Noise Ratio \( \hat{\gamma} \).

Another interesting finding comes from analysing the Noise Ratio \( \gamma = \sqrt{\hat{\lambda} \cdot \hat{m}} \) introduced by Oomen (2006). As discussed in Christensen et al. (2010), the parameter evaluates the ratio between market microstructure noise and the integrated variance component; strong of a direct relation to the bias of the realized variance \( \mathbb{E}(RV) = IV(1 + 2\gamma^2) \). Therefore, we can identify the level of noise to which each series is subject. On this aspect strikes the relative size of the estimated \( \gamma \) for the E-Mini S&P 500 compared to the other series considered, showing that both the September and December series are subject to markedly higher level of noise. Additionally, we consider table 9 of particular interest as it seems to provide empirical evidence in support of the noise structure assumed in 2.2. In fact, from the reported noise ratios, we can see how futures with different maturity but written on the same underlying are affected by very similar level of noise but different from the level estimated for the underlying asset. This is compatible with the theory of a second source of noise which affects the futures as in (12).

Figures 8 and 9 show the daily average ACF and PACF. A simple eye balling of the graphs allows us to find empirical evidence against the independent noise assumption. In fact, all the series analysed show some degree of correlation with wide variations across different series. The only exception is given by the EMD futures but this absence of correlation can reasonably be explained by the low liquidity recorded, which is especially marked for the contract with December expiration. In fact, given that the series are analysed in transaction time, a low liquid series will tend to have observations distant in time from each other and, consequently, will exhibit a lower degree of correlation (assumed that the series is stationary). Further, the higher degree of autocorrelation exhibited by both ES series relates back to the higher noise ratio evidenced in 9.
Figure 8. ACF and PACF for S&P 500 and NASDAQ 100

Note: The figure shows the average ACF (○○○) and PACF (×××) of two different underlying (SPY and QQQ) and the relative Futures contracts.
Figure 9. ACF and PACF for Dow Jones and S&P MidCap 400

Note: The figure shows the average ACF (○○○○) and PACF (××××) of two different underlying (DIA and MDY) and the relative Futures contracts.
5.5 Findings

Tables 10 to 12 present the estimated average daily volatility for each series and every estimator considered. A quick look at the results shows how our theoretical assumptions are further validated by the empirical findings. We can see how, for each estimator we obtain fairly close estimates of the integrated variance between the futures series and the ETF used as proxy for the underlying index. As expected, we observe different precision levels across different estimators with the simple realized variance clearly reporting widely different estimates. Tables 10, 11 and 12 present the average daily estimates for the different realized measures considered. While Table 12 consider the period subsequent the expiration of the September contract, Tables 10 and 11 refer to the period 01/07 - 20/09. Hence, in the latter case both the September and December average values are reported. However, due to the restriction on the minimum number of daily observations\(^{26}\), the actual number of days used to compute the average value for the December expiration series is very low (approximately less than 10 days are used) and concentrates over the last days of the considered period explains the difference in the reported values.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RPA:</td>
<td>0.4825</td>
<td>0.4920</td>
<td>0.5449</td>
<td>0.5616</td>
<td>0.5643</td>
<td>0.5387</td>
</tr>
<tr>
<td>RK:</td>
<td>0.4828</td>
<td>0.4919</td>
<td>0.5455</td>
<td>0.5630</td>
<td>0.5662</td>
<td>0.5375</td>
</tr>
<tr>
<td>RVTS:</td>
<td>0.4740</td>
<td>0.5016</td>
<td>0.5311</td>
<td>0.5363</td>
<td>0.5679</td>
<td>0.5325</td>
</tr>
<tr>
<td>RVMS:</td>
<td>0.4755</td>
<td>0.5098</td>
<td>0.5385</td>
<td>0.5494</td>
<td>0.5676</td>
<td>0.5253</td>
</tr>
<tr>
<td>RV5:</td>
<td>0.4656</td>
<td>0.4750</td>
<td>0.5519</td>
<td>0.5501</td>
<td>0.5466</td>
<td>0.5071</td>
</tr>
<tr>
<td>RV:</td>
<td>0.4780</td>
<td>0.9731</td>
<td>0.7701</td>
<td>0.6440</td>
<td>0.6184</td>
<td>0.5750</td>
</tr>
</tbody>
</table>

The table reports the average daily volatility for the S&P 500 and NASDAQ estimated over the period 01-July to 20-Sept. with the six different estimators considered.

\(^{26}\)As discussed in section 5.1 only days with more than 4680 observations are used.
Table 11  Estimated average daily volatility (%) July - September

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RPA:</td>
<td>0.4559</td>
<td>0.4592</td>
<td>0.4731</td>
<td>0.6334</td>
<td>0.6421</td>
<td>0.5574</td>
</tr>
<tr>
<td>RK:</td>
<td>0.4559</td>
<td>0.4592</td>
<td>0.4754</td>
<td>0.6302</td>
<td>0.6424</td>
<td>0.5541</td>
</tr>
<tr>
<td>RVTS:</td>
<td>0.4517</td>
<td>0.4707</td>
<td>0.4301</td>
<td>0.6111</td>
<td>0.6347</td>
<td>0.5749</td>
</tr>
<tr>
<td>RVMS:</td>
<td>0.4557</td>
<td>0.4713</td>
<td>0.4341</td>
<td>0.6192</td>
<td>0.6376</td>
<td>0.5651</td>
</tr>
<tr>
<td>RV5:</td>
<td>0.4468</td>
<td>0.4511</td>
<td>0.4494</td>
<td>0.6250</td>
<td>0.6246</td>
<td>0.5488</td>
</tr>
<tr>
<td>RV:</td>
<td>0.4410</td>
<td>0.5076</td>
<td>0.4545</td>
<td>0.5988</td>
<td>0.6302</td>
<td>0.5686</td>
</tr>
</tbody>
</table>

The table reports the average daily volatility for the S&P MidCap 400 and Dow Jones estimated over the period 01-July to 20-Sept. with the six different estimators considered.

Figure 10 offers a visual explanation of the problem as it shows the time series of the daily estimates across the whole sample period. A quick eye balling of the figure further allows to appreciate the precision of the estimators (the closer are the dots to each other the best the performance of the estimator). Of particular interest is the different performance registered by the realized variance, for different series. In line with the estimated values of $\gamma$ and $\lambda$, RV shows the worst performance on the E-mini S&P 500 series (the ones affected by a higher degree of microstructure noise) while behaving somehow better for the other series. \(^{27}\)

Table 12  Estimated average daily volatility (%) September - October

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RPA:</td>
<td>0.5368</td>
<td>0.5426</td>
<td>0.6478</td>
<td>0.6410</td>
<td>0.5002</td>
<td>0.5019</td>
<td>0.6350</td>
<td>0.6602</td>
</tr>
<tr>
<td>RK:</td>
<td>0.5376</td>
<td>0.5436</td>
<td>0.6494</td>
<td>0.6439</td>
<td>0.4997</td>
<td>0.5044</td>
<td>0.6374</td>
<td>0.6616</td>
</tr>
<tr>
<td>RVTS:</td>
<td>0.5270</td>
<td>0.5502</td>
<td>0.6140</td>
<td>0.6365</td>
<td>0.4984</td>
<td>0.5076</td>
<td>0.6461</td>
<td>0.6572</td>
</tr>
<tr>
<td>RVMS:</td>
<td>0.5299</td>
<td>0.5593</td>
<td>0.6323</td>
<td>0.6422</td>
<td>0.5034</td>
<td>0.5126</td>
<td>0.6503</td>
<td>0.6626</td>
</tr>
<tr>
<td>RV5:</td>
<td>0.5179</td>
<td>0.5374</td>
<td>0.6377</td>
<td>0.6238</td>
<td>0.4881</td>
<td>0.4905</td>
<td>0.6387</td>
<td>0.6315</td>
</tr>
<tr>
<td>RV:</td>
<td>0.5155</td>
<td>0.9333</td>
<td>0.6896</td>
<td>0.6602</td>
<td>0.4721</td>
<td>0.5318</td>
<td>0.6270</td>
<td>0.6618</td>
</tr>
</tbody>
</table>

The table reports the average daily volatility estimated over the period 21-Sept.to 20-Oct. with the six different estimators considered.

\(^{27}\)The same fact can be observed for the E-mini Dow series, which, as evidenced by the reported noise ratio on table 9, are affected by a higher degree of noise compared to the considered proxy (DIA).
Figure 10. Time series plot of the estimated Realized Measures

Note: The figure shows the Time Series of daily realized measures estimated. The top figure displays the (SPY), the (ES Sept.) and the (ES Dec.). While the bottom figure reports the (QQQ), the (NQ Sept.) and the (NQ Dec.)
Given the clear indication of autocorrelation in the data indicated by the ACF and PACF analysis, we opted for a block bootstrap in the model confidence set procedure to be able to replicate this feature. Different loss functions have been tested. However, as expected, the QLIKE loss function demonstrated more power in rejecting less performing estimators, narrowing more the optimal set of elements compared to the other loss functions tested.

Table 13 reports the included estimators and the elimination order for both the $T_R$ and $T_{\text{max}}$ tests with 10,000 bootstrap replications at a 95% confidence level. Despite the short sample period considered, the data are informative enough to allow the model confidence set procedure to adequately restrict the set of optimal elements. While the results changes according to the series analysed, we can see that (in line with the results obtained in simulation) both the Realized Kernel and the Pre-averaged Realized Variance are consistently selected as superior estimators. However, due to the remarkable low liquidity observed, both EMD series and the YM December series do not allow the MCS to properly determine any difference in the performance of the estimates; consequently, the procedure acts conservatively without restricting the set of superior models as much as expected.

Table 13  MCS Ranking

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_R$</td>
<td>$T_R$</td>
<td>$T_R$</td>
<td>$T_R$</td>
<td>$T_R$</td>
<td>$T_R$</td>
<td>$T_R$</td>
<td>$T_R$</td>
</tr>
<tr>
<td>RPA:</td>
<td>$\mathcal{M}_4$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\mathcal{M}_4$</td>
<td>$\mathcal{M}_2$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>RK:</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>RVTS:</td>
<td>$\mathcal{M}_2$</td>
<td>$\mathcal{M}_2$</td>
<td>$\mathcal{M}_0$</td>
<td>$\mathcal{M}_1$</td>
<td>$\mathcal{M}_1$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>RVMS:</td>
<td>$\mathcal{M}_1$</td>
<td>$\mathcal{M}_1$</td>
<td>$\mathcal{M}_3$</td>
<td>$\mathcal{M}_3$</td>
<td>$\mathcal{M}_2$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>RV5:</td>
<td>$\mathcal{M}_3$</td>
<td>$\mathcal{M}_3$</td>
<td>$\mathcal{M}_2$</td>
<td>$\mathcal{M}_2$</td>
<td>$\mathcal{M}_3$</td>
<td>$\mathcal{M}_1$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
<tr>
<td>RV:</td>
<td>$\mathcal{M}_0$</td>
<td>$\mathcal{M}_0$</td>
<td>$\mathcal{M}_1$</td>
<td>$\mathcal{M}_0$</td>
<td>$\mathcal{M}_0$</td>
<td>$\times$</td>
<td>$\times$</td>
<td>$\times$</td>
</tr>
</tbody>
</table>

Included estimators ($\times$) and elimination order for the $T_R$ test with 10,000 bootstrap replications at a 95% confidence level.

Table 14 presents sample results of the Diebold-Mariano type test for a representative asset and both its expirations. The test reports the values obtained using the QLIKE loss function. Specifically, for each row, the values display the performance of the estimator relatively to the intersecting estimator on each column. Hence, a positive value implies a superior performance of the row estimator over the estimator on the intersecting column and vice versa. The results are in line with the Model Confidence Set findings, evidencing the consistent over-performance of both pre-averaged variance and realized kernel over the other estimators considered.

---

28 The linex function introduced by Varian (1975) and a standard quadratic loss function have also been tested. The results are available upon request.
29 The complete results are available upon request.
30 Results for different loss functions are available upon request.
<table>
<thead>
<tr>
<th>Dec.</th>
<th>RPA</th>
<th>RK</th>
<th>RVTS</th>
<th>RVMS</th>
<th>RV5</th>
<th>RV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sept.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPA:</td>
<td>0</td>
<td>-3.8743**</td>
<td>9.1409**</td>
<td>8.0964**</td>
<td>1.4858*</td>
<td>11.3561***</td>
</tr>
<tr>
<td>RK:</td>
<td>0.4975</td>
<td>0</td>
<td>9.2017**</td>
<td>8.1672**</td>
<td>1.7077**</td>
<td>11.3535***</td>
</tr>
<tr>
<td>RVTS:</td>
<td>-5.9919***</td>
<td>-6.3310***</td>
<td>0</td>
<td>2.2688**</td>
<td>-1.6758**</td>
<td>11.3158***</td>
</tr>
<tr>
<td>RVMS:</td>
<td>-7.7719***</td>
<td>-8.2971***</td>
<td>-2.2222***</td>
<td>0</td>
<td>-2.3404**</td>
<td>11.3528***</td>
</tr>
<tr>
<td>RV5:</td>
<td>-2.8538***</td>
<td>-2.8763***</td>
<td>0.0848</td>
<td>0.9364</td>
<td>0</td>
<td>10.9825***</td>
</tr>
<tr>
<td>RV:</td>
<td>-7.2875***</td>
<td>-7.2926***</td>
<td>-7.2045***</td>
<td>-7.2518***</td>
<td>-7.0904***</td>
<td>0</td>
</tr>
</tbody>
</table>

The table presents the pairwise comparison test results for a representative asset and both expirations. The upper triangular matrix displays the results for the contract with September expiration while the lower triangular reports the ones for the contract with December expiration.

10%, 5% and 1% significance levels are reported respectively with *, **, ***.

### 6 Conclusions

We consider the log-linear relationship between futures and their underlying assets and show that in absence of arbitrage, in a BSM framework, the two series must share the same integrated variance. However, due to the presence of noise the use of noise robust estimators becomes of primary importance in retrieving the latent parameter. This finding opens up to numerous applications and in this paper we develop one in detail; setting out a methodology, robust to microstructure noise, to test and evaluate the noise cancellation accuracy and performance of several noise-robust estimators of the integrated variance.

We conduct a thorough simulation to analyse the behaviour of the considered estimators under different combinations of noise and trading intensities and, noise structures. Further, six different models have been generated to investigate the effect of different stochastic components and their combinations. The results, point out a better performance of the realized pre-averaged variance over the competing estimators with the realized kernel being typically selected for lower values of $\lambda$. However, in line with the empirical findings, as a stochastic interest rate process is introduced the MCS procedure doesn’t evidence significant differences between the realized kernel and the pre-averaged variance, including both in the set of superior elements for high levels of trading intensity.

The introduction of multiple stochastic components doesn’t seem to affect the results; only a minor variation in the elimination order of lower performing estimators is evidenced.

We finally verify the empirical accuracy over a set of index futures. While we might expect our results to be more precise on samples larger than the restricted set of assets and time span considered (8 assets over a 4 months period) in our empirical application, the results are nonetheless robust and satisfactory. The Model Confidence Set procedure adequately restrict the set of superior models for five of the eight considered contracts, acting conservatively only for the three less liquid series. The Diebold-Mariano type test used to obtain a pairwise comparison further supports the results, showing once again consistent better performances for both the Realized Kernel and the Pre-averaged Variance estimators.
References


<table>
<thead>
<tr>
<th>Year</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017-07</td>
<td>Oskar Knapik: Modeling and forecasting electricity price jumps in the Nord Pool power market</td>
<td>Oskar Knapik</td>
</tr>
<tr>
<td>2017-08</td>
<td>Malene Kallestrup-Lamb and Carsten P.T. Rosenskjold: Insight into the Female Longevity Puzzle: Using Register Data to Analyse Mortality and Cause of Death Behaviour Across Socio-economic Groups</td>
<td>Malene Kallestrup-Lamb and Carsten P.T. Rosenskjold</td>
</tr>
<tr>
<td>2017-10</td>
<td>Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Dynamics of Variance Risk Premia, Investors’ Sentiment and Return Predictability</td>
<td>Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante</td>
</tr>
<tr>
<td>2017-11</td>
<td>Søren Johansen and Morten Nyboe Tabor: Cointegration between trends and their estimators in state space models and CVAR models</td>
<td>Søren Johansen and Morten Nyboe Tabor</td>
</tr>
<tr>
<td>2017-12</td>
<td>Lukasz Gatarek and Søren Johansen: The role of cointegration for optimal hedging with heteroscedastic error term</td>
<td>Lukasz Gatarek and Søren Johansen</td>
</tr>
<tr>
<td>2017-13</td>
<td>Niels S. Grønborg, Asger Lunde, Allan Timmermann and Russ Wermers: Picking Funds with Confidence</td>
<td>Niels S. Grønborg, Asger Lunde, Allan Timmermann and Russ Wermers</td>
</tr>
<tr>
<td>2017-15</td>
<td>Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante: A Non-Structural Investigation of VIX Risk Neutral Density</td>
<td>Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante</td>
</tr>
<tr>
<td>2017-16</td>
<td>Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris: Does the ARFIMA really shift?</td>
<td>Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris</td>
</tr>
<tr>
<td>2017-17</td>
<td>Massimo Franchi and Søren Johansen: Improved inference on cointegrating vectors in the presence of a near unit root using adjusted quantiles</td>
<td>Massimo Franchi and Søren Johansen</td>
</tr>
<tr>
<td>2017-19</td>
<td>Daniel Borup and Martin Thyrgaard: Statistical tests for equal predictive ability across multiple forecasting methods</td>
<td>Daniel Borup and Martin Thyrgaard</td>
</tr>
<tr>
<td>2017-20</td>
<td>Tommaso Proietti and Alessandro Giovannelli: A Durbin-Levinson Regularized Estimator of High Dimensional Autocovariance Matrices</td>
<td>Tommaso Proietti and Alessandro Giovannelli</td>
</tr>
<tr>
<td>2017-21</td>
<td>Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Variance swap payoffs, risk premia and extreme market conditions</td>
<td>Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante</td>
</tr>
<tr>
<td>2017-24</td>
<td>Giorgio Mirone: Inference from the futures: ranking the noise cancelling accuracy of realized measures</td>
<td>Giorgio Mirone</td>
</tr>
</tbody>
</table>