DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

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

Giorgio Mirone

CREATES Research Paper 2017-24

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

Email: oekonomi@au.dk
Tel: +45 8716 5515

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.

May 2017

ABSTRACT

We consider the log-linear relationship between futures contracts and their underlying assets and show that in the classical Brownian semi-martingale ($\mathcal{BSM}$) framework the two series must, by no-arbitrage, have the same integrated variance. We then introduce the concept of noise cancelling 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. Using E-mini index futures contracts, we carry out formal testing of several realized measures in the presence of noise. Moreover, a thorough simulation analysis is employed to evaluate the estimators' sensitivity to different price and noise processes, and sampling frequencies.

JEL classification: C52, C58.

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

*I am grateful to Kim Christensen 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). The author acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

†*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 $\left(\int_0^T \sigma_t^2 \mathrm{d}t\right)$ 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. While 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 we propose a methodology to evaluate and compare estimation errors over 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) a niche literature with a focus on estimation performances has developed presenting methods to obtain consistent rankings using a proxy for the latent variable (see, Patton, 2011; Liu, Patton, and Sheppard, 2015). We add to the existing literature providing a new method with focus on the evaluation of the noise cancelling performances (defined in section 3) which complement the current literature on estimation accuracy of latent variables.

The article is structured as follows: the second section 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 three we set out the details of the methodology to be used in the subsequent analysis. The fourth section 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 five 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 six 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 ($Y \in \mathcal{SM}$) on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}\right)$. 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 ($Y \in \mathcal{BSM}$) of the form:

$$Y_t = Y_0 + \int_0^t a_u \mathrm{d}u + \int_0^t \sigma_u \mathrm{d}W_u, \tag{1}$$

where Y_0 is $\mathcal{F}_0$-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 $Y \in \mathcal{BSM}$ then, the integrated variance (our object of interest) is well defined:

$$IV = \int_0^t \sigma_u^2 \mathrm{d}u. \tag{2}$$

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

$$[Y]_t = \operatorname*{plim}_{n\to\infty} \sum_{i=0}^{t_i \leq t-1} \left(Y_{t_{i+1}} - Y_{t_i}\right)^2. \tag{3}$$

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 example, Andersen, Bollerslev, Diebold, and Labys 2001; Barndorff-Nielsen and Shephard 2002, 2007).

Hence, as $Y \in \mathcal{BSM}$:

$$[Y]_t = \int_0^t \sigma_u^2 \mathrm{d}u. \tag{4}$$

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.

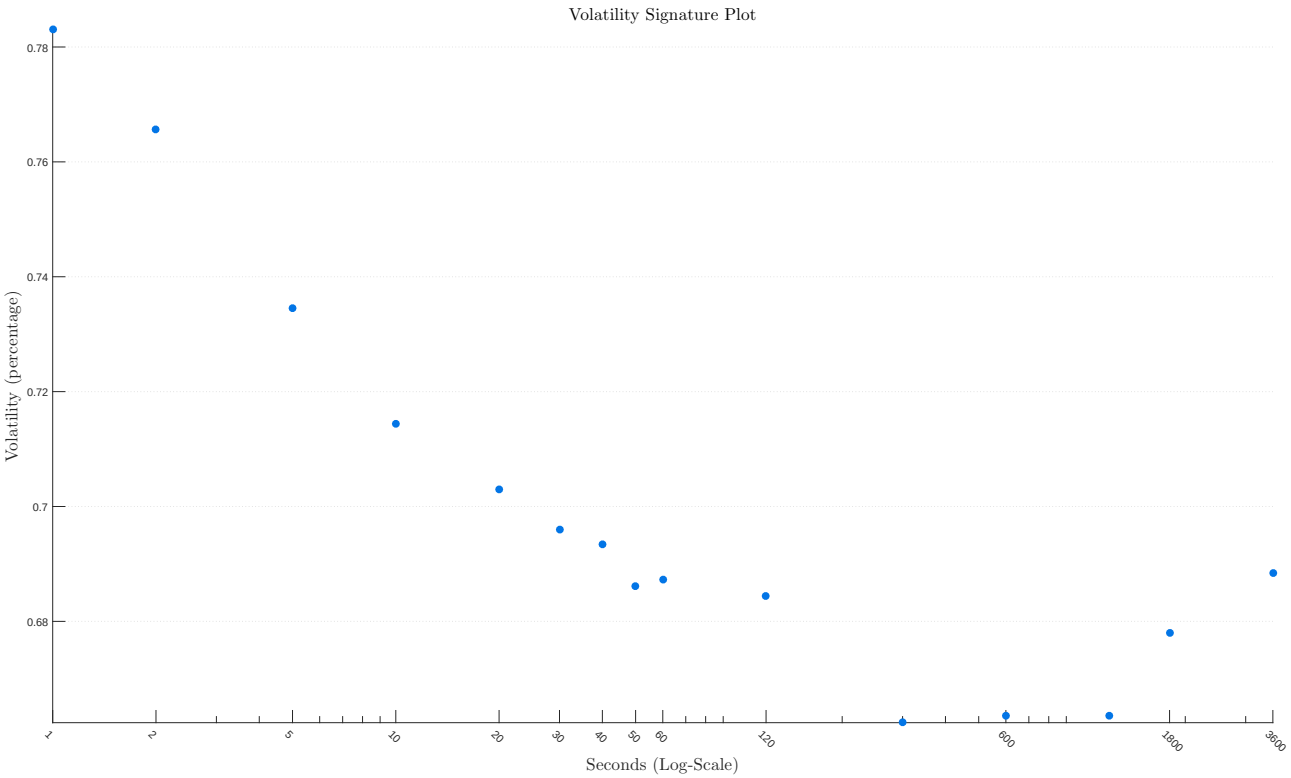

Figure 1. Volatility signature plot for the E-mini Nasdaq 100 with September maturity: as initially shown by (Andersen et al., 1999), the values estimated using the RV estimator explodes as the sampling frequency increases.

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:

$$X_t = Y_t + U_t. \tag{5}$$

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:

$$\mathbb{E}\left(U_t\right) = 0, \quad \mathbb{E}\left(U_t^2\right) = \omega^2, \quad U_t \perp\!\!\!\perp U_s,\, t \neq s. \tag{6}$$

While the white noise assumption is not realistic, 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 ($U \perp\!\!\!\perp Y$). 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 fre-

quencies 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:

$$RV_t^{(m)} \xrightarrow{p} \infty, \text{ as } m \to \infty. \tag{7}$$

Particularly, given (6), we will have:

$$RV_t^{(m)} \simeq 2m\omega^2. \tag{8}$$

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 to verify the accuracy of the obtained estimate. 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:

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

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:

$$P_{t,T} = e^{Y_t + \int_t^T r_u \mathrm{d}u}, \tag{9}$$

where r_t is a deterministic, continuously compounded interest rate[3].
Taking logarithms, we obtain:

$$F_{t,T} = Y_0 + \int_0^t a_u \mathrm{d}u + \int_0^t \sigma_u \mathrm{d}W_u + \int_t^T r_u \mathrm{d}u. \tag{10}$$

Evidently, from (10), we can see that the integrated variance of a futures contract is given by:

$$IV_{F_{t,T}} = \int_0^t \sigma_u^2 \mathrm{d}u.$$

Hence, it directly follows[4]:

$$IV_{F_{t,T}} = IV_{Y_t}. \tag{11}$$

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 $U_t = \omega_1 Z_{1_t}$, 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:

$$V_t = \omega_2 \rho Z_{1_t} + \omega_2 \sqrt{1-\rho^2} Z_{2_t}, \tag{12}$$

where, $Z_{2_t} \sim \mathcal{N}(0,1)$ is correlated with Z_{1_t} with correlation parameter $\rho = \mathrm{corr}(Z_{1_t}, Z_{2_t})$.
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 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_t} and idiosyncratic component Z_{2_t} of the noise term affecting F_t.
We then define:

$$K_{t,T} = F_{t,T} + V_t. \tag{13}$$

Consequently, we will have again:

$$RV_{K,t}^{(m)} \xrightarrow{p} \infty, \text{ as } m \to \infty.$$

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).

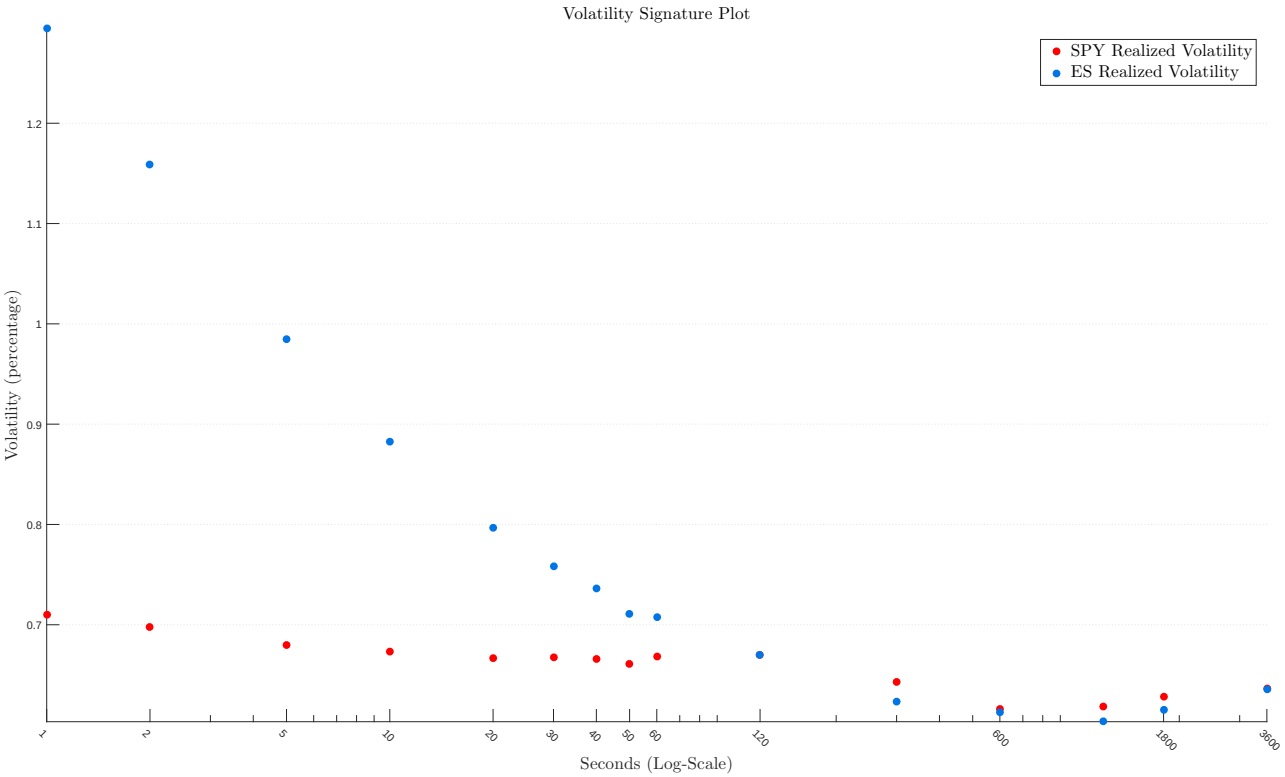

Figure 2. The figure shows the signature plot for the Emini S&P 500 with December expiration together with the underlying asset proxied by the SPY. It can be easily seen how the two series are differently affected by the microstructure noise

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 constant and entirely imputable to the different discount rate used to price the futures (Chen, Cuny, and Haugen, 1995).

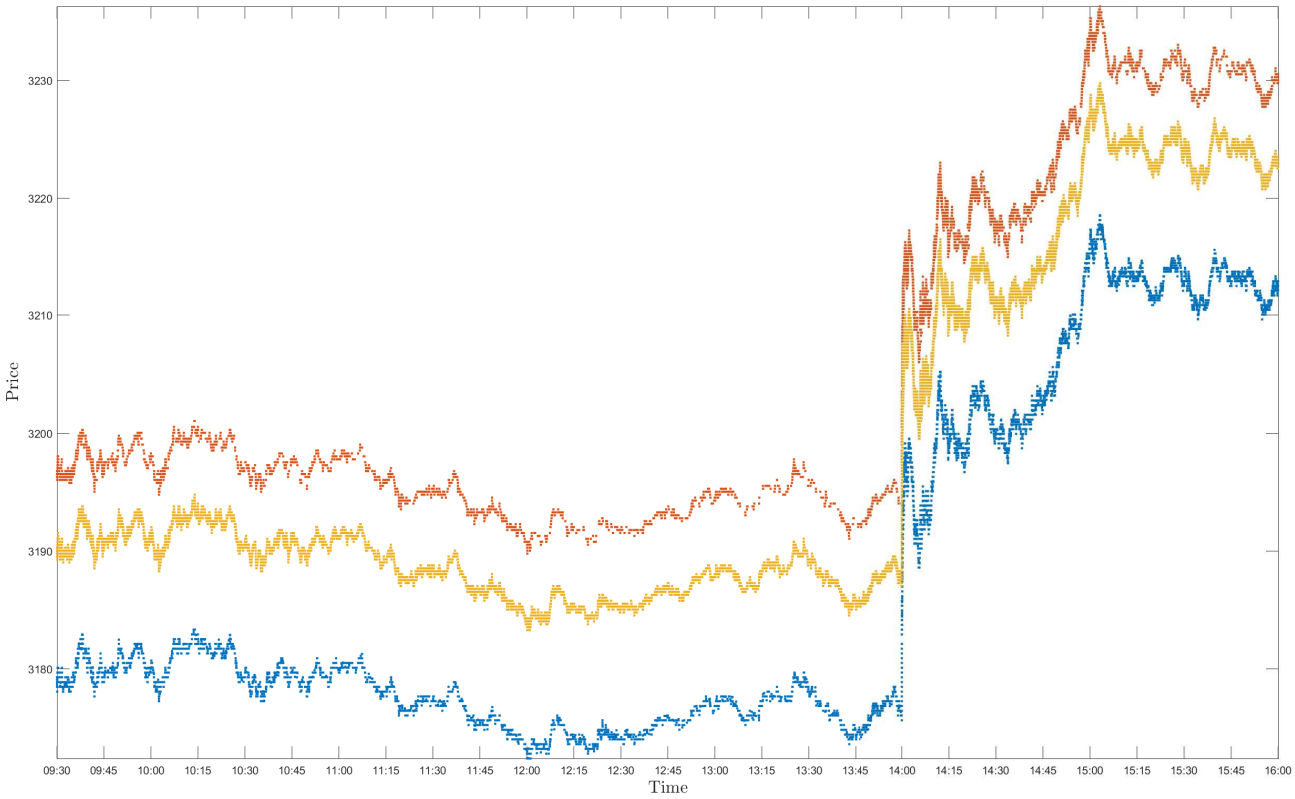

Figure 3. The day trade occurred on September 18 2013 of the QQQ, the NQ Sept. and the NQ Dec. are plotted. It can easily be seen that the price of the underlying and the two futures move identically. Hence, when correctly eliminating the bias caused by microstructure noise, the two estimated volatilities should not be significantly different from each other. The price of the QQQ has been scaled up to allow the comparison

3 Methodology

In this section we introduce and formalize the procedures to consistently evaluate the noise cancelling accuracy of a given estimator and compare its relative performance with respect to the competing estimators.

We define the noise cancelling performance as the 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.). However, if a realized measure is not able to completely remove the effect of the noise, an additional source of inaccuracy will be present, affecting the overall performance of the estimator. 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.
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. This grants the existence of a proxy for the integrated variance which is both unbiased and accurate.
Unfortunately, the noise cancelling ranking is not guaranteed to provide a consistent ranking for the IV estimates. That is to say, defined $L(\cdot, \cdot)$ any appropriate loss function:

$$\mathbb{E}\left[L\left(\Theta_{t,1}, \Gamma_{t,1}\right)\right] \lesseqgtr \mathbb{E}\left[L\left(\Theta_{t,2}, \Gamma_{t,2}\right)\right] \not\Leftrightarrow \mathbb{E}\left[L\left(IV_t, \Gamma_{t,1}\right)\right] \lesseqgtr \mathbb{E}\left[L\left(IV_t, \Gamma_{t,2}\right)\right]$$

In fact, following Patton (2011)[6], we can easily see that the log-linear relationship between underlying and futures implies a perfect correlation of the estimation error of Θ and Γ with the variable of interest which breaks the equality $\mathbb{E}\left[\Delta L\left(\Theta_t, \Gamma_t\right)\right] = \mathbb{E}\left[\Delta L\left(IV_t, \Gamma_t\right)\right]$.
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).

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

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:

$$d_{ij,t} = L_{t,i} - L_{t,j}. \tag{14}$$

Additionally,

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

will provide a consistent estimate of the difference in accuracy between any two realized measures (as $N \to \infty$) and under standard regularity conditions[7] we can implement bootstrap methods to test for differences in the estimators' performance.

3.1 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.

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{\text{var}}(\bar{d}_{ij})}}, \tag{15}$$

where $\widehat{\text{var}}(\bar{d}_{ij})$ 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.

3.2 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 $\mathcal{M}^* = \left\{ i \in \mathcal{M}^0 \,|\, \mathbb{E}\left(d_{ij,t}\right) \leq 0 \; \forall \; j \in \mathcal{M}^0 \right\}$, which will contain the best object with a given level of confidence (α). 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 α 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: $\mathcal{M}^*_{1-\alpha} = \mathcal{M}$.
Following Hansen et al. (2011) we will adopt two different equivalence tests and their coherent elimination rules. 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 either one of the two following statistics:

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

where $t_{i,j}$ has been introduced in equation (15) and tests the null $H_{0,ij} : \mathbb{E}\left[d_{ij}\right] = 0$.
The second t-statistic, t_i, is defined as:

$$t_i = \frac{\bar{d}_i}{\sqrt{\widehat{\mathrm{var}}(\bar{d}_i)}}$$

and is associated to the null hypothesis: $H_{0,i} : \mathbb{E}\left[\bar{d}_i\right] = 0$. With $\bar{d}_i = \frac{1}{M} \sum_{j \in \mathcal{M}} \bar{d}_{ij}$ measuring the sample loss of the ith model relative to the average across models in $\mathcal{M}$. While the asymptotic distributions of these statistics are non-standard, the use of appropriate bootstrap methods allows to retrieve the relevant distributions.
Finally, the associated elimination rules follow naturally; for $T_{\max,\mathcal{M}}$ we will use:

$$e_{max,\mathcal{M}} = \arg\max_{i \in \mathcal{M}} t_i,$$

which implies that a rejection of H_0 would identify as false $\mathbb{E}\left[\bar{d}_k\right] = 0$ for $k = e_{max,\mathcal{M}}$. Consequently, the elimination rule would remove the element k which contributes the most to the test statistic outcome.
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 $t_{e_{R,\mathcal{M}} k} = T_{\mathrm{R},\mathcal{M}}$ 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)[8]) 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 $\gamma_H \approx 5$)[9] 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 no-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 representative 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 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, 2009; Liu et al., 2015).

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 components 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.

Table 1 The unrestricted model

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

With S_t the log price process of the underlying asset.

Table 2 Prameters values

		Including DGP:					
	Value	DGP(1)	DGP(2)	DGP(3)	DGP(4)	DGP(5)	DGP(6)
μ	$2.4e^{-3}$	✓	✓	✓	✓	✓	✓
σ_S^2	$3.6e^{-4}$	✓	✗	✗	✗	✓	✗
r	$3.6e^{-4}$	✓	✓	✓	✓	✗	✗
κ_S	0.5		✓	✓	✓		✓
θ_S	$3.6e^{-4}$		✓	✓	✓		✓
σ_v	$3.0e^{-3}$		✓	✓	✓		✓
σ_{J_S}	$9.0e^{-7}$			✓			✓
λ_{J_S}	3			✓			✓
σ_{J_v}	$9.0e^{-4}$				✓		✓
λ_{J_v}	3				✓		✓
κ_r	2					✓	✓
θ_r	$4.0e^{-4}$					✓	✓
σ_r	0.02					✓	✓

We mark with ✓ a parameter included by the model and with ✗ when the parameter is implemented as initial value of the stochastic process. Following Christensen et al. (2014), we choose the size of σ_{J_v} such that $JV = \frac{\sum J^2}{[X]_1} \approx 0.2$ where $[X]_1 = \int_0^1 \sigma_u^2 \mathrm{d}u + \sum J^2$.

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). Additionaly, following Ramaswamy and Sundaresan (1985), we can expand this finding to obtain a closed-form solution for a futures written on any asset when we assume zero correlation between the stochastic process driving the interest rate and the one driving the underlying price process. Thus, to generate DGP(5) and DGP(6) we chose to model the interest rate dynamics as a CIR process (Cox, Ingersoll, and Ross, 1985). Consequently, instead of (9) we have:

$$F = Sa(\tau)e^{b(\tau)r} \tag{16}$$

where:

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

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

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 (16) and (1) respectively. 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[10].

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. Additionally, the results obtained using the T_R and T_{max} tests do not show any relevant difference, for ease of exposition only the T_R test results for the unrestricted model will be presented.

Table 3 MCS Ranking - DGP(6) - T_R

		All		$\lambda = 23400$		$\lambda = 7800$		$\lambda = 1560$		$\lambda = 390$		$\lambda = 78$		$\lambda = 39$	
		IID	MA(1)	IID	MA(1)	IID	MA(1)	IID	MA(1)	IID	MA(1)	IID	MA(1)	IID	MA(1)
RPA:	γ_H	X	X	X	X	X	X	X	X	X	X	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$
	γ_M	X	X	X	X	X	X	X	X	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$
	γ_L	X	X	X	X	X	X	X	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$
RK:	γ_H	X	X	X	X	$e_{\mathcal{M}_5}$	X	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	X	X	X	X	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$
	γ_M	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	X	X	X	X	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$
	γ_L	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	X	X	X	X	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$
RVTS:	γ_H	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$
	γ_M	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$
	γ_L	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$
RVMS:	γ_H	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$
	γ_M	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$
	γ_L	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$
RV5:	γ_H	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$
	γ_M	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$
	γ_L	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$
RV:	γ_H	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	X	X	X	X
	γ_M	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	X	$e_{\mathcal{M}_5}$	X	X	X	X
	γ_L	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	X	X	X	X

Included estimators (**X**) and elimination order for the T_R test with 10.000 bootstrap replications at a 95% confidence level. $\gamma_H = 5.22$, $\gamma_M = 2.43$ and $\gamma_L = 0.57$. 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.

Table 3, reports the MCS results for the unrestricted model; following the notation introduced by Hansen and Lunde (2006a) $e_{\mathcal{M}_1}$ characterize the first element to be eliminated (*i.e.* the one identified as worst performing), $e_{\mathcal{M}_2}$ identifies the second element to be eliminated 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 γ 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 gets often 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 λ 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 λ 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 cancelling 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[11]. 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 unsurprisingly given the asymptotic equivalence of realized kernel, two and multi-scale realized variance and realized pre-averaged variance.

Table 4 Estimated average standardized volatility over 10.000 Simulations

		All		$\lambda = 23400$		$\lambda = 7800$		$\lambda = 1560$		$\lambda = 390$		$\lambda = 78$		$\lambda = 39$	
		IID	MA(1)	IID	MA(1)	IID	MA(1)	IID	MA(1)	IID	MA(1)	IID	MA(1)	IID	MA(1)
RPA:	γ_H	1.012	1.027	1.012	1.027	1.012	1.017	1.01	1.011	1.008	1.008	**0.992**	**0.992**	0.962	0.962
	γ_M	**0.999**	1.002	**0.999**	1.002	**0.999**	**1**	**0.998**	**0.998**	**0.995**	**0.995**	0.981	0.981	0.952	0.952
	γ_L	0.996	**0.998**	0.996	**0.998**	**0.997**	**0.997**	**0.995**	0.995	0.993	0.993	**0.979**	**0.979**	0.95	0.95
RK:	γ_H	1.007	1.009	1.007	1.009	**1.003**	**1.004**	1.008	1.009	**1.006**	**1.006**	0.975	0.976	0.923	0.923
	γ_M	0.997	**1**	0.997	**1**	0.996	0.997	0.992	0.992	0.987	0.987	0.96	0.96	0.913	0.913
	γ_L	0.995	0.995	0.995	0.995	0.993	0.993	0.988	0.988	0.982	0.982	0.956	0.956	0.911	0.911
RVTS:	γ_H	**0.998**	3.224	**0.998**	3.214	0.996	1.369	**0.994**	1.033	0.989	0.994	0.976	0.977	0.949	0.949
	γ_M	0.997	2.163	0.997	2.156	0.997	1.186	0.993	1.01	0.99	0.991	0.973	0.974	0.947	0.947
	γ_L	**0.998**	1.24	**0.998**	1.238	**0.997**	1.028	0.994	**0.996**	0.991	0.991	0.973	0.973	0.947	0.947
RVMS:	γ_H	0.997	**0.997**	0.997	**0.997**	0.996	**0.996**	0.992	**0.992**	0.988	0.988	0.979	0.978	0.96	0.959
	γ_M	0.997	0.997	0.997	0.997	0.997	0.996	0.993	0.993	0.992	0.992	0.979	0.978	0.958	0.958
	γ_L	**0.998**	**0.998**	**0.998**	**0.998**	**0.997**	**0.997**	0.994	0.994	0.992	0.992	0.978	0.978	**0.958**	**0.958**
RV5:	γ_H	1.19	1.19	1.19	1.19	1.19	1.189	1.189	1.189	1.182	1.182	1.088	1.088	**1.012**	**1.013**
	γ_M	1.023	1.023	1.023	1.023	1.022	1.022	1.021	1.021	1.016	1.015	0.983	0.983	0.947	0.948
	γ_L	0.987	0.987	0.987	0.987	0.986	0.986	0.985	0.985	0.98	0.98	0.961	0.961	0.934	0.934
RV:	γ_H	63.352	33.011	63.186	33.006	21.777	18.41	5.151	5.016	2.029	2.021	1.178	1.177	1.047	1.047
	γ_M	12.486	6.955	12.455	6.954	4.826	4.213	1.761	1.737	1.182	1.181	**1.01**	**1.01**	**0.964**	**0.964**
	γ_L	1.762	1.462	1.76	1.462	1.253	1.219	1.047	1.045	**1.004**	**1.004**	0.974	0.974	0.947	0.947

The table reports the average daily standardized volatility for DGP(6) computed for different trading intensities and different level of noises.

5 Empirical Application

The dataset of futures used consists of transaction data recorded over a 4 months period (from July 1^{st} - October 31^{st} 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.

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 mainly 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.

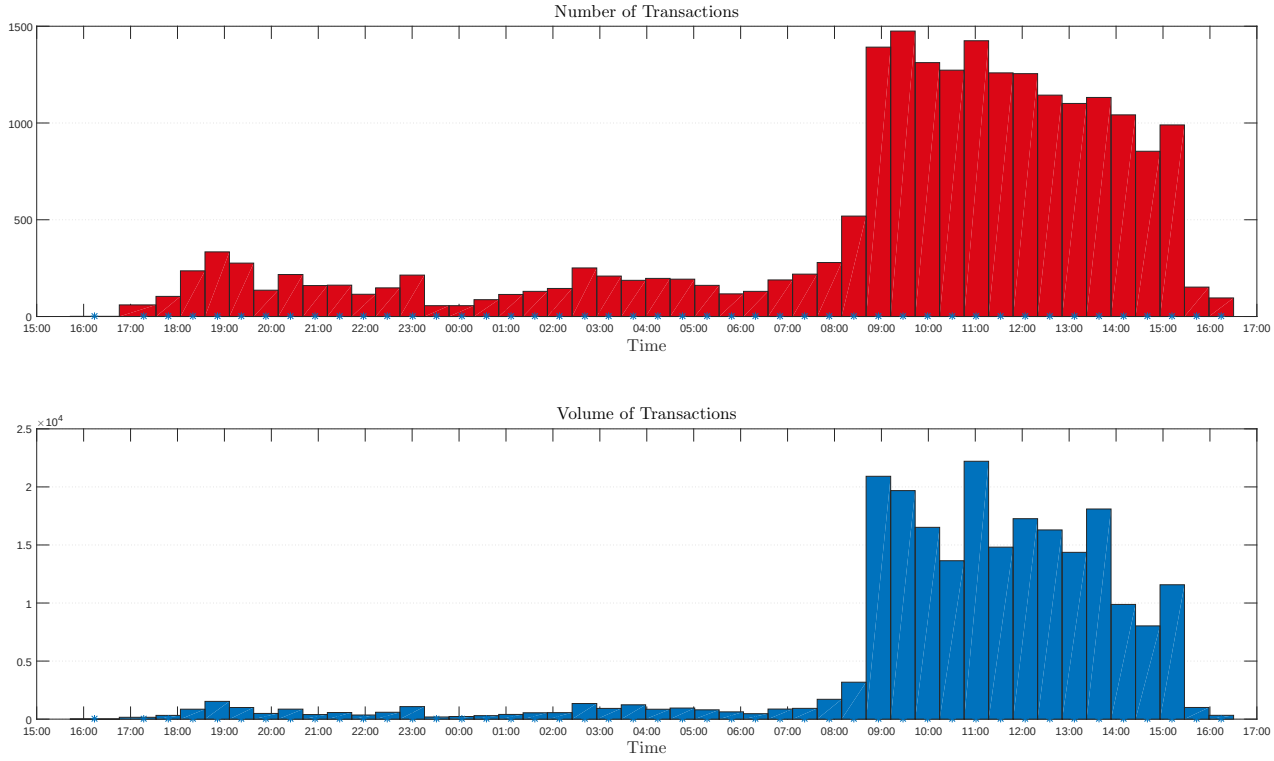

Figure 4. The figure shows a trading day for the E-mini Dow Jones with December expiration. Each bar represents a 30 minutes interval with the top graph showing the number of transactions occurred and the bottom graph showing the cumulative traded volume per interval.

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[12].

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 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)[13] , 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

	SP		NQ		YM		MD	
	Sept.	Dec.	Sept.	Dec.	Sept.	Dec.	Sept.	Dec.
Avg. time between trades (sec.)	0.15	0.24	0.44	0.26	0.63	0.42	1.82	1.68
Avg. number of trades	268,352	303,584	86,193	109,665	58,589	74,289	13,753	15,543
Max number of trades	510,497	531,992	163,069	212,007	112,338	137,590	33,265	37,254
Min number of trades	8,520	7,174	7,043	15,357	5,166	6,965	7,514	8,361
Avg. Noise Variance (e^{-7})	0.482	0.547	0.523	3.851	0.356	3.547	0.698	1.823
Number of observations	14,759,380	11,153,203	4,568,248	3,947,955	3,105,214	2,674,428	673,885	544,014

Table 6 ETF Data: Summary Statistics

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

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[14] are removed before applying a final, more elaborated filtering algorithm. 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 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 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 (λ) and average Noise Ratio (γ) 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 $\bar{\omega}^2 = n^{-1}\sum_{t=1}^{n}\widehat{\omega}_t^2$ and $\bar{IV} = n^{-1}\sum_{t=1}^{n}\widehat{IV}_t$. 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 7 we can observe that $\hat{\lambda}$ tends to be very similar across series even though Table 5 presented very different values for $\widehat{\omega}^2$. 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 7 Noise-to-Signal (e^{-2}) and Noise Ratio

	SPY	SP		QQQ	NQ		DIA	YM		MDY	MD	
		Sept.	Dec.		Sept.	Dec.		Sept.	Dec.		Sept.	Dec.
Avg. noise-to-signal	0.18	0.22	0.22	0.18	0.17	0.18	0.18	0.19	0.19	0.17	0.17	0.19
Max noise-to-signal	0.28	0.71	0.36	0.25	0.24	0.24	0.30	0.26	0.25	0.29	0.28	0.28
Min noise-to-signal	0.09	0.08	0.12	0.11	0.11	0.13	0.09	0.07	0.13	0.07	0.08	0.09
Avg. Noise Ratio	0.38	1.41	1.15	0.48	0.42	0.39	0.31	0.46	0.41	0.34	0.32	0.33
Max Noise Ratio	0.8	2.36	2.45	0.99	0.77	0.62	0.82	0.78	0.75	0.66	0.56	0.52
Min Noise Ratio	0.03	0.62	0.19	0.04	0.03	0.10	0.03	0.02	0.06	0.04	0.04	0.06

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 ($\gamma = \sqrt{\lambda * m}$) 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 (as $\mathbb{E}(RV) = IV(1 + 2\gamma^2)$). Therefore, we can identify the level of noise to which each series is subject. On this aspect strikes the relative size of the estimated γ 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 7 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 equation (12).

Figures (5) and (6) 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).

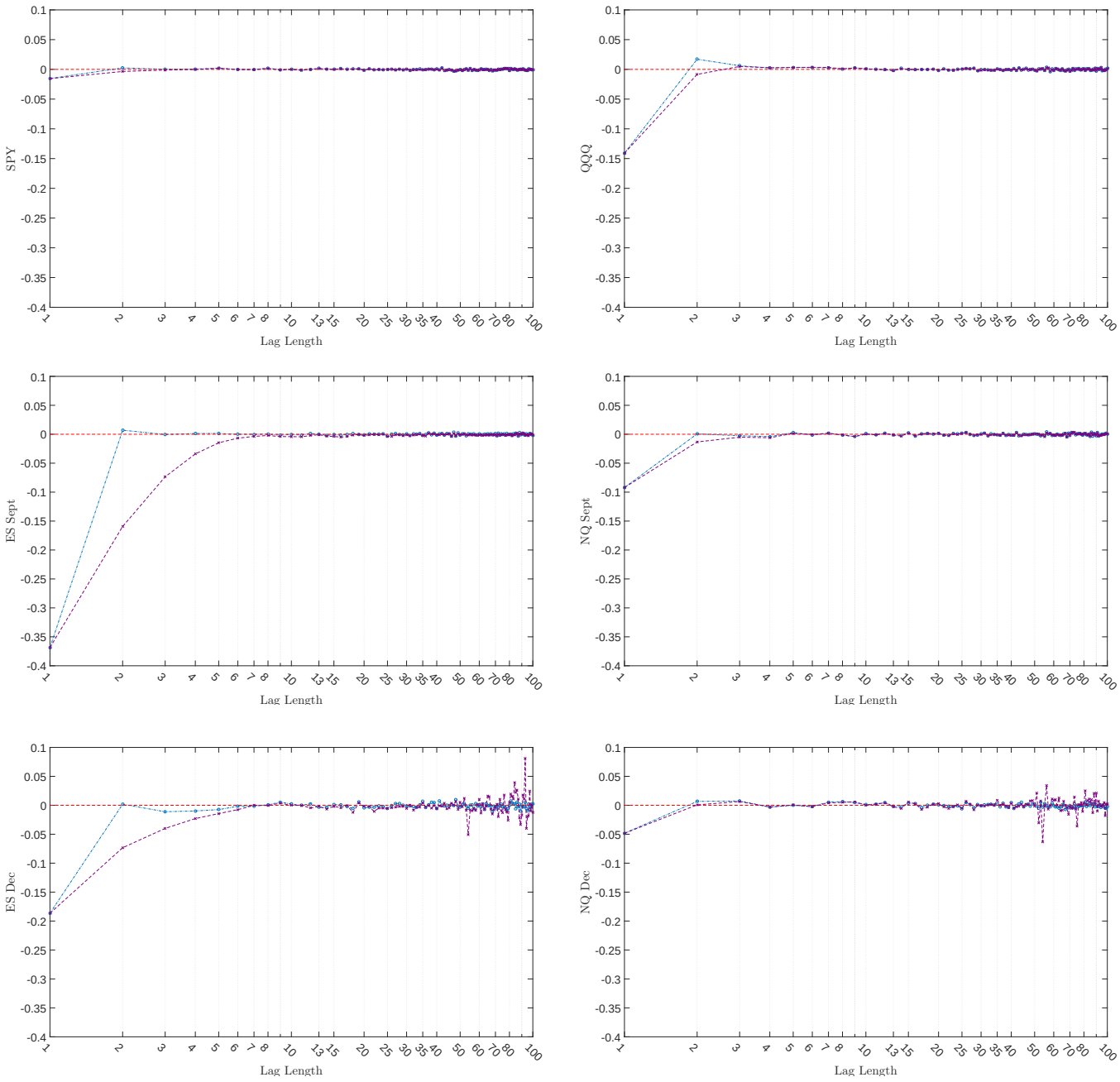

Figure 5. The figure shows the average ACF (o–o–o) and PACF (×-·-×-·-×) of two different underlying (SPY and QQQ) and the relative Futures contracts.

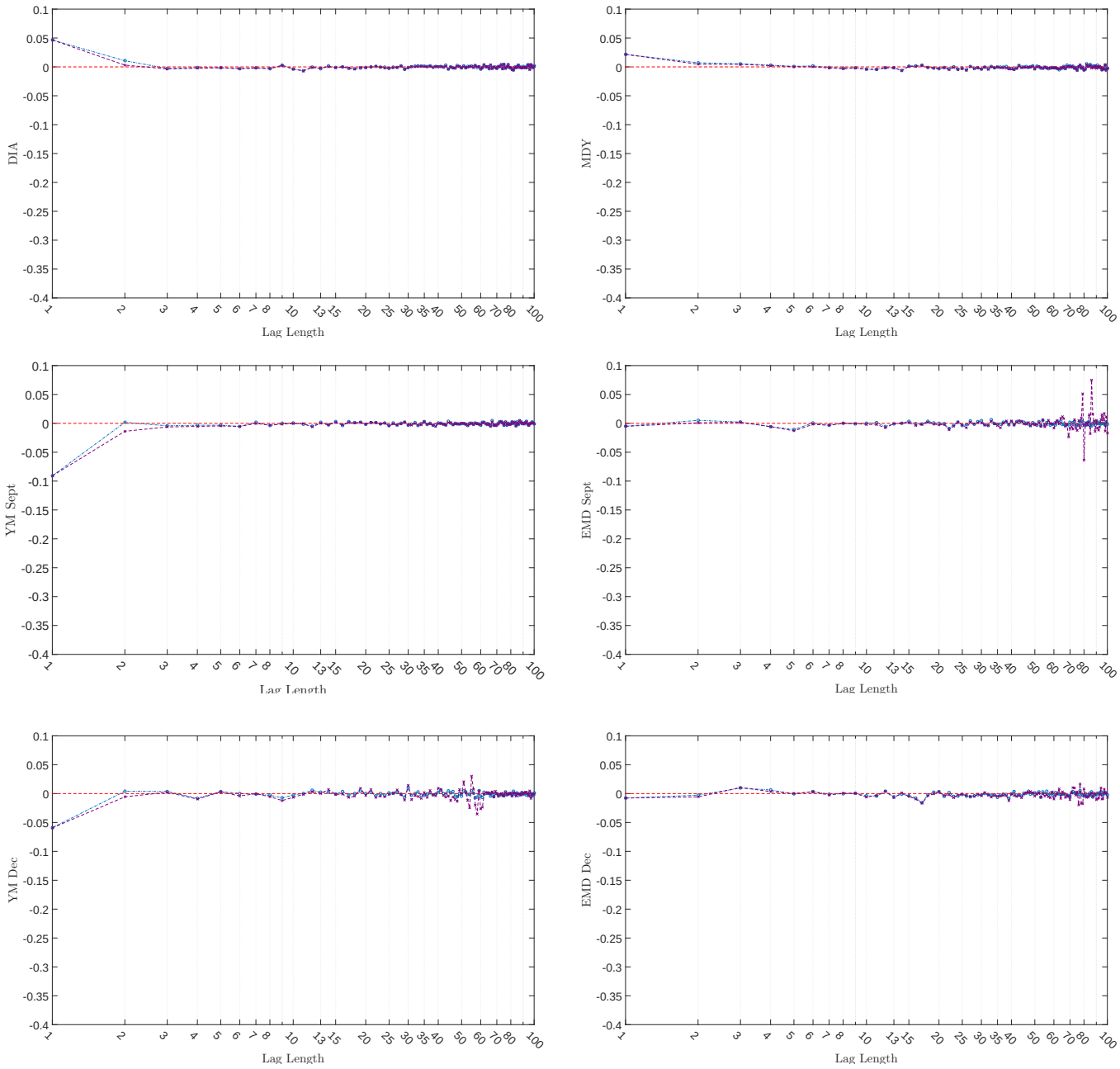

Figure 6. The figure shows the average ACF (o–o–o) and PACF (×·–×·–×) of two different underlying (DIA and MDY) and the relative Futures contracts.

5.3 Findings

Tables 8 to 10 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 8, 9 and 10 present the average daily estimates for the different realized measures considered. While Table 10 consider the period subsequent the expiration of the September contract, Tables 8 and 9 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[15], 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 8 Estimated average daily volatility (%) July - September

	SPY	ES Sept.	ES Dec.	QQQ	NQ Sept.	NQ Dec.
RPA:	0.4825	0.4920	0.5449	0.5616	0.5643	0.5387
RK:	0.4828	0.4919	0.5455	0.5630	0.5662	0.5375
RVTS:	0.4740	0.5016	0.5311	0.5363	0.5679	0.5325
RVMS:	0.4755	0.5098	0.5385	0.5494	0.5676	0.5253
RV5	0.4656	0.4750	0.5519	0.5501	0.5466	0.5071
RV	0.4780	0.9731	0.7701	0.6440	0.6184	0.5750

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.

Table 9 Estimated average daily volatility (%) July - September

	DIA	YM Sept.	YM Dec.	MDY	EMD Sept.	EMD Dec.
RPA:	0.4559	0.4592	0.4731	0.6334	0.6421	0.5574
RK:	0.4559	0.4592	0.4754	0.6302	0.6424	0.5541
RVTS:	0.4517	0.4707	0.4301	0.6111	0.6347	0.5749
RVMS:	0.4557	0.4713	0.4341	0.6192	0.6376	0.5651
RV5	0.4468	0.4511	0.4494	0.6250	0.6246	0.5488
RV	0.4410	0.5076	0.4545	0.5988	0.6302	0.5686

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.

Figures 7 to 9 offer a visual explanation of the problem as they show the time series of the daily estimates for each series and across the whole sample period. A quick eye balling of the figures 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 γ and λ, 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. The same fact can be observed for the E-mini Dow series, which, as evidenced by the reported noise ratio on table 7, are affected by a higher degree of noise compared to the considered proxy (DIA).

Table 10 Estimated average daily volatility (%) September - October

	SPY	ES Dec.	QQQ	NQ Dec.	DIA	YM Dec.	MDY	EMD Dec.
RPA:	0.5368	0.5426	0.6478	0.6410	0.5002	0.5019	0.6350	0.6602
RK:	0.5376	0.5436	0.6494	0.6439	0.4997	0.5044	0.6374	0.6616
RVTS:	0.5270	0.5502	0.6140	0.6365	0.4984	0.5076	0.6461	0.6572
RVMS:	0.5299	0.5593	0.6323	0.6422	0.5034	0.5126	0.6503	0.6626
RV5	0.5179	0.5374	0.6377	0.6238	0.4881	0.4905	0.6387	0.6315
RV	0.5155	0.9333	0.6896	0.6602	0.4721	0.5318	0.6270	0.6618

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

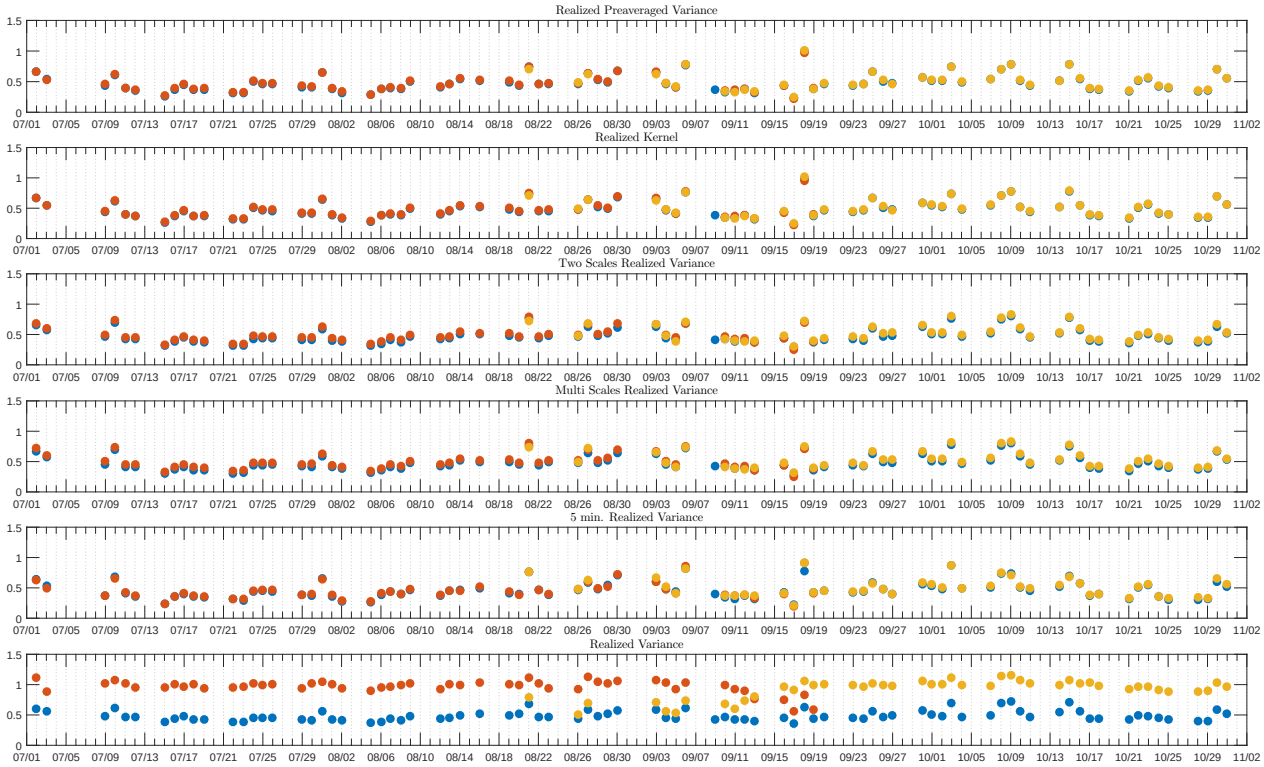

Figure 7. **The figure shows the Time Series of daily realized measures estimated for the (SPY), the (ES Sept.) and the (ES Dec.).**

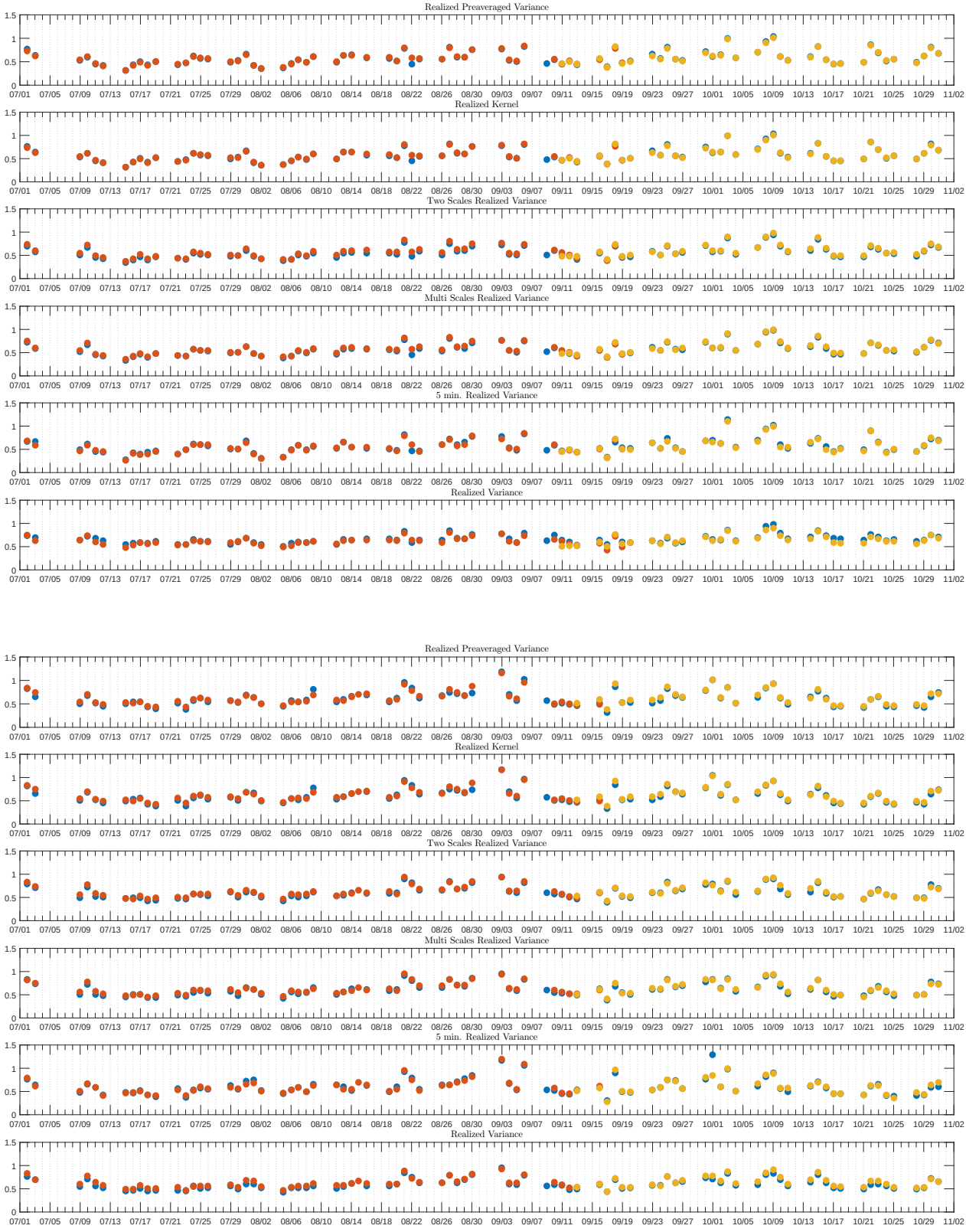

Figure 8. The figure shows the Time Series of daily realized measures estimated. The top figure displays the (QQQ), the (NQ Sept.) and the (NQ Dec.). While the bottom figure reports the (MDY), the (EMD Sept.) and the (EMD Dec.)

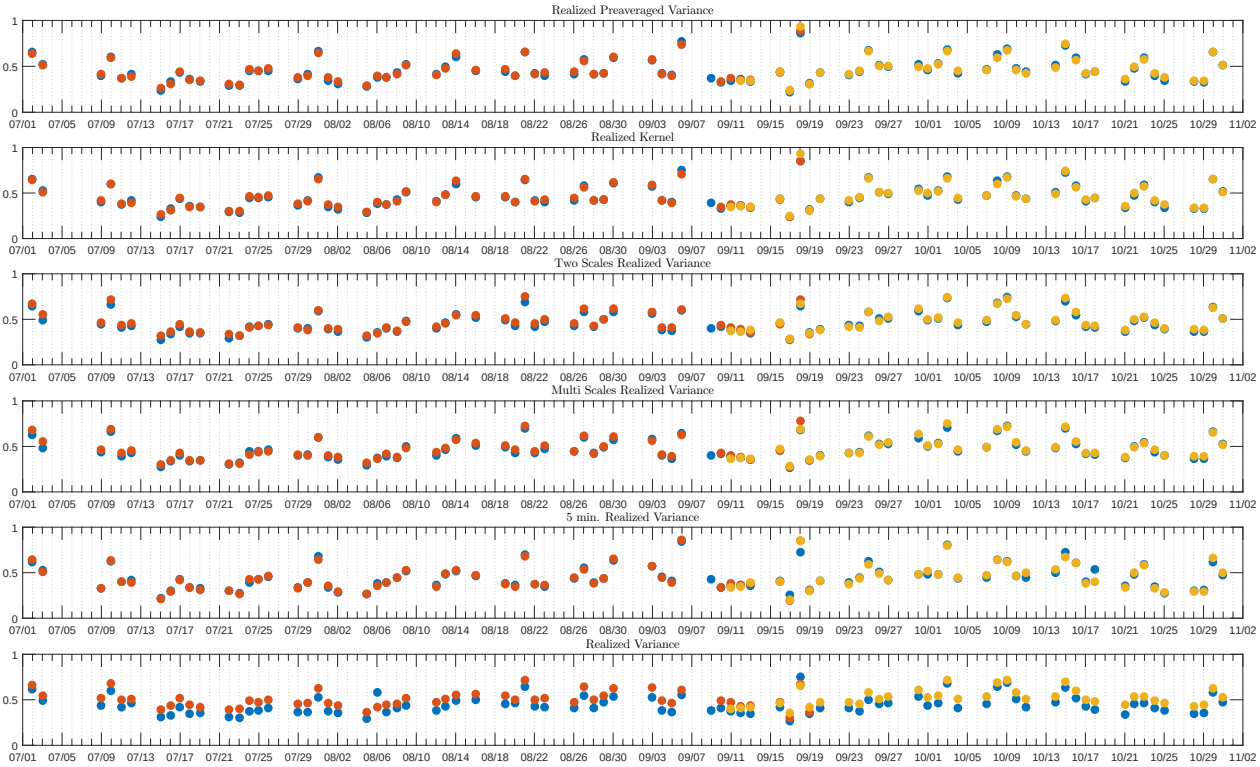

Figure 9. The figure shows the Time Series of daily realized measures estimated for the (DIA), the (YM Sept.) and the (YM Dec.).

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[16]. 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 11 reports the included estimators and the elimination order for both the T_R and $T_{\max}$ 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 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 11 MCS Ranking

	ES Sept.		ES Dec.		NQ Sept.		NQ Dec.		YM Sept.		YM Dec.		EMD Sept.		EMD Dec.	
	T_{max}	T_R	T_{max}	T_R	T_{max}	T_R	T_{max}	T_R	T_{max}	T_R	T_{max}	T_R	T_{max}	T_R	T_{max}	T_R
RPA:	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	X	X	X	X	X	X	$e_{\mathcal{M}_5}$	$e_{\mathcal{M}_5}$	X	$e_{\mathcal{M}_3}$	X	X	X	X
RK:	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
RVTS:	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	X	X	X	X	X	X
RVMS:	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	X	X	X	X	X	X
RV5:	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_3}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_4}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	X	X	X	X
RV:	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_2}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	$e_{\mathcal{M}_1}$	X	X

Included estimators (X) and elimination order for both the T_R and $T_{\max}$ tests with 10.000 bootstrap replications at a 95% confidence level.

Table 12 presents sample results of the Diebold-Mariano type test for a representative asset and both its expirations[17]. The test reports the values obtained using the QLIKE loss function[18]. 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.

Table 12 Pairwise comparison test results: E-mini S&P 500 Sept.

Dec. \Sept.	RPA	RK	RVTS	RVMS	RV5	RV
RPA:	0	-3.8743^{***}	9.1409^{***}	8.0964^{***}	1.4858^{*}	11.3561^{***}
RK:	0.4975	0	9.2017^{***}	8.1672^{***}	1.7077^{**}	11.3535^{***}
RVTS:	-5.9919^{***}	-6.3310^{***}	0	2.2688^{**}	-1.6758^{**}	11.3158^{***}
RVMS:	-7.7719^{***}	-8.2971^{***}	-2.2222^{***}	0	-2.3404^{**}	11.3528^{***}
RV5:	-2.8538^{***}	-2.8763^{***}	0.0848	0.9364	0	10.9825^{***}
RV:	-7.2875^{***}	-7.2926^{***}	-7.2045^{***}	-7.2518^{***}	-7.0904^{***}	0

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 $\mathcal{BSM}$ 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 to test and evaluate the noise cancelling 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 the noise structures, noise and trading intensities. 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 consistently selected for lower values of λ. However, in line with the empirical findings, as a stochastic interest rate process is introduced the MCS procedure consistently includes in the optimum set the RK also 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 registered.

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.

Notes

[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.

[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).

[6]For details refer to: Patton 2011, Appendix, Proof of Proposition 1.

[7]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).

[8]with MA parameter: $\phi_{U,1} = \phi_{V,1} = 0.8$

[9]We define the Noise Ratio γ introduced by Oomen (2006) in section 5.2.

[10]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%.

[11]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.

[12]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.

[13]Consult Barndorff-Nielsen et al. (2008) Section 5.1. "Estimating ω^2" for details.

[14]These are, for example, observations with a bid, ask or transaction price equal to zero or entries marked as corrected trades (CORR$\neq$ 0). A detailed procedure can be found in Barndorff-Nielsen et al. (2009).

[15]As discussed in section 5.1 only days with more than 4680 observations are used.

[16]The linex function introduced by Varian (1975) and a standard quadratic loss function have also been tested. The results are available upon request.

[17]The complete results are available upon request.

[18]Results for different loss functions are available upon request.

References

Aït-Sahalia, Yacine, and Loriano Mancini, 2008, Out of sample forecasts of quadratic variation, *Journal of Econometrics* 147, 17 – 33, Econometric modelling in finance and risk management: An overview.

Aït-Sahalia, Yacine, Per A. Mykland, and Lan Zhang, 2011, Ultra high frequency volatility estimation with dependent microstructure noise, *Journal of Econometrics* 160, 160 – 175, Realized Volatility.

Andersen, Torben G., 2000, Some reflections on analysis of high-frequency data, *Journal of Business & Economic Statistics* 18, 146–153.

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold, 2007, Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility, *The Review of Economics and Statistics* 89, 701–720.

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold, 2010, Parametric and nonparametric volatility measurement, in *Handbook of Financial Econometrics, Vol 1*, chapter 2, 67–137 (Elsevier B.V.).

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 1999, (understanding, optimizing, using and forecasting) realized volatility and correlation, Manuscript, Northwestern University, Duke University and University of Pennsylvania., Published in revised form as "Great Realizations," Risk, March 2000, 105-108.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2001, The distribution of realized exchange rate volatility, *Journal of the American Statistical Association* 96, 42–55.

Andersen, Torben G., Tim Bollerslev, and Nour Meddahi, 2011, Realized volatility forecasting and market microstructure noise, *Journal of Econometrics* 160, 220 – 234, Realized Volatility.

Bandi, F. M., and J. R. Russell, 2008, Microstructure noise, realized variance, and optimal sampling, *The Review of Economic Studies* 75, 339–369.

Barndorff-Nielsen, O. E., P. Reinhard Hansen, Asger Lunde, and Niel Shephard, 2009, Realized kernels in practice: trades and quotes, *Econometrics Journal* 12, C1–C32.

Barndorff-Nielsen, Ole E., Peter Reinhard Hansen, Asger Lunde, and Neil Shephard, 2008, Designing realized kernels to measure the ex post variation of equity prices in the presence of noise, *Econometrica* 76, 1481–1536.

Barndorff-Nielsen, Ole E., and Neil Shephard, 2002, Estimating quadratic variation using realized variance, *Journal of Applied Econometrics* 17, 457–477.

Barndorff-Nielsen, Ole E., and Neil Shephard, 2007, Variation, jumps, and high-frequency data in financial econometrics, in Richard Blundell, Whitney Newey, and Torsten Persson, eds., *Advances in Economics and Econometrics*, volume 3, 328–372 (Cambridge University Press), Cambridge Books Online.

Bianco, S., F. Corsi, and R. Renò, 2009, Intraday LeBaron effects, *Proceedings of the National Academy of Science* 106, 11439–11443.

Björk, Tomas, 2009, *Arbitrage Theory in Continuous Time*, third edition (Oxford Finance).

Brownlees, Christian T., and Giampiero M. Gallo, 2006, Financial econometric analysis at ultra-high frequency: Data handling concerns, *Computational Statistics & Data Analysis* 51, 2232–2245.

Cakici, Nusret, and Sris Chatterjee, 1991, Pricing stock index futures with stochastic interest rates, *Journal of Futures Markets* 11, 441–452.

Chen, Nai-fu, Charles J. Cuny, and Robert A. Haugen, 1995, Stock volatility and the levels of the basis and open interest in futures contracts, *The Journal of Finance* 50, 281–300.

Christensen, Kim, Roel Oomen, and Mark Podolskij, 2010, Realised quantile-based estimation of the integrated variance, *Journal of Econometrics* 159, 74 – 98.

Christensen, Kim, Roel C.A. Oomen, and Mark Podolskij, 2014, Fact or friction: Jumps at ultra high frequency, *Journal of Financial Economics* 114, 576 – 599.

Cornell, Bradford, and Kenneth R. French, 1983, The pricing of stock index futures, *Journal of Futures Markets* 3, 1–14.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross, 1985, A theory of the term structure of interest rates, *Econometrica* 53, 385–407.

Delbaen, Freddy, and Walter Schachermayer, 1994, A general version of the fundamental theorem of asset pricing., *Mathematische Annalen* 300, 463–520.

Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business & Economic Statistics* 13, 253–263.

Hansen, Peter R., and Asger Lunde, 2005, A forecast comparison of volatility models: does anything beat a garch(1,1)?, *Journal of Applied Econometrics* 20, 873–889.

Hansen, Peter R., and Asger Lunde, 2006a, Realized variance and market microstructure noise, *Journal of Business & Economic Statistics* 24, 127–161.

Hansen, Peter R., Asger Lunde, and James M. Nason, 2011, The model confidence set, *Econometrica* 79, 453–497.

Hansen, Peter Reinhard, and Asger Lunde, 2006b, Consistent ranking of volatility models, *Journal of Econometrics* 131, 97 – 121.

Jacod, Jean, Yingying Li, Per A. Mykland, Mark Podolskij, and Mathias Vetter, 2009, Microstructure noise in the continuous case: The pre-averaging approach, *Stochastic Processes and their Applications* 119, 2249 – 2276.

Jacod, Jean, and Albert N. Shiryaev, 2003, *Limit Theorems for Stochastic Processes*, Grundlehren der mathematischen Wissenschaften 288, second edition (Springer-Verlag Berlin Heidelberg).

Kalnina, Ilze, and Oliver Linton, 2008, Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error, *Journal of Econometrics* 147, 47 – 59, Econometric modelling in finance and risk management: An overview.

Li, J., and A. J. Patton, 2015, Asymptotic inference about predictive accuracy using high frequency data, Working paper, Duke University.

Liu, Lily Y., Andrew J. Patton, and Kevin Sheppard, 2015, Does anything beat 5-minute rv? a comparison of realized measures across multiple asset classes, *Journal of Econometrics* 187, 293 – 311.

Munk, Claus, 2011, *Fixed Income Modelling* (Oxford University Press, USA).

Mykland, Per A., and Lan Zhang, 2009, *The Econometrics of High Frequency Data*.

Oomen, Roel C. A., 2006, Properties of realized variance under alternative sampling schemes, *Journal of Business & Economic Statistics* 24, 219 – 237.

Patton, Andrew J., 2011, Data-based ranking of realised volatility estimators, *Journal of Econometrics* 161, 284 – 303.

Patton, Andrew J., and Kevin Sheppard, 2009, Optimal combinations of realised volatility estimators, *International Journal of Forecasting* 25, 218 – 238, Forecasting Returns and Risk in Financial Markets using Linear and Nonlinear Models.

Podolskij, Mark, and Mathias Vetter, 2009, Bipower-type estimation in a noisy diffusion setting, *Stochastic Processes and their Applications* 119, 2803 – 2831.

Protter, Philip E., 2004, *Stochastic integration and differential equations*, Applications of mathematics 21 0172-4568, second edition (Springer).

Ramaswamy, Krishna, and Suresh M. Sundaresan, 1985, The valuation of options on futures contracts, *The Journal of Finance* 40, 1319–1340.

Rossi, Eduardo, and Paolo Santucci de Magistris, 2013, A no-arbitrage fractional cointegration model for futures and spot daily ranges, *Journal of Futures Markets* 33, 77–102.

U.S. Commodity Futures Trading Commission and U.S. Securities & Exchange Commission, 2010, Findings regarding the market events of may 6, 2010, Unpublished.

Varian, Hal R, 1975, A bayesian approach to real estate assessment, *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage* 195–208.

Zhang, Lan, 2006, Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach, *Bernoulli* 12, 1019–1043.

Zhang, Lan, Per A. Mykland, and Yacine Aït-Sahalia, 2005, A tale of two time scales: Determining integrated volatility with noisy high-frequency data, *Journal of the American Statistical Association* 100, 1394–1411.

Research Papers 2016

2017-07: Oskar Knapik: Modeling and forecasting electricity price jumps in the Nord Pool power market

2017-08: 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

2017-09: Thomas Quistgaard Pedersen and Erik Christian Montes Schütte: Testing for Explosive Bubbles in the Presence of Autocorrelated Innovations

2017-10: Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Dynamics of Variance Risk Premia, Investors' Sentiment and Return Predictability

2017-11: Søren Johansen and Morten Nyboe Tabor: Cointegration between trends and their estimators in state space models and CVAR models

2017-12: Lukasz Gatarek and Søren Johansen: The role of cointegration for optimal hedging with heteroscedastic error term

2017-13: Niels S. Grønborg, Asger Lunde, Allan Timmermann and Russ Wermers: Picking Funds with Confidence

2017-14: Martin M. Andreasen and Anders Kronborg: The Extended Perturbation Method: New Insights on the New Keynesian Model

2017-15: Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante: A Non-Structural Investigation of VIX Risk Neutral Density

2017-16: Davide Delle Monache, Stefano Grassi and Paolo Santucci de Magistris: Does the ARFIMA really shift?

2017-17: Massimo Franchi and Søren Johansen: Improved inference on cointegrating vectors in the presence of a near unit root using adjusted quantiles

2017-18: Matias D. Cattaneo, Michael Jansson and Kenichi Nagasawa: Bootstrap-Based Inference for Cube Root Consistent Estimators

2017-19: Daniel Borup and Martin Thyrsgaard: Statistical tests for equal predictive ability across multiple forecasting methods

2017-20: Tommaso Proietti and Alessandro Giovannelli: A Durbin-Levinson Regularized Estimator of High Dimensional Autocovariance Matrices

2017-21: Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Variance swap payoffs, risk premia and extreme market conditions

2017-22: Jakob Guldbæk Mikkelsen: Testing for time-varying loadings in dynamic factor models

2017-23: Roman Frydman, Søren Johansen, Anders Rahbek and Morten Nyboe Tabor: The Qualitative Expectations Hypothesis: Model Ambiguity, Concistent Representations of Market Forecasts, and Sentiment

2017-24: Giorgio Mirone: Inference from the futures: ranking the noise cancelling accuracy of realized measures