Cross-sectional noise reduction and more efficient estimation of Integrated Variance

Giorgio Mirone

CREATES Research Paper 2018-18

Cross-sectional noise reduction and more efficient estimation of Integrated Variance[*]

Giorgio Mirone[†]

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

June 19, 2018

Abstract

In this paper we propose a straightforward approach to obtain a more efficient estimate of the integrated variance of an asset through a cross-sectional combination with a futures contract written on it. Our method constructs a variance-preserving series with reduced noise size as a linear combination of the underlying asset and the futures and base measurement of the integrated variance on this new series. We first illustrate how a theoretically but infeasible optimal series can be obtained and then suggest a feasible procedure to attain noise reduction. In a simulation study we verify how prevalent estimators of integrated variance applied to such noise-reduced series outperform estimators applied directly to the asset price. Finally, we apply the method to an empirical data set and, through the stabilized signature plot, we show how the noise reduced series provides consistent integrated variance estimates using naive realized measures at very high frequencies.

JEL classification: C10, C60, C80.

Keywords: Realized Covariance, High-frequency data, Volatility Estimation, Market Microstructure Noise, Noise reduction, Volatility Signature Plot, Realized Variance.

1 Introduction

Since the pioneering paper by Andersen et al. (2000) the high-frequency econometrics literature has seen a spurt of noise robust estimators of the QV of an asset; among the others we can name the aforementioned Jacod et al. (2009), Barndorff-Nielsen et al. (2008), Christensen et al. (2010b), Zhang et al. (2005), Zhang (2006), Bandi and Russell (2008). These estimators all try to smooth out the noise with a weight or kernel function working on the time dimension. Further, some of these estimators also enjoy an optimal rate of convergence. In the present paper we do not aim at providing a new estimator of the IV, rather our objective is to show that the problem can be approached (and somehow solved) from a totally different perspective.

We propose a novel noise reducing methodology obtained via linear combination of different price series with the same variance. This creates a variance-preserving series whose noise variance can be minimized with an appropriate weighting function. In this paper we specifically focus on estimating volatility of the S&P 500 index using high-frequency data of the SPY obtained from the TAQ database as a proxy;[1] this allows us to work on two different directions. We begin by discussing in details the relationship between a futures and its underlying asset; following Mirone (2017) we show that under the classical semimartingale framework the two financial instruments must, by no arbitrage, share the same integrated variance (IV) or, more generally, the same quadratic variation (QV). We then introduce the noise reduction methodology offering both a theoretical and a feasible approach on how to construct the noise reduced series (Z). Further, we discuss how the same approach can be implemented on an asset traded on different markets as long as the microstructure noise processes differ.

The essence of our method is akin to the preaveraging methodology (Jacod et al., 2009), with the main difference that our proposal aims at smoothing out the noise by averaging over the cross-section of returns instead of the time dimension. Our novel methodology produces a variance preserving series less affected by microstructure noise, which allows practitioners to implement their favourite realized measure in order to more efficiently retrieve the quadratic variation of the series.

In Section 3 and Section 4 we show that the trivial realized variance computed on the noise reduced series (Z) grants consistent results at frequencies as high as five - ten seconds, allowing to take advantage of an incredibly higher number of observations compared to the commonly used five minutes frequency.

As our approach requires estimation of the covariance of returns in a high-frequency and

---

[1]The use of an ETF as a proxy for the market index is common practice in the literature.

asynchronous framework, we will build on the work of Renò (2003); Hayashi and Yoshida (2005); Christensen et al. (2010a) and discuss to which extent our methodology is affected by the Epps effect (Epps, 1979). In this context, two different synchronization methods are investigated; the so-called previous tick interpolation and the refresh time discussed in Barndorff-Nielsen et al. (2011).

The remainder of the paper is structured as follows. Section 2 introduces the framework and thoroughly illustrates the noise reduction methodology. Section 3 presents a comprehensive simulation study to appraise and investigate the sensitivity of the procedure to different factors. Empirical results on the SPY and E-mini S&P 500 futures are given in Section 4. Conclusions are discussed in Section 5

2 Framework and Method

In this section we delineate our framework in which price and volatility processes are modelled. We also introduce an infeasible method of noise reduction that is of theoretical interest and a feasible method to be applied in practice. Henceforth, we work on a filtered probability space $(\Omega, \mathscr{F}, \mathbf{F} = (\mathscr{F}_t), \mathbb{P})$ that satisfies the usual hypotheses, that is, it is right continuous and complete.

2.1 The relation between the integrated variances of futures and its underlying asset

In consonance with the prevalent literature, we assume that the *latent efficient log-price process* S follows an Itô semimartingale; that is,

$$S_t = S_0 + \int_0^t b_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}W_s, \tag{1}$$

where b is locally bounded, σ is càdlàg and predictable, and W is a standard Wiener process. It is worth mentioning that the assumption on S can be generalized to accommodate jumps.[2] Our method, which serves as a prewhitening filter, is not affected by the assumption imposed on S, but relies on the noise structure as will be discussed in the sequel. The primary interest of this paper concerns the estimation of the *integrated variance*, defined as quadratic variation of S, i.e.,

---

[2] For example, S can take the form of Grigelionis representation as in theorem 2.1.2 in Jacod and Protter (2011).

$$IV_t(S) = [S,S]_t = \int_0^t \sigma_s^2 \, \mathrm{d}s. \tag{2}$$

A well known result in the semimartingale theory is that the integrated variance can be consistently estimated by the *realized variance*, which is constructed as the sum of squared difference of the process on a grid with vanishing mesh. In details, let $\tau_i^{(n)} : 0 \leq i \leq n$ denote the sampling times of the price process, which is a grid of finite partitions of the interval $[0, t]$. The realized variance spanning the same time window is computed as

$$RV_t(S) = \sum_{i=1}^n (S_{\tau_i^{(n)}} - S_{\tau_{i-1}^{(n)}})^2. \tag{3}$$

Then, as long as the mesh of the observations grid goes to zero the realized variance converges in probability to IV: $\max_{1 \leq i \leq n} |\tau_i^{(n)} - \tau_{i-1}^{(n)}| \to 0$ as $n \to \infty$, $RV_t(S) \xrightarrow{\mathbb{P}} IV_t(S)$. However, it has long been documented that realized variance does not converge when applied to empirical data. The volatility signature plot introduced by Andersen et al. (2000) typically visualizes the explosive behaviour of realized variance at ultra high frequencies, suggesting that the price process cannot simply be modelled as a semimartingale. Instead, it is argued that though the latent efficient price can be modelled as a semimartingale, the observed price has an extra *noise* component arising from market microstructure features, such as bid-ask bounces, discreteness of prices and outliers. In other words, if $S^\star$ denotes the observed price, then

$$S^\star = S + \varepsilon^{(S)}, \tag{4}$$

where $\varepsilon^{(S)}$ captures all sources of microstructure noise. The "contaminated" price process is no longer a semimartingale and thus the realized variance constructed in (3) no longer consistently estimate the integrated variance. At mild sampling frequencies, say 5, 10 or 30 minutes, the effect of noise is typically moderate to negligible; hence, in practice, realized variance computed at these frequencies is often used as estimator of the integrated variance. A less heuristic choice of the sampling frequency has also been proposed by Bandi and Russell (2008) with their optimal sampling RV (RV^*). Various others estimators have been proposed to address the issue of microstructure noise, including the two-scale realized variance (TSRV) (Zhang et al., 2005) and the multi-scale realized variance (MSRV) (Zhang, 2006), which consistently estimate the integrated variance (provided the noise process is temporally independent and identically distributed) or the realized kernel (RK) (Barndorff-Nielsen et al., 2008) and the pre-averaging (PA) estimator (Jacod et al., 2009) which are robust to temporally dependent noise.

It should be pointed out that all aforementioned methods tackle noise by a transformation

of the (observed) price process, say $S^\star$. In this paper, however, we adopt a different perspective. Instead of transforming the price process $S^\star$, we seek to mitigate the effect of noise by combining $S^\star$ with a different time series which enjoys the same variance as S. To this end, we will expand on Mirone (2017) and specifically focus on a futures, F, written on the series of interest, S.

Suppose that in the market there are two additional assets: a risk-free asset, B, whose dynamics is given by $dB_t = r_t dt$, where r is the short rate process, and a futures written on S_t with maturity T greater than t. We denote the *latent efficient log futures price* by F_t and analogously denote by $F_t^\star$ the observed futures price. The noise contaminating the futures is denoted by $\varepsilon_t^{(F)}$, and the processes are linked by the following equation:

$$F_t^\star = F_t + \varepsilon_t^{(F)}. \tag{5}$$

Our method is based on the close relation between the integrated variance of the underlying and the futures written on it. The efficient futures price, F_t, is the unique price that would be determined in a market that admits no free lunch. The well known fundamental theorem of asset pricing (Delbaen and Schachermayer, 1994) entails existence of the risk neutral measure $\mathbb{Q}$. Moreover, if $X_t = \exp(S_t)$ and $Y_t = \exp(F_t)$ are the price processes of the underlying asset and the futures respectively, then we have the following relation (see, e.g., Björk, 2009)

$$Y_s = \mathbb{E}^{\mathbb{Q}}[X_T \,|\, \mathscr{F}_s], \ s \leq T. \tag{6}$$

In general, Equation (6) admits no closed form solution unless we impose additional assumptions on the interest rate process r. An appealing specification of r might be the CIR process (Cox et al., 1985), which together with other assumptions leads to a closed-form solution for F (see, Ramaswamy and Sundaresan, 1985). A detailed discussion about such specifications can be found in Mirone (2017) or Lunde et al. (2017). In this section, we make the simple assumption that the short rate process r is *deterministic*. This assumption seems to be overly strict as it completely disregards the stochasticity of the short rate. However, simulation results in Section 3 suggest that assuming r to be deterministic might be innocuous and, more importantly, this assumption allows us to obtain estimators easy to use and, at the same time, more efficient. When the short rate is deterministic, Equation (6) leads to

$$Y_s = \mathbb{E}^{\mathbb{Q}}[X_T e^{-\int_0^T r_u \, du} \,|\, \mathscr{F}_s] e^{\int_0^T r_u du} = X_s e^{-\int_0^s r_u du} e^{\int_0^T r_u du} = X_s e^{\int_s^T r_u du}. \tag{7}$$

Taking logarithms on both sides of (7) we see that the log-prices F and S are linked through

$$F_s = S_s + \int_s^T r_u \mathrm{d}u \tag{8}$$

A straightforward consequence of Equation (8), is that the integrated variances, or equivalently, the quadratic variations of F and S coincide; i.e.,

$$IV_t(F) = [F, F]_t = [S, S]_t = IV_t(S). \tag{9}$$

Similar evidence has also been studied in Rossi and Santucci de Magistris (2013), where the authors assume the interest rate to have negligible variation.

An immediate corollary of (8) is that, $IV_t(F)$ and $IV_t(S)$ are also identical to the quadratic covariation between F and S, that is,

$$IV_t(F) = IV_t(S) = [F, S]_t. \tag{10}$$

Hence, any estimator of the integrated variance of the futures or the quadratic covariation between the underlying and a futures written on it also serves as an estimator of the integrated variance of the underlying asset, and vice versa.

2.2 Infeasible noise reduction by virtue of futures

Certainly, estimation of $IV_t(F)$ is complicated by the presence of a further source of noise proper of the futures market (possibly correlated with the noise process affecting S). Nonetheless, a reliable estimator can be constructed for $IV_t(F)$ directly. Yet, we propose to tackle the problem from another perspective. Equations (8) and (9) illustrate the close relation between an asset and its corresponding futures contracts. Naturally, noises contaminating the futures price and the underlying asset price, namely $\varepsilon^{(F)}$ and $\varepsilon^{(S)}$ respectively, might also be closely related. Consider the extreme but heuristic case where $\varepsilon^{(F)}$ and $\varepsilon^{(S)}$ are perfectly negatively correlated, that is, $\rho(\varepsilon_t^{(F)}, \varepsilon_t^{(S)}) = -1$. In this case, the series $Z = \frac{1}{2}S^\star + \frac{1}{2}F^\star$ is *noise-free* and has exactly the same integrated variance as S (and F), i.e., $IV_t(Z) = IV_t(S)$. The integrated variance of S can then be consistently estimated by $RV_t(Z)$, the realized variance of Z, which is also the most straightforward estimator available.

No doubt perfect negative correlation between $\varepsilon^{(S)}$ and $\varepsilon^{(F)}$ is an unrealistic assumption. Nevertheless, in the forgoing paragraph we investigate the possibility of *noise reduction* through a simple linear combination of S and F. We formulate the idea as follows. Let $Z(\alpha) = \alpha S + (1 - \alpha)F$, which is the latent efficient price of a "portfolio" consisting of only the underlying asset and the futures, with α shares invested in the former and

$(1 - \alpha)$ shares in the latter. Correspondingly, the observed "noisy" price of this portfolio is $Z^{\star}(\alpha) = \alpha S^{\star} + (1-\alpha)F^{\star}$, where the noise contaminating $Z(\alpha)$ is $\varepsilon^{(Z(\alpha))} = \alpha\varepsilon^{(S)} + (1-\alpha)\varepsilon^{(F)}$. Evidently, for any share $\alpha \in \mathbb{R}$, the integrated variance of $Z_t(\alpha)$ equals $IV_t(S)$. Hence, to properly tackle the noise, the bottom line is in the appropriate selection of α.

We start by noting that the variance of $\varepsilon_t^{(Z(\alpha))}$ is given by

$$\sigma_{\varepsilon_t^{(Z(\alpha))}}^2 = \alpha^2\sigma_{\varepsilon_t^{(S)}}^2 + (1-\alpha)^2\sigma_{\varepsilon_t^{(F)}}^2 + 2\alpha(1-\alpha)\rho\sigma_{\varepsilon_t^{(S)}}\sigma_{\varepsilon_t^{(F)}}, \tag{11}$$

where $\sigma_{\varepsilon_t^{(S)}}^2$ and $\sigma_{\varepsilon_t^{(F)}}^2$ denote the variance of $\varepsilon_t^{(S)}$ and $\varepsilon_t^{(F)}$, respectively, and ρ is the correlation coefficient between $\varepsilon_t^{(S)}$ and $\varepsilon_t^{(F)}$. We can then rearrange the terms to express $\sigma_{\varepsilon_t^{(Z(\alpha))}}^2$ as an upward opening parabolic function of α

$$\sigma_{\varepsilon_t^{(Z(\alpha))}}^2 = (\sigma_{\varepsilon_t^{(S)}}^2 + \sigma_{\varepsilon_t^{(F)}}^2 - 2\rho\sigma_{\varepsilon_t^{(S)}}\sigma_{\varepsilon_t^{(F)}})\alpha^2 + 2(\rho\sigma_{\varepsilon_t^{(S)}}\sigma_{\varepsilon_t^{(F)}} - \sigma_{\varepsilon_t^{(F)}}^2)\alpha + \sigma_{\varepsilon_t^{(F)}}^2, \tag{12}$$

which is uniquely minimized at its vertex

$$\alpha^{\dagger} = \frac{\sigma_{\varepsilon_t^{(F)}}^2 - \rho\sigma_{\varepsilon_t^{(S)}}\sigma_{\varepsilon_t^{(F)}}}{\sigma_{\varepsilon_t^{(S)}}^2 + \sigma_{\varepsilon_t^{(F)}}^2 - 2\rho\sigma_{\varepsilon_t^{(S)}}\sigma_{\varepsilon_t^{(F)}}}. \tag{13}$$

The minimum of $\sigma_{\varepsilon_t^{(Z(\alpha))}}^2$ is given by

$$\sigma_{\varepsilon_t^{(Z(\alpha^{\dagger}))}}^2 = \frac{(1 - \rho^2)\sigma_{\varepsilon_t^{(S)}}^2\sigma_{\varepsilon_t^{(F)}}^2}{\sigma_{\varepsilon_t^{(S)}}^2 + \sigma_{\varepsilon_t^{(F)}}^2 - 2\rho\sigma_{\varepsilon_t^{(S)}}\sigma_{\varepsilon_t^{(F)}}}. \tag{14}$$

From (14), we see that as long as $\sigma_{\varepsilon_t^{(S)}}^2 \neq \sigma_{\varepsilon_t^{(F)}}^2$, the minimal variance of the noise $\varepsilon_t^{(Z(\alpha^{\dagger}))}$ tends to 0 as $|\rho|$ tends to 1. It appears evident that due to the effect of the covariance term, the convergence rate is *asymmetric*: $\sigma_{\varepsilon_t^{(Z(\alpha^{\dagger}))}}^2$ tends to 0 more rapidly when $\rho \to -1$ than the case where $\rho \to 1$ as it can easily be inferred by noting that the covariance term in the denominator will increase the overall value of the denominator for $\rho < 0$ and decrease it for $\rho > 0$. As a consequence, the method of noise reduction performs better when $\varepsilon_t^{(S)}$ and $\varepsilon_t^{(F)}$ are negatively correlated. Figure 1 allows to visualize the behavior of Equation (14) for all possible correlation values and for different noise variance ratios. We can clearly see that the asymmetric behavior tends to get more pronounced when the ratio of the two noise variances approaches one; with the extreme case being $\sigma_{\varepsilon_t^{(S)}}^2 = \sigma_{\varepsilon_t^{(F)}}^2$. In such case, the residual noise variance will simply linearly increases in $\rho_{\varepsilon^{(S)},\varepsilon^{(F)}}$ and reach its maximum $\sigma_{\varepsilon_t^{(Z(\alpha^{\dagger}))}}^2 = \sigma_{\varepsilon_t^{(S)}}^2$ for $\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 1$.

Figure 1: Surface plot of the minimum residual variance
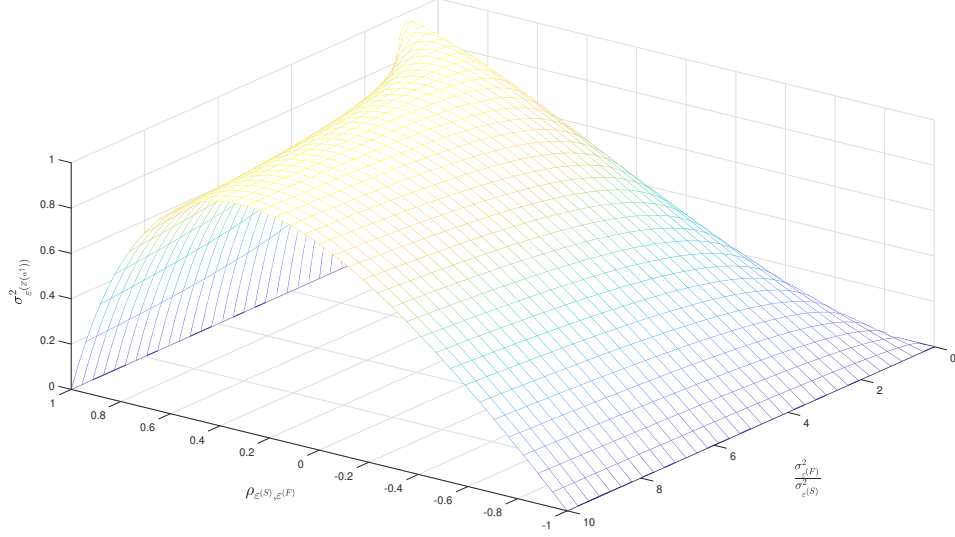
The plot displays the minimum residual variance $(\sigma^2_{\varepsilon_t^{(Z(\alpha^\dagger))}})$ for values of $\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} \in [-1,1]$ and for varying ratios of the two noise variance.

The method now boils down to estimate $IV_t(S)$ on the noise reduced series $Z(\alpha^\dagger)$ which, being affected by a dampened microstructure noise, will (in principle) result in a more precise estimate. In practice the optimal "share", $\alpha^\dagger$, is a latent quantity which has to be estimated, since we have no a priori knowledge about the covariance of $\varepsilon_t^{(S)}$ and $\varepsilon_t^{(F)}$. For this reason we refer to $Z(\alpha^\dagger)$ as the *infeasible* optimal series. Nonetheless, $Z(\alpha^\dagger)$ is of theoretical interest as it can be interpreted as a benchmark, i.e., the *optimal* linear combination with *minimal* noise size. In a simulation study estimation based on the infeasible series can be computed and employed to measure performance of estimation conducted on a feasible series, as we shall present in Section 3.

2.3   Feasible noise reduction by virtue of futures

In this subsection we propose a feasible way to estimate $\alpha^\dagger$ and so construct the series $Z_t(\alpha^\dagger)$. Here, we deliberately refrain from considering the possible implications of the Epps effect and leave the discussion to Section 2.5. A precise estimate of $\alpha^\dagger$ is complicated by the latent nature of the noise processes and requires retrieving both the noise variances $\sigma^2_{\varepsilon_t^{(S)}}$ and $\sigma^2_{\varepsilon_t^{(F)}}$, as well as the noise correlation $\rho_{\varepsilon_t^{(F)},\varepsilon_t^{(S)}}$. With the practitioner in mind we propose a trivial way to estimate these values. As previously mentioned, we know that for any generic asset $X_t^\star$ affected by microstructure noise, as the number of observations, n, goes to infinity,

7

$RV_t(X^\star)$ is not a consistent estimator for $[X, X]_t$. In this context $RV_t(X^\star)$ is found to be a consistent and asymptotically normal estimator of the quantity $2n\mathbb{E}\left[\varepsilon_t^{(X^*)2}\right]$ rather than the quadratic variation of the price process.[3] I.e., when the frequency of observations increases realized volatility does not estimate the true integrated volatility but rather the variance of the contaminating noise process. It immediately follows that

$$\widetilde{\sigma}^2_{\varepsilon_t^{(X)}} = \frac{RV_t(X^\star)}{2n}, \tag{15}$$

which is a biased, but consistent estimator when $n \to \infty$.

Further, a widely employed alternative to $\widetilde{\sigma}^2_{\varepsilon_t^{(X)}}$ exploits the relation between return autocovariance and the noise variance

$$\widehat{\sigma}^2_{\varepsilon_t^{(X)}} = -\frac{r_i^\star r_{i-1}^\star}{n-1}, \tag{16}$$

where $r_i^\star = X_{\tau_i}^\star - X_{(\tau_{i-1})}^\star$. Both estimators are extremely trivial but suffer from the same drawback; they are not consistent for coloured noise processes. While a simple adjustment to $\widehat{\sigma}^2_{\varepsilon_t^{(X)}}$ is indeed possible, it further requires the evaluation of the correlation structure of the noise process which can be quite burdensome (*e.g.* see, Jacod et al. (2016)).[4] As of today, the general consensus in the literature is to simply use one of the estimators presented in Equation (15) and Equation (16). Additionally, in Section 3 we will see that even though do not accounting for autocorrelation of the error term leads to less precise estimates, the feasible estimator will still perform as well as the infeasible one with no appreciable differences between the independent and identically distributed (IID) and autocorrelated noise cases (indifferently of the degree, and relative degree, of autocorrelation of the two noise series). The expansion of Equation (15) and (16) to the multivariate setting is trivial. For simplicity we assume that the two observations are perfectly synchronized (*i.e.* $\tau_i^{(S)} \equiv \tau_i^{(F)} \; \forall \; i \in [0, t]$).[5] Consider the trivial estimator of $[S, F]_t$

$$RC_t(S, F) = \sum_{i}^{n} \left(S_{\tau_i^{(n)}} - S_{\tau_{i-1}^{(n)}}\right) \left(F_{\tau_i^{(n)}} - F_{\tau_{i-1}^{(n)}}\right), \tag{17}$$

which is known as the *realized covariance* and, in the absence of noise, as $n \to \infty$ and $\max_{1 \leq i \leq n} |\tau_i^{(n)} - \tau_{i-1}^{(n)}| \to 0$, $RC_t(S, F) \xrightarrow{\mathbb{P}} [F, S]_t$.

---

[3] A detailed discussion on the argument can be found, among others, in Zhang et al. (2005), Aït-Sahalia and Yu (2009).

[4] The methodology proposed in Jacod et al. (2016) have been investigated and no superior performance has been registered. Additionally, as the estimation of the noise covariance matrix heavily relies on an heuristic approach to tailor some key parameter to the series of interest and can be quite convoluted we decided to disregard this procedure in favour of other, more convenient, estimators. The obtained results are available upon request.

[5] We will discuss how to deal with asynchronous observations in Section 2.5.

$RC_t(S, F)$ suffer an analogous inconsistency due to microstructure noise. Hence, under the previously stated assumptions, given (10)

$$RC_t(S^\star, F^\star) \to IV_t(S) + 2n\mathbb{E}[\Sigma_{\varepsilon_t^{(S)}, \varepsilon_t^{(F)}}], \tag{18}$$

as $n \to \infty$.

Similarly, we have that

$$\mathbb{E}[r_{i,s}^\star * r_{i-1,f}^\star] = -\Sigma_{\varepsilon_t^{(S)}, \varepsilon_t^{(F)}}.$$

Consequently, we can obtain a biased (but consistent) estimate of the covariance of the two noise series as

$$\widetilde{\Sigma}_{\varepsilon_t^{(S)}, \varepsilon_t^{(F)}} = \frac{RC_t(S^\star, F^\star)}{2 * n} \tag{19}$$

or, rather,

$$\widehat{\Sigma}_{\varepsilon_t^{(S)}, \varepsilon_t^{(F)}} = -\frac{1}{n-1} \sum_{i=2}^{n} r_{i,s}^\star * r_{i-1,f}^\star \tag{20}$$

and easily back out the correlation parameter $\rho_{\varepsilon_t^{(S)}, \varepsilon_t^{(F)}}$. It is important to point out that as long as we have consistency, the bias in the estimated variance-covariance matrix will not affect the noise reduction methodology as we are interested in the *ratio* of the two noise variances rather than the exact values.

All the estimators presented in this section have been investigated both in our simulation environment and on the empirical dataset and no relevant difference among them has been evidenced. For the sake of brevity we decide to only report results obtained using the autocovariance estimator (Equation (16) and (20)).

Hence, our estimator for $\alpha^\dagger$ will be

$$\widehat{\alpha} = \frac{\widehat{\sigma}_{\varepsilon_t^{(F)}}^2 - \widetilde{\rho}\widehat{\sigma}_{\varepsilon_t^{(S)}}\widehat{\sigma}_{\varepsilon_t^{(F)}}}{\widehat{\sigma}_{\varepsilon_t^{(S)}}^2 + \widehat{\sigma}_{\varepsilon_t^{(F)}}^2 - 2\widetilde{\rho}\widehat{\sigma}_{\varepsilon_t^{(S)}}\widehat{\sigma}_{\varepsilon_t^{(F)}}}.$$

Consequently we will construct $Z_t(\widehat{\alpha}) = \widehat{\alpha}S_t^\star + (1 - \widehat{\alpha})F_t^\star$.

2.4 Noise reduction through cross-sectional combination over different markets

It is of interest to notice that the methodology discussed in Section 2.2 and Section 2.3 straightforwardly applies to any price series with the same integrated variance. Thus, an immediate expansion would be to use the time series of the same asset traded on different financial markets. Since the microstructure noise is thought to be dependent on the market

microstructure itself and is heavily influenced by factors as liquidity, bid-ask spread, tick size, etc. we can confidently assume that the noise affecting the same asset on two different markets will, to a certain extent, differ. Leading us to the same framework previously discussed for the stock-futures relation. Hence, in this situation, we will generally estimate a noise-reduced time series $Z(\widehat{\alpha}) = \widehat{\alpha}S_i^\star + (1 - \widehat{\alpha})S_j^\star$, with $i \neq j$ identifying two different markets.

$$\widehat{\alpha} = \frac{\widehat{\sigma}^2_{\varepsilon^{(S)}_{t,j}} - \widehat{\rho}\widehat{\sigma}_{\varepsilon^{(S)}_{t,i}}\widehat{\sigma}_{\varepsilon^{(S)}_{t,j}}}{\widehat{\sigma}^2_{\varepsilon^{(S)}_{t,i}} + \widehat{\sigma}^2_{\varepsilon^{(S)}_{t,j}} - 2\widehat{\rho}\widehat{\sigma}_{\varepsilon^{(S)}_{t,i}}\widehat{\sigma}_{\varepsilon^{(S)}_{t,j}}}.$$

Additionally, focusing on different markets opens up to the possibility of smoothing over a larger cross-section to further strengthen the noise reduction. However, liquidity and asynchronicity problems will become increasingly more relevant when working with more than two time series. A detailed discussion of such issues is presented in the following section.

2.5 Synchronization methods and the Epps effect

In this section we consider the case where $S^\star$ and $F^\star$ are observed asynchronously. A required step to implement the procedure presented in Section 2.3 is to deal with the non-synchronous observations in order to be able to estimate the correlation between the two noise processes. We present two methodologies to synchronize the observations and discuss their implications.

We now work with the following framework. Define the observation times of $S^\star$ as $\mathcal{T}_S^n$ and those of $F^\star$ as $\mathcal{T}_F^m$. The elements of $\mathcal{T}_S^n$ and $\mathcal{T}_F^m$ will then be denoted as $\tau_j^{n,S}$ and $\tau_j^{m,F}$ respectively, where the observed sequences are assumed to be independent of the price process and can be irregularly spaced and even random. To ease the notation, we will often suppress the superscript n and m keeping in mind that the number of observations of S and F differ.

2.5.1 Previous-tick sampling

Consider now a set of grid points $(\mathcal{V}_N = \nu_i)_{i=0,...,M}$ with $\nu_i \in [0, T]$ (say, $T = 1$ day), $\nu_0 = 0$, $\nu_T = T$ and M is the sampling frequency. It is common practice to choose $\mathcal{V}_N$ equally spaced. In this case $M = T/\Delta$ with $\Delta = \nu_i - \nu_{i-1}$, for all i being the chosen sampling interval.[6] Once our grid points have been identified, we can define our sampling

---

[6]Typical sampling intervals adopted in the finance literature are 5, 15, 30 minutes and one hour.

points as

$$\tau_{\nu_i}^S = \max\{\tau^S \in \mathcal{T}_n \ : \ \tau^S \leq \nu_i\},$$
$$\tau_{\nu_i}^F = \max\{\tau^F \in \mathcal{T}_n \ : \ \tau^F \leq \nu_i\},$$

that is, for each security we select the largest observation time up to and including the grid point time ν_i as shown in Figure 2. Assumptions may be set to ensure consistency of the realized covariance estimator using previous tick interpolation. These assumptions mainly require the two observed time series to be "liquid enough". [7] We decide not to work under such assumptions but rather investigate the behaviour of the previous tick sampling to different liquidity levels.

Figure 2: Previous Tick sampling
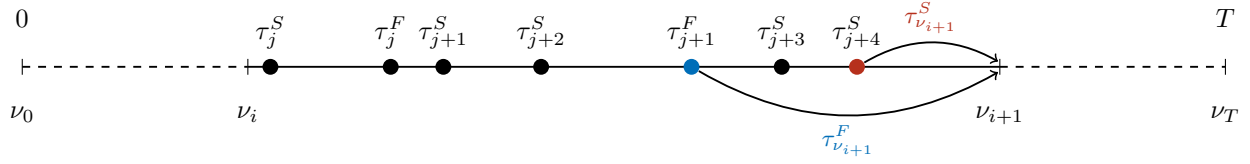
2.5.2 Refresh time sampling

We also consider a second sampling methodology, the refresh time introduced by (Barndorff-Nielsen et al., 2011). In this case we let the data choose the grid points $(\mathcal{V}_N = \nu_i)_{i=0,\dots,M}$. We begin by identifying the first refresh time as $\nu_1 = \max(\tau_1^S, \tau_1^F)$, and then define each subsequent grid point as

$$\nu_{i+1} = \max(\tau_{\tau_{\nu_i}^S+1}^S, \tau_{\tau_{\nu_i}^F+1}^F)$$

Figure 3: Refresh Time sampling
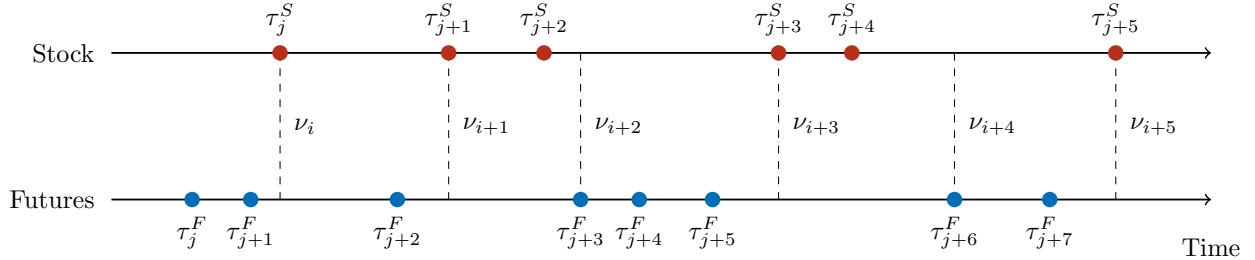
The procedure, as illustrated in Figure 3, leads to an unequally spaced time series where each grid point is taken as soon as all asset prices have been refreshed. In this situation

---

[7]Interested readers are referred to Zhang (2011).

11

we assume that the observed prices are seen exactly at the refreshed times $\mathcal{V}_n$. Unlike the Previous Tick methodology, when synchronizing using Refresh Time, the sample size M depends on the degree of asynchronicity of the observed price series.

2.5.3 The Epps Effect

The empirical evidence that stocks correlation decrease as the sampling frequency increase (that is to say, $\rho_{S,F} \to 0$ as $M \to \infty$) has been initially reported by Epps (1979) using intraday returns data on four US stocks. After Epps' initial work this phenomenon has been named Epps effect and has been subsequently observed in several markets and reported by numerous papers in the literature. Under the previously stated assumptions, it is easy to see that in our framework we have

$$\sigma_Z^2 = \alpha^2\sigma_S^2 + (1-\alpha)^2\sigma_F^2 + 2\alpha(1-\alpha)\rho_{S,F}\sigma_S\sigma_F,$$

where $\rho_{S,F} = 1$ and $\sigma_S = \sigma_F$ by definition and, consequently, $\sigma_Z^2 = \sigma_S^2$. However, asynchronicity of the observations might trigger the Epps effect and reduce the observed correlation between the series ($\rho_{S,F}^\star < 1$). In this situation it is easy to see that $\sigma_Z^2 < \sigma_S^2$ for $\alpha > 0$ and $\sigma_Z^2 > \sigma_S^2$ for $\alpha < 0$. Fortunately, many papers show that the potential bias arising from asynchronicity is strongly correlated to the frequency of the observations (see, *e.g* Hayashi and Yoshida, 2005; Renò, 2003; Zhang, 2011). This implies that for sufficiently liquid series, the Epps effect will be small or even negligible.

In Section 3.1 we present a throughout simulation study and investigate the sensitivity of our methodology to the Epps effect. The overall results are comforting as the bias in the observed correlation between the two assets becomes significant only for extremely illiquid series when the previous tick sampling technique is used. Conversely, the realized covariance constructed using the refresh time procedure demonstrates to provide consistent estimates even when the trading intensity gets very low. In addition to asynchronicity, lead-lag relationships are a second statistical feature of the data which is sometimes suggested as a source of bias in the correlation measurement. In other words, if the correlations between assets are lagged, working at a sampling frequency lower than this lag would lead to undervalued correlations estimates. However, this second source of bias is shown to have a marginal effect compared to the impact of asynchronous observations. Additionally, in our framework we can rule out any lead-lag relationship between the correlation of the two asset prices (although we cannot exclude the possibility of lagged correlation between the two noise processes).

2.6 Estimation of integrated variance with noise-reduced series

Having obtained a feasible "portfolio", $Z_t(\hat{\alpha})$, one may proceed with the estimation of the integrated variance $IV_t(Z)$. The same considerations used in the choice of an estimator when working with a standard price series apply when working with Z_t. The most straightforward estimator is the realized variance RV computed using $Z_t(\hat{\alpha})$, whose signature plot should be more stable than that of $RV_t(S)$. $RV_t(Z(\hat{\alpha}))$ is typically inconsistent, and to achieve consistency one may refer to more sophisticated estimators, such as the realized kernel or pre-averaging estimator. Whichever estimator is selected, the ones computed on $Z_t(\hat{\alpha})$ are expected to outperform the ones computed on S_t directly. In Section 3 we illustrate the difference in performance in a simulation study.

3 Simulation Study

We perform a rich simulation study intended to investigate whether the feasible method proposed in Section 2.3 leads to noise-reduced series and appraise how much gain we can derive from using such series to estimate $IV_t(S)$. To these ends we impose parametric structures on S and r so that S and F can be simulated with closed-form solutions. In contrast to what we assume when developing the method, we *do not* use a deterministic short rate in the simulation. Instead, we model r as a CIR process (Cox et al., 1985), whose driving Brownian motion is independent of the Brownian motions generated on S and σ. Under such setting, a closed-form solution is available for F (see Ramaswamy and Sundaresan (1985) for details or Munk (2011) for a general discussion in the case of affine models). Table 1 reports the dynamics assumed for the simulated price process S_t; a simulation design which allows for realistic features such as jumps in the price process and stochastic volatility with jumps.

Table 1: The model

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

We allow $W_{1,t}$ and $W_{2,t}$ to be correlated and fix $\rho_{W_{1,t}, W_{2,t}} = 0.6$.

Having simulated both S and F and contaminated them with temporally dependent

noises, we apply the feasible method discussed in Section 2.3 to the pair $S^\star$ and $F^\star$, deliberately omitting stochasticity in r. The purpose of such designation is to check whether our misspecified model is still capable of producing noise-reduced series despite considering r to be deterministic. Having constructed $Z(\hat{\alpha})$ by the feasible method, we compute several popular estimators of IV with $Z(\hat{\alpha})$ and investigate whether they outperform estimates computed directly with S and F.

We assume a standard day-trade time of 6.5 trading hours and simulate equally spaced observations at one second interval. We proceed with the following simulation design for $R = 1,000$ replications:

- Generate $N = 23,400$ observations for the underlying, S, and the futures, F, price processes.
- Contaminate both simulated paths with two different noise processes (IID and MA(100)).
  - For both structures we allow cross-sectional correlation of the noise processes affecting S and F through the parameter $\rho_{\varepsilon(S),\varepsilon(F)}$.
- Apply the feasible method to estimate the noise structures and generate the linear combination series $Z(\hat{\alpha})$.
  - We take advantage of the simulation environment to obtain the optimal linear combination, $Z(\alpha^\dagger)$, using the true noise parameters.
- Compute the panel of realized measures and diagnostic tests.

Following Jacod et al. (2016) the autocorrelated noise processes are generated as MA series which approximate fractionally differenced processes. This allows the simulation design to replicate the slow autocorrelation decay empirically evidenced in the aforementioned paper. Throughout the study we allow several other parameters to vary across a wide range of values so as to investigate how the method of noise-reduction is affected by these variations. More specifically, we analyse different trading intensities $\lambda = \{23,400; 7,800; 1,560; 390; 78; 39\}$ and different levels of noise ($\gamma_H = 5; \gamma_M = 2; \gamma_L = 0.5$).[8] Further, as the noise reduction procedure is heavily influenced by the cross-sectional correlation between the two noise processes, we let $\rho_{\varepsilon(S),\varepsilon(F)}$ vary between (-0.9 and 0.9) at 0.1 intervals. Finally, we investigate the sensitivity to different synchronization methodologies; the naive *previous tick* interpolation and the more structured *refresh time* procedure presented in Section 2.5.

As part of our investigation we test the noise reduction performance on a selected group

---

[8]Following Christensen et al. (2014) we compute $\gamma = \sqrt{\frac{\sigma_\varepsilon^2}{IV} * N}$ as a measure of the intensity of the noise process. The reported γ values represent the average between the noise processes affecting S and F.

of popular realized measures. Further, our selection aims to cover different classes of integrated variance estimators. The first estimator we include is the previously discussed realized variance (RV), our main results will investigate the gain on this well known (and biased) estimator as it provides a clear information on the precision gain produced by our cross-sectional noise reduction. The second estimator considered is the optimal sampling RV (RV^*) of Bandi and Russell (2008). The third and fourth estimators are the two-scale realized variance (TSRV) of Zhang et al. (2005) and the multi-scale realized variance (MSRV) of Zhang (2006). The fifth selected realized measure is the realized kernel (RK) (Barndorff-Nielsen et al., 2008) and the sixth is the pre-averaged realized variance (PA) of Jacod et al. (2009).[9] Table 2 and 3 respectively show the root mean square error and normalized integrated variance estimate for different levels of trading intensity and the two considered synchronization methodologies. From the displayed results we can clearly see a marked, consistent improvement in the estimation accuracy for both methodologies and for the different trading intensities considered. Unsurprisingly, the proposed noise reduction methodology becomes less reliable when the number of observations gets too low i.e., $\lambda < 1,560$. Indeed, the construction of $Z(\alpha)$ relies on the cross-sectional correlation between the noise processes affecting S and F, which will undoubtedly be lost if the observations of the two series are temporally far away from each other. Additionally, the tables suggest the refresh time sampling to be more stable across trading intensities and to provide more precise estimations of the integrated variance despite generally showing a higher root mean square error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{R}(\widehat{IV}_i - IV)^2}{R}}.$$

---

[9] A brief discussion on the differences between these estimators and details for their implementation can be found in Mirone (2017).

Table 2: Normalized IV estimates - Varying λ

		Refresh Time						Previous Tick					
		PA	RK	TSRV	MSRV	RV^*	RV5	PA	RK	TSRV	MSRV	RV^*	RV5
All	$Z(\alpha^\dagger)$	1.0245	1.0232	1.3698	1.3522	1.1098	1.0273	1.0245	1.0232	1.3698	1.3522	1.1098	1.0273
	$Z(\hat{\alpha})$	1.0247	1.0234	1.3716	1.3532	1.1127	1.0269	1.0247	1.0234	1.3716	1.3532	1.1127	1.0269
	S	1.0785	1.0678	1.8843	1.6765	1.4736	1.1411	1.0785	1.0678	1.8843	1.6765	1.4736	1.1411
	F	1.1142	1.0985	2.1141	1.8583	1.6597	1.1701	1.1142	1.0985	2.1141	1.8583	1.6597	1.1701
$\lambda = 23400$	$Z(\alpha^\dagger)$	1.0244	1.0236	1.3714	1.3524	1.0957	1.0246	1.0246	1.0233	1.3721	1.3524	1.1107	1.0284
	$Z(\hat{\alpha})$	1.0246	1.0238	1.3735	1.3533	1.0978	1.0246	1.0248	1.0235	1.3737	1.3533	1.1094	1.0279
	S	1.0783	1.0678	1.8817	1.6763	1.4720	1.1414	1.0783	1.0678	1.8817	1.6763	1.4720	1.1414
	F	1.1140	1.0983	2.1126	1.8569	1.6606	1.1701	1.1140	1.0983	2.1126	1.8569	1.6606	1.1701
$\lambda = 7800$	$Z(\alpha^\dagger)$	1.0240	1.0218	0.8863	1.0710	1.0700	0.9419	1.0403	1.0370	2.5830	2.4733	1.0960	1.0376
	$Z(\hat{\alpha})$	1.0240	1.0218	0.8860	1.0711	1.0675	0.9417	1.0404	1.0370	2.5849	2.4774	1.0963	1.0379
	S	1.0473	1.0437	1.2605	1.2416	1.1994	1.1441	1.0473	1.0437	1.2605	1.2416	1.1994	1.1441
	F	1.0672	1.0620	1.3816	1.3235	1.2738	1.1795	1.0672	1.0620	1.3816	1.3235	1.2738	1.1795
$\lambda = 1560$	$Z(\alpha^\dagger)$	1.0134	1.0075	0.5900	0.9968	1.0640	0.9144	1.0764	1.0645	1.1450	1.7423	1.0687	1.0326
	$Z(\hat{\alpha})$	1.0134	1.0074	0.5901	0.9964	1.0508	0.9149	1.0762	1.0644	1.1442	1.7412	1.0793	1.0327
	S	1.0207	1.0191	0.8057	1.0486	1.0626	1.1391	1.0207	1.0191	0.8057	1.0486	1.0626	1.1391
	F	1.0288	1.0280	0.8650	1.0746	1.0866	1.1737	1.0288	1.0280	0.8650	1.0746	1.0866	1.1737
$\lambda = 390$	$Z(\alpha^\dagger)$	1.0192	0.9859	0.5086	0.9898	1.0356	0.0190	0.9505	0.9505	0.4380	0.8795	1.0389	0.9392
	$Z(\hat{\alpha})$	1.0194	0.9864	0.5085	0.9912	1.0474	0.0192	0.9487	0.9488	0.4360	0.8759	1.0389	0.9370
	S	1.0163	1.0120	0.5892	1.0007	1.0290	1.1363	1.0163	1.0120	0.5892	1.0007	1.0290	1.1363
	F	1.0177	1.0138	0.6226	1.0083	1.0138	1.1689	1.0177	1.0138	0.6226	1.0083	1.0138	1.1689
$\lambda = 78$	$Z(\alpha^\dagger)$	0.9944	0.9508	0.4915	1.0183	1.0393	-	0.6455	0.6517	0.2861	0.5745	0.8690	0.7176
	$Z(\hat{\alpha})$	0.9914	0.9462	0.4888	1.0143	1.0084	-	0.6341	0.6398	0.2804	0.5632	0.8733	0.7057
	S	0.9838	0.9652	0.4980	0.9739	0.9696	1.0547	0.9838	0.9652	0.4980	0.9739	0.9696	1.0547
	F	0.9840	0.9713	0.5013	0.9705	0.9893	1.0803	0.9840	0.9713	0.5013	0.9705	0.9893	1.0803
$\lambda = 39$	$Z(\alpha^\dagger)$	0.9401	0.8734	0.5038	0.9480	0.9485	-	0.5598	0.5667	0.2655	0.5317	0.7143	0.6076
	$Z(\hat{\alpha})$	0.9260	0.8607	0.4936	0.9425	0.9479	-	0.5438	0.5499	0.2578	0.5162	0.7230	0.5908
	S	0.9643	0.9274	0.4951	0.9560	0.9759	0.9917	0.9643	0.9274	0.4951	0.9560	0.9759	0.9917
	F	0.9660	0.9261	0.4910	0.9583	0.9611	1.0079	0.9660	0.9261	0.4910	0.9583	0.9611	1.0079

Where $\{\varepsilon_t^{(S)}\}$ and $\{\varepsilon_t^{(F)}\}$ are $MA(100)$ processes and we fix: $\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.4$, $\gamma_H = 5$, $\sigma^2_{\varepsilon^{(S)}} = 3.3985e^{-7}$, $\sigma^2_{\varepsilon^{(F)}} = 4.3215e^{-7}$.

Table 3: RMSE (e^{-3})

|		Refresh Time						Previous Tick					
		PA	RK	TSRV	MSRV	RV^*	RV5	PA	RK	TSRV	MSRV	RV^*	RV5
All	$Z(\alpha^\dagger)$	0.0275	0.0275	0.1343	0.1282	0.0581	0.0511	0.0275	0.0275	0.1343	0.1282	0.0581	0.0511
	$Z(\hat{\alpha})$	0.0276	0.0276	0.1349	0.1286	0.0584	0.0511	0.0276	0.0276	0.1349	0.1286	0.0584	0.0511
	S	0.0430	0.0423	0.3215	0.2464	0.1775	0.0850	0.0430	0.0423	0.3215	0.2464	0.1775	0.0850
	F	0.0532	0.0508	0.4047	0.3123	0.2440	0.0930	0.0532	0.0508	0.4047	0.3123	0.2440	0.0930
$\lambda = 23400$	$Z(\alpha^\dagger)$	0.0276	0.0275	0.1348	0.1282	0.0588	0.0566	0.0276	0.0275	0.1351	0.1283	0.0604	0.0512
	$Z(\hat{\alpha})$	0.0276	0.0275	0.1355	0.1286	0.0592	0.0567	0.0276	0.0276	0.1357	0.1286	0.0581	0.0512
	S	0.0430	0.0423	0.3206	0.2464	0.1773	0.0852	0.0430	0.0423	0.3206	0.2464	0.1773	0.0852
	F	0.0531	0.0507	0.4042	0.3117	0.2445	0.0928	0.0531	0.0507	0.4042	0.3117	0.2445	0.0928
$\lambda = 7800$	$Z(\alpha^\dagger)$	0.0412	0.0390	0.0446	0.0346	0.0745	0.1103	0.0297	0.0292	0.5709	0.5332	0.0646	0.0537
	$Z(\hat{\alpha})$	0.0412	0.0391	0.0447	0.0346	0.0726	0.1103	0.0297	0.0292	0.5715	0.5348	0.0670	0.0538
	S	0.0450	0.0451	0.0972	0.0922	0.0932	0.0864	0.0450	0.0451	0.0972	0.0922	0.0932	0.0864
	F	0.0494	0.0494	0.1405	0.1213	0.1160	0.0951	0.0494	0.0494	0.1405	0.1213	0.1160	0.0951
$\lambda = 1560$	$Z(\alpha^\dagger)$	0.0655	0.0589	0.1502	0.0360	0.1751	0.2546	0.0391	0.0368	0.0662	0.2682	0.0782	0.0493
	$Z(\hat{\alpha})$	0.0655	0.0589	0.1502	0.0361	0.1741	0.2547	0.0391	0.0368	0.0660	0.2678	0.0781	0.0493
	S	0.0640	0.0597	0.0752	0.0471	0.1173	0.0846	0.0640	0.0597	0.0752	0.0471	0.1173	0.0846
	F	0.0643	0.0609	0.0570	0.0511	0.1075	0.0935	0.0643	0.0609	0.0570	0.0511	0.1075	0.0935
$\lambda = 390$	$Z(\alpha^\dagger)$	0.1086	0.0878	0.1790	0.0766	0.1845	0.3565	0.0332	0.0333	0.2026	0.0475	0.1128	0.0549
	$Z(\hat{\alpha})$	0.1087	0.0879	0.1790	0.0767	0.1877	0.3565	0.0336	0.0337	0.2033	0.0489	0.1134	0.0555
	S	0.0938	0.0821	0.1526	0.0703	0.1797	0.0856	0.0938	0.0821	0.1526	0.0703	0.1797	0.0856
	F	0.0934	0.0824	0.1418	0.0734	0.1662	0.0933	0.0934	0.0824	0.1418	0.0734	0.1662	0.0933
$\lambda = 78$	$Z(\alpha^\dagger)$	0.1755	0.1523	0.1965	0.1915	0.3036	0.3601	0.1379	0.1362	0.2579	0.1585	0.1178	0.1212
	$Z(\hat{\alpha})$	0.1747	0.1511	0.1970	0.1905	0.3034	0.3601	0.1415	0.1399	0.2599	0.1624	0.1154	0.1246
	S	0.1476	0.1287	0.1900	0.1518	0.2405	0.0924	0.1476	0.1287	0.1900	0.1518	0.2405	0.0924
	F	0.1484	0.1317	0.1900	0.1553	0.2497	0.1005	0.1484	0.1317	0.1900	0.1553	0.2497	0.1005
$\lambda = 39$	$Z(\alpha^\dagger)$	0.1852	0.1783	0.2017	0.2381	0.2866	0.3601	0.1674	0.1654	0.2656	0.1756	0.1429	0.1533
	$Z(\hat{\alpha})$	0.1876	0.1804	0.2046	0.2382	0.2874	0.3601	0.1729	0.1712	0.2684	0.1810	0.1435	0.1595
	S	0.1798	0.1716	0.2009	0.2177	0.2675	0.1199	0.1798	0.1716	0.2009	0.2177	0.2675	0.1199
	F	0.1788	0.1679	0.2018	0.2157	0.2656	0.1230	0.1788	0.1679	0.2018	0.2157	0.2656	0.1230

Where $\{\varepsilon_t^{(S)}\}$ and $\{\varepsilon_t^{(F)}\}$ are $MA(100)$ processes and we fix: $\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.4$, $\gamma_H = 5$, $\sigma^2_{\varepsilon^{(S)}} = 3.3985e^{-7}$, $\sigma^2_{\varepsilon^{(F)}} = 4.3215e^{-7}$.

Figure 4: Minimum residual variance for $\frac{\sigma^2_{\varepsilon(F)}}{\sigma^2_{\varepsilon(S)}} = 1.27$
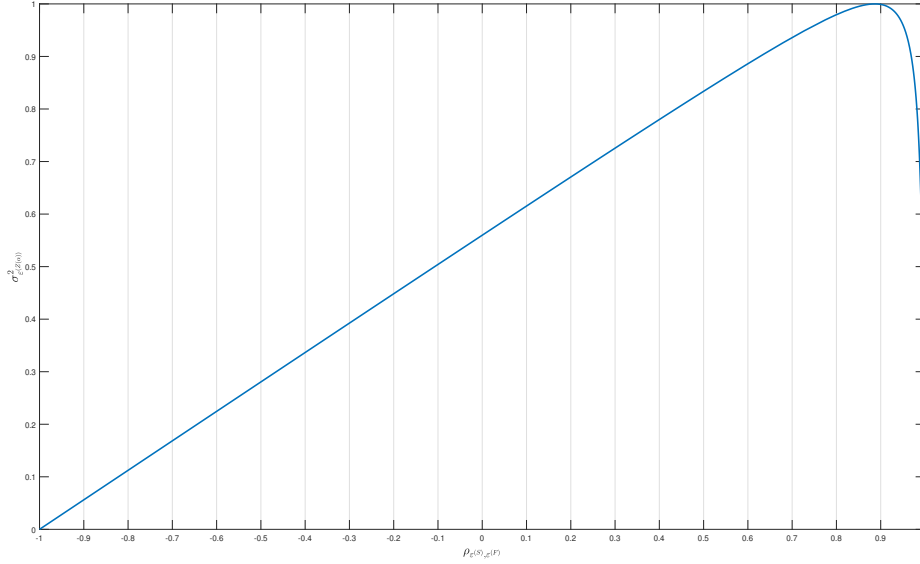

The figure shows the minimum residual variance (expressed as a ratio over $\sigma^2_{\varepsilon(S)}$) for varying values of $\rho_{\varepsilon(S),\varepsilon(F)}$. For given noise variance values $\sigma^2_{\varepsilon(S)} = 3.3985e^{-7}$ and $\sigma^2_{\varepsilon(F)} = 4.3215e^{-7}$.

We investigate the sensitivity of the noise reduction procedure for different levels of correlation between the two noise processes in Table 4 and 5. From a quick look emerges that the residual bias in the estimated values tends to increase for increasing values of $\rho_{\varepsilon(S),\varepsilon(F)}$, reaching its maximum at $\rho_{\varepsilon(S),\varepsilon(F)} = 0.9$. Similarly, the estimation error variation tends to get wider as reported in Table 5. Nonetheless, regardless of the correlation value, the results strikingly show that both $Z(\alpha)$ and $Z(\alpha^{\dagger})$ provide more precise IV estimates compared to the values obtained using the same estimators on S and F. Figure 3 helps to interpret our findings as it shows the asymmetric behaviour of $\sigma^2_{\varepsilon_t^{(Z(\alpha^{\dagger}))}}$ for varying noise correlation values fixing the noise variances to the value selected in our simulation. As the ratio between $\sigma^2_{\varepsilon_t^{(S)}}$ and $\sigma^2_{\varepsilon_t^{(F)}}$ is close to the unit value (1.2716) the asymmetry of $\sigma^2_{\varepsilon_t^{(Z(\alpha^{\dagger}))}}$ is extremely marked with the function increasing almost linearly in $\rho_{\varepsilon(S),\varepsilon(F)}$ until a sharp drop after $\rho_{\varepsilon(S),\varepsilon(F)} = 0.9$. The practical effect of such behaviour can be seen in Figure 5, which displays the stabilization of the signature plot of the series obtained through noise reduction. Increasing values of $\rho_{\varepsilon(S),\varepsilon(F)}$ have the effect of reducing the negative downward slope exhibited by the realized covariance estimate and, consequently, the efficiency of the noise reduction. This effect can be further apprised from Table 6, where the increase in the residual noise variance for increasing noise correlation values appears evident. Even so, these results are of no concern as they show anew that only in the worst case scenario

considered ($\rho_{\varepsilon^{(S)}, \varepsilon^{(F)}} = 0.9$) the noise reduction fails to be significant. Consequently, the feasible methodology can be confidently implemented without the risk of worsening the obtained estimates.

Figure 5: Signature plot for different values of $\rho_{\varepsilon^{(S)}, \varepsilon^{(F)}}$
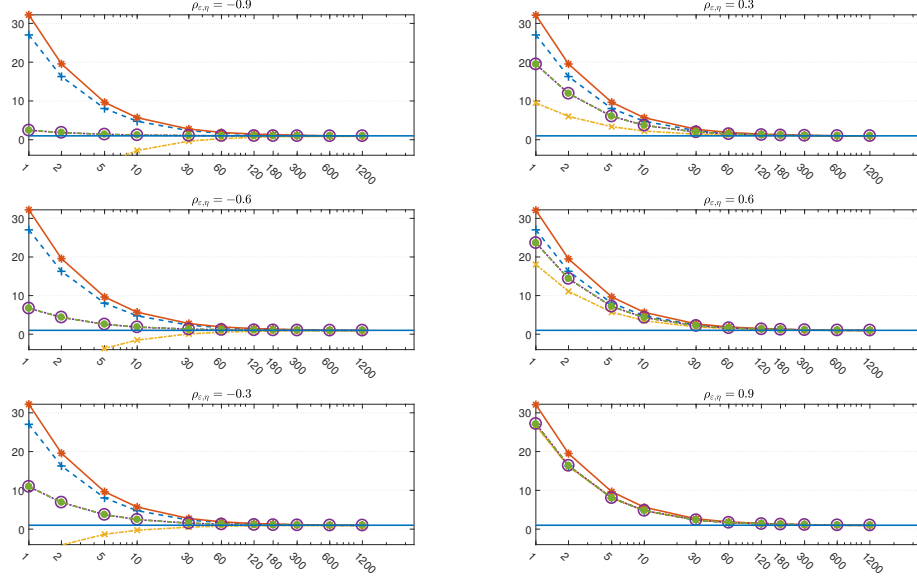
 The figure shows the classic signature plot. We report the realized variance estimates (on the Y axis) for different sampling frequencies (on the X axis). The reported series are: $RV_{Z(\hat{\alpha})}$ ($\bigcirc\!-\!\cdot\bigcirc\!-\!\cdot\bigcirc$), $RV_{Z(\alpha^\dagger)}$ ($\bullet\!-\!-\!\bullet\!-\!-\!\bullet$), RC ($\times\!\cdot\times\!\cdot\times$), RV_S ($+\!-\!-\!+\!-\!-\!+$), RV_F ($*\!-\!*\!-\!*$) for varying values of $\rho_{\varepsilon^{(S)}, \varepsilon^{(F)}}$.

Table 4: Normalized IV estimates - Varying $\rho_{\varepsilon^{(S)},\varepsilon^{(F)}}$

		PA	RK	TSRV	MSRV	RV^*	RV5
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.9$	$Z(\alpha^\dagger)$	1.0055	1.0062	0.7945	1.0934	1.0773	0.9870
	$Z(\hat{\alpha})$	1.0058	1.0064	0.7957	1.0945	1.0735	0.9867
	S	1.0786	1.0664	1.8817	1.6728	1.4780	1.1276
	F	1.1127	1.0950	2.1167	1.8608	1.6596	1.1665
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.7$	$Z(\alpha^\dagger)$	1.0131	1.0130	1.1131	1.2201	1.0713	1.0023
	$Z(\hat{\alpha})$	1.0133	1.0132	1.1142	1.2211	1.0589	1.0020
	S	1.0786	1.0664	1.8817	1.6728	1.4780	1.1276
	F	1.1122	1.0946	2.1143	1.8626	1.6627	1.1664
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.5$	$Z(\alpha^\dagger)$	1.0210	1.0201	1.2961	1.3122	1.0826	1.0191
	$Z(\hat{\alpha})$	1.0212	1.0203	1.2980	1.3132	1.0889	1.0187
	S	1.0786	1.0664	1.8817	1.6728	1.4780	1.1276
	F	1.1119	1.0943	2.1152	1.8628	1.6672	1.1663
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.3$	$Z(\alpha^\dagger)$	1.0291	1.0273	1.4334	1.3875	1.1142	1.0367
	$Z(\hat{\alpha})$	1.0293	1.0275	1.4353	1.3886	1.1133	1.0362
	S	1.0786	1.0664	1.8817	1.6728	1.4780	1.1276
	F	1.1117	1.0941	2.1161	1.8608	1.6676	1.1662
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.1$	$Z(\alpha^\dagger)$	1.0373	1.0345	1.5456	1.4575	1.1736	1.0549
	$Z(\hat{\alpha})$	1.0375	1.0346	1.5481	1.4587	1.1850	1.0544
	S	1.0786	1.0664	1.8817	1.6728	1.4780	1.1276
	F	1.1115	1.0939	2.1160	1.8595	1.6656	1.1660
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.1$	$Z(\alpha^\dagger)$	1.0455	1.0417	1.6422	1.5256	1.2298	1.0737
	$Z(\hat{\alpha})$	1.0457	1.0418	1.6439	1.5261	1.2265	1.0732
	S	1.0786	1.0664	1.8817	1.6728	1.4780	1.1276
	F	1.1114	1.0938	2.1158	1.8596	1.6653	1.1659
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.3$	$Z(\alpha^\dagger)$	1.0538	1.0487	1.7273	1.5735	1.3012	1.0930
	$Z(\hat{\alpha})$	1.0540	1.0489	1.7295	1.5755	1.3004	1.0925
	S	1.0786	1.0664	1.8817	1.6728	1.4780	1.1276
	F	1.1113	1.0937	2.1148	1.8595	1.6635	1.1657
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.5$	$Z(\alpha^\dagger)$	1.0619	1.0556	1.8030	1.6233	1.3557	1.1128
	$Z(\hat{\alpha})$	1.0622	1.0559	1.8046	1.6252	1.3496	1.1123
	S	1.0786	1.0664	1.8817	1.6728	1.4780	1.1276
	F	1.1113	1.0938	2.1150	1.8584	1.6607	1.1656
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.7$	$Z(\alpha^\dagger)$	1.0694	1.0620	1.8649	1.6645	1.4270	1.1327
	$Z(\hat{\alpha})$	1.0698	1.0623	1.8715	1.6700	1.4230	1.1322
	S	1.0786	1.0664	1.8817	1.6728	1.4780	1.1276
	F	1.1114	1.0939	2.1144	1.8590	1.6627	1.1656
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.9$	$Z(\alpha^\dagger)$	1.0722	1.0643	1.8754	1.6623	1.4865	1.1487
	$Z(\hat{\alpha})$	1.0732	1.0652	1.8850	1.6718	1.4885	1.1484
	S	1.0786	1.0664	1.8817	1.6728	1.4780	1.1276
	F	1.1117	1.0942	2.1149	1.8581	1.6567	1.1659

Here $\{\varepsilon_t^{(S)}\}$ and $\{\varepsilon_t^{(F)}\}$ are $MA(100)$ processes and we fix: $\gamma_H = 5$, $\sigma_{\varepsilon^{(S)}}^2 = 3.3985e^{-7}$, $\sigma_{\varepsilon^{(F)}}^2 = 4.3215e^{-7}$.

Table 5: RMSE - Varying $\rho_{\varepsilon^{(S)},\varepsilon^{(F)}}$

		PA	RK	TSRV	MSRV	RV^*	RV5
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.9$	$Z(\alpha^\dagger)$	0.0262	0.0264	0.0763	0.0353	0.0705	0.0500
	$Z(\hat\alpha)$	0.0263	0.0265	0.0759	0.0356	0.0694	0.0501
	S	0.0426	0.0413	0.3203	0.2450	0.1787	0.0787
	F	0.0516	0.0484	0.4053	0.3127	0.2440	0.0889
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.7$	$Z(\alpha^\dagger)$	0.0265	0.0267	0.0421	0.0806	0.0604	0.0498
	$Z(\hat\alpha)$	0.0266	0.0267	0.0425	0.0809	0.0577	0.0499
	S	0.0426	0.0413	0.3203	0.2450	0.1787	0.0787
	F	0.0514	0.0482	0.4044	0.3134	0.2452	0.0889
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.5$	$Z(\alpha^\dagger)$	0.0271	0.0272	0.1075	0.1137	0.0589	0.0504
	$Z(\hat\alpha)$	0.0272	0.0272	0.1082	0.1141	0.0587	0.0504
	S	0.0426	0.0413	0.3203	0.2450	0.1787	0.0787
	F	0.0513	0.0481	0.4048	0.3136	0.2467	0.0890
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.3$	$Z(\alpha^\dagger)$	0.0280	0.0279	0.1571	0.1409	0.0608	0.0518
	$Z(\hat\alpha)$	0.0280	0.0279	0.1578	0.1413	0.0589	0.0518
	S	0.0426	0.0413	0.3203	0.2450	0.1787	0.0787
	F	0.0512	0.0481	0.4050	0.3128	0.2468	0.0891
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.1$	$Z(\alpha^\dagger)$	0.0291	0.0288	0.1979	0.1663	0.0759	0.0542
	$Z(\hat\alpha)$	0.0292	0.0289	0.1988	0.1667	0.0785	0.0541
	S	0.0426	0.0413	0.3203	0.2450	0.1787	0.0787
	F	0.0512	0.0480	0.4051	0.3124	0.2460	0.0892
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.1$	$Z(\alpha^\dagger)$	0.0305	0.0300	0.2331	0.1909	0.0917	0.0574
	$Z(\hat\alpha)$	0.0306	0.0300	0.2338	0.1911	0.0913	0.0573
	S	0.0426	0.0413	0.3203	0.2450	0.1787	0.0787
	F	0.0512	0.0480	0.4050	0.3124	0.2462	0.0893
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.3$	$Z(\alpha^\dagger)$	0.0322	0.0313	0.2639	0.2080	0.1187	0.0614
	$Z(\hat\alpha)$	0.0322	0.0313	0.2647	0.2088	0.1185	0.0612
	S	0.0426	0.0413	0.3203	0.2450	0.1787	0.0787
	F	0.0512	0.0481	0.4047	0.3124	0.2452	0.0894
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.5$	$Z(\alpha^\dagger)$	0.0340	0.0327	0.2915	0.2265	0.1351	0.0662
	$Z(\hat\alpha)$	0.0341	0.0328	0.2920	0.2272	0.1334	0.0660
	S	0.0426	0.0413	0.3203	0.2450	0.1787	0.0787
	F	0.0513	0.0481	0.4049	0.3121	0.2448	0.0895
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.7$	$Z(\alpha^\dagger)$	0.0359	0.0342	0.3138	0.2412	0.1605	0.0718
	$Z(\hat\alpha)$	0.0359	0.0343	0.3163	0.2432	0.1599	0.0715
	S	0.0426	0.0413	0.3203	0.2450	0.1787	0.0787
	F	0.0515	0.0482	0.4046	0.3123	0.2454	0.0894
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.9$	$Z(\alpha^\dagger)$	0.0366	0.0347	0.3176	0.2403	0.1803	0.0772
	$Z(\hat\alpha)$	0.0369	0.0350	0.3211	0.2437	0.1816	0.0769
	S	0.0426	0.0413	0.3203	0.2450	0.1787	0.0787
	F	0.0517	0.0485	0.4050	0.3120	0.2429	0.0894

Here $\{\varepsilon_t^{(S)}\}$ and $\{\varepsilon_t^{(F)}\}$ are $MA(100)$ processes and we fix: $\gamma_H = 5$, $\sigma^2_{\varepsilon^{(S)}} = 3.3985e^{-7}$, $\sigma^2_{\varepsilon^{(F)}} = 4.3215e^{-7}$.

Table 6: Residual noise variance (e^-7)

		IID	$MA(100)$	IID	$MA(100)$		
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.9$	$Z(\alpha^\dagger)$	0.1909	0.1909	2.0911	2.0910	$Z(\alpha^\dagger)$	$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.1$
	$Z(\hat{\alpha})$	0.1909	0.1918	2.0910	2.0932	$Z(\hat{\alpha})$	
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.8$	$Z(\alpha^\dagger)$	0.3817	0.3817	2.2789	2.2788	$Z(\alpha^\dagger)$	$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.2$
	$Z(\hat{\alpha})$	0.3817	0.3827	2.2788	2.2813	$Z(\hat{\alpha})$	
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.7$	$Z(\alpha^\dagger)$	0.5724	0.5724	2.4656	2.4655	$Z(\alpha^\dagger)$	$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.3$
	$Z(\hat{\alpha})$	0.5724	0.5735	2.4656	2.4683	$Z(\hat{\alpha})$	
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.6$	$Z(\alpha^\dagger)$	0.7630	0.7630	2.6507	2.6507	$Z(\alpha^\dagger)$	$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.4$
	$Z(\hat{\alpha})$	0.7630	0.7642	2.6507	2.6539	$Z(\hat{\alpha})$	
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.5$	$Z(\alpha^\dagger)$	0.9535	0.9535	2.8333	2.8333	$Z(\alpha^\dagger)$	$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.5$
	$Z(\hat{\alpha})$	0.9535	0.9548	2.8332	2.8371	$Z(\hat{\alpha})$	
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.4$	$Z(\alpha^\dagger)$	1.1438	1.1438	3.0113	3.0114	$Z(\alpha^\dagger)$	$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.6$
	$Z(\hat{\alpha})$	1.1438	1.1452	3.0113	3.0160	$Z(\hat{\alpha})$	
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.3$	$Z(\alpha^\dagger)$	1.3340	1.3340	3.1807	3.1808	$Z(\alpha^\dagger)$	$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.7$
	$Z(\hat{\alpha})$	1.3340	1.3354	3.1806	3.1867	$Z(\hat{\alpha})$	
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.2$	$Z(\alpha^\dagger)$	1.5238	1.5237	3.3285	3.3288	$Z(\alpha^\dagger)$	$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.8$
	$Z(\hat{\alpha})$	1.5238	1.5254	3.3284	3.3370	$Z(\hat{\alpha})$	
$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = -0.1$	$Z(\alpha^\dagger)$	1.7134	1.7133	3.3948	3.3953	$Z(\alpha^\dagger)$	$\rho_{\varepsilon^{(S)},\varepsilon^{(F)}} = 0.9$
	$Z(\hat{\alpha})$	1.7133	1.7151	3.3947	3.4090	$Z(\hat{\alpha})$	

Residual noise variance for different values of $\rho_{\varepsilon^{(S)},\varepsilon^{(F)}}$ and noise structures. For the autocorrelated case $\{\varepsilon_t^{(S)}\}$ and $\{\varepsilon_t^{(F)}\}$ are $MA(100)$ processes and we fix: $\gamma_H = 5$, $\sigma_{\varepsilon^{(S)}}^2 = 3.3985e^{-7}$, $\sigma_{\varepsilon^{(F)}}^2 = 4.3215e^{-7}$.

Another important factor influencing the estimation behaviour is the noise "intensity", which we define through the noise ratio. Following Christensen et al. (2014) we identify three different levels of noise; a high level $\gamma_H = 5$, a medium level $\gamma_M = 2.5$ and a low level $\gamma_L = 0.5$. We examine the impact of different γs in Table 7 and 8. The results for both γ_H and γ_M are clearly in favour of the noise reduced approach. However, we can notice that lower levels of noise tend to produce more precise estimates on all series, to the point that for γ_L most estimators provide results statistically not significantly different across the four different series. However, both the feasible and infeasible approaches still present a markedly lower RMSE and, as shown in Table 8, they still greatly reduce the noise variance affecting the series.

Table 7: Varying noise ratio (γ)

		PA	RK	TSRV	MSRV	RV^*	RV5
$\gamma_H = 5$	$Z(\alpha^\dagger)$	1.0245	1.0232	1.3698	1.3522	1.1098	1.0273
		(0.0275)	(0.0275)	(0.1343)	(0.1282)	(0.0581)	(0.0511)
	$Z(\hat{\alpha})$	1.0247	1.0234	1.3716	1.3532	1.1127	1.0269
		(0.0276)	(0.0276)	(0.1349)	(0.1286)	(0.0584)	(0.0511)
	S	1.0785	1.0678	1.8843	1.6765	1.4736	1.1411
		(0.0430)	(0.0423)	(0.3215)	(0.2464)	(0.1775)	(0.0850)
	F	1.1142	1.0985	2.1141	1.8583	1.6597	1.1701
		(0.0532)	(0.0508)	(0.4047)	(0.3123)	(0.2440)	(0.0930)
$\gamma_M = 2.5$	$Z(\alpha^\dagger)$	1.0063	1.0068	0.8998	1.1321	1.0437	0.9915
		(0.0261)	(0.0263)	(0.0391)	(0.0491)	(0.0686)	(0.0496)
	$Z(\hat{\alpha})$	1.0064	1.0068	0.9003	1.1324	1.0490	0.9913
		(0.0262)	(0.0263)	(0.0390)	(0.0492)	(0.0680)	(0.0496)
	S	1.0256	1.0226	1.3494	1.3321	1.0892	1.0417
		(0.0327)	(0.0338)	(0.1270)	(0.1214)	(0.0579)	(0.0642)
	F	1.0217	1.0194	1.2432	1.2856	1.0599	1.0244
		(0.0326)	(0.0338)	(0.0885)	(0.1045)	(0.0564)	(0.0619)
$\gamma_L = 0.5$	$Z(\alpha^\dagger)$	1.0021	1.0031	0.5118	1.0051	1.0124	0.9827
		(0.0259)	(0.0261)	(0.1761)	(0.0091)	(0.1030)	(0.0501)
	$Z(\hat{\alpha})$	1.0021	1.0031	0.5118	1.0052	1.0101	0.9826
		(0.0259)	(0.0261)	(0.1761)	(0.0091)	(0.1031)	(0.0501)
	S	0.9999	0.9984	0.5208	1.0077	1.0078	0.9841
		(0.0300)	(0.0310)	(0.1732)	(0.0096)	(0.0918)	(0.0556)
	F	1.0015	0.9998	0.6273	1.0415	1.0079	0.9867
		(0.0300)	(0.0309)	(0.1375)	(0.0179)	(0.0824)	(0.0557)

Normalized IV estimates and RMSE e^{-3} in parenthesis. Here $\{\varepsilon_t^{(S)}\}$ and $\{\varepsilon_t^{(F)}\}$ are $MA(100)$ processes and the noise variances are respectively: $\sigma^2_{\varepsilon(S)} = 3.3985e^{-7}$, $\sigma^2_{\varepsilon(F)} = 4.3215e^{-7}$, $\sigma^2_{\varepsilon(S)} = 1.122e^{-7}$, $\sigma^2_{\varepsilon(F)} = 0.841e^{-7}$, $\sigma^2_{\varepsilon(S)} = 1.385e^{-9}$, $\sigma^2_{\varepsilon(F)} = 7.538e^{-9}$.

In order to assess the performance of Equation 20 under a colored noise structure, the reported results have mainly focused on the autocorrelated noise case. As final step of our analysis we investigate whether the noise reduction methodology exhibits major differences in the case of IID noise. Table 6 and Table 8 report our findings and allow for a comparison between the IID and coloured noise cases. A glance at the tables shows that the difference between the two noise structures is indeed negligible. Consequently, in empirical applications, disregarding the autocorrelated structure of the microstructure noise series would not hinder the noise reduction performance. Therefore, the correct assessment of the actual noise structure appear to be a minor concern in the adoption of the feasible noise reduction methodology.

Table 8: Residual noise variance (e^{-7}) varying γ

		γ_H	γ_M	γ_L
$MA(100)$	$\sigma^2_{\varepsilon^{Z(\alpha\dagger)}}$	1.1438	0.2840	0.0076
	$\sigma^2_{\varepsilon^{Z(\hat{\alpha})}}$	1.1452	0.2843	0.0076
	$\sigma^2_{\varepsilon^{(S)}}$	4.3215	1.1215	0.0138
	$\sigma^2_{\varepsilon^{(F)}}$	3.3985	0.8138	0.0754
IID	$\sigma^2_{\varepsilon^{Z(\alpha\dagger)}}$	1.1435	0.2839	0.0076
	$\sigma^2_{\varepsilon^{Z(\hat{\alpha})}}$	1.1435	0.2839	0.0076
	$\sigma^2_{\varepsilon^{(S)}}$	4.3215	1.1215	0.0138
	$\sigma^2_{\varepsilon^{(F)}}$	3.3985	0.8138	0.0754

Residual noise variances for different noise structures and noise intensities. For the coloured noise case $\{\varepsilon_t^{(S)}\}$ and $\{\varepsilon_t^{(F)}\}$ are $MA(100)$ processes. We set: $\gamma_H = 5$, $\sigma^2_{\varepsilon^{(S)}} = 3.3985e^{-7}$, $\sigma^2_{\varepsilon^{(F)}} = 4.3215e^{-7}$.

3.1 Investigating the Epps effect

In this section we expand our simulation analysis to investigate the impact of the Epps effect on the noise reduction methodology. The simulation design closely follows the one introduced in the previous section but we will modify some key setting to analyse the behavior in extremely illiquid scenarios. We will simulate a true latent process generating at unobservable frequencies (often, as in the present paper, the timestamp precision of empirical datasets is a single second). To this extent we will consider a single trading day but generate the process with an increasingly shrinking interval between observations, dt, of one second, $\frac{1}{10}$ of a second, $\frac{1}{100}$ of a second and one millisecond. This setup corresponds to a latent process with an increasingly higher number of daily realizations $N = \{23,400; 234,000; 2,340,000; 23,400,000\}$. We will then verify the impact of a relatively lower liquidity on our correlation estimates and, consequently, on our noise reduction procedure. We will show that, given the series not being too illiquid, the estimated noise correlation will not be excessively biased and will produce sub-optimal but consistent results. Taking advantage of the simulation environment, we will verify the overall impact of the Epps effect by providing correlation and covariance values of the noise free processes for different trading intensities as well as the average distance between the "observed" realizations of the two series.

Table 9 presents the estimated realized variance (normalized for the true IV) for different trade intensities and processes generated at increasingly high frequencies. From a first look at the table emerges immediately that the previous tick methodology is extremely sensi-

tive to the amount of available observations and the obtained estimates tend to worsen dramatically for both higher frequencies (smaller dt) and lower trading intensities (smaller λ). Nonetheless, we can see that the results are comforting as also in the extreme case of $dt = 1$ millisecond, the trivial realized variance computed on $Z(\hat{\alpha})$ is able to provide precise estimates for sufficiently liquid time series ($\lambda = 7,800$)[10] although its performance deteriorates quickly. For higher values of dt the estimates obtained using the refresh time sampling scheme are even more encouraging as they remain very close to the latent IV for values of λ as low as $1,560$. However, it is important to remark that, as the series approach critical liquidity levels, the effect of microstructure noise becomes marginal and implementing the simple realized variance on the underlying series will likely provide optimal results. Further, as already evinced in the previous section, the results obtained via the feasible noise reduction technique barely differ from the optimal noise reduced series $Z(\alpha^{\dagger})$. This finding hints at the fact that at the lowest trading intensities the correlation between the two series is invariably lost and, even knowing the latent noise structure, it is not possible to obtain any noise reduction through cross-sectional combination. Indeed, inspecting the residual noise variances (expressed as a ratio over $\sigma^2_{\varepsilon(S)}$) displayed in Table 10 we can see that after a striking reduction for very high level of λ, both for the optimal and the feasible noise reduced series, the residual noise variances register a sharp increase; to the point of becoming larger than the original noise variance. The problem, as expected, is given by the correlation between the two assets fading away as the observations get more and more asynchronous. This fact can be evinced from Table 11, where we can see that, as the distance between the observations upon which the correlation gets computed widens, the correlation (and hence the covariance) quickly fades away. Table 11 also helps explaining the behavior of the noise reduction methodology under the refresh time synchronization scheme; we can see that the correlation reaches a soft bound around 0.65. This grants that Z will not deviate substantially from the true IV, which can then be measured with any appropriate realized measure.

---

[10]It should be noted that $\lambda = 7,800$ corresponds to an incredibly low trading intensity of $\frac{N}{3,000}$ given a "latent" process generating at 1ms frequency.

Table 9: Normalized RV 5min. estimates

		All	$\lambda = 23400$	$\lambda = 7800$	$\lambda = 1560$	$\lambda = 390$	$\lambda = 78$	$\lambda = 39$
	$Z(\alpha^\dagger)$	1.0116	1.0115	1.0442	1.0309	0.9536	0.7182	0.6193
		-	1.0240	1.0069	0.9879	0	0	0
	$Z(\hat{\alpha})$	1.0116	1.0115	1.0442	1.0306	0.9528	0.7108	0.6052
		-	1.0242	1.0069	0.9878	0	0	0
$dt = 1$	S	1.1641	1.1635	1.1652	1.1686	1.1410	1.0738	1.0441
		-	1.1635	1.1652	1.1686	1.1410	1.0738	1.0441
	F	1.1231	1.1234	1.1350	1.1237	1.1215	1.0486	1.0046
		-	1.1234	1.1350	1.1237	1.1215	1.0486	1.0046
	$Z(\alpha^\dagger)$	1.0459	1.0550	1.0200	0.8516	0.6451	0.5374	0.5129
		-	1.0037	1.0272	1.0063	0	0	0
	$Z(\hat{\alpha})$	1.0460	1.0550	1.0200	0.8503	0.6428	0.5372	0.5215
		-	1.0038	1.0273	1.0062	0	0	0
$dt = \frac{1}{10}$	S	1.1814	1.1869	1.1726	1.1438	1.0707	1.0310	0.9923
		-	1.1821	1.1726	1.1438	1.0707	1.0310	0.9923
	F	1.1447	1.1459	1.1420	1.1267	1.0618	1.0021	1.0066
		-	1.1441	1.1420	1.1267	1.0618	1.0021	1.0066
	$Z(\alpha^\dagger)$	1.0514	0.9192	0.7375	0.5642	0.5162	0.5009	0.4802
		-	0.9985	0.9958	0.9788	0	0	0
	$Z(\hat{\alpha})$	1.0515	0.9186	0.7358	0.5614	0.5140	0.5008	0.4830
		-	0.9985	0.9957	0.9692	0	0	0
$dt = \frac{1}{100}$	S	1.1891	1.1775	1.1176	1.0362	1.0055	0.9938	0.9467
		-	1.1765	1.1176	1.0362	1.0055	0.9938	0.9467
	F	1.1494	1.1407	1.0925	1.0225	0.9994	0.9902	0.9550
		-	1.1402	1.0925	1.0225	0.9994	0.9902	0.9550
	$Z(\alpha^\dagger)$	1.0561	0.9243	0.5933	0.5422	0.5177	0.4880	0.0617
		-	1.0454	0.9967	0.8770	0	0	0
	$Z(\hat{\alpha})$	1.0562	0.9237	0.5902	0.5376	0.5122	0.4759	0.0549
		-	1.0455	0.9966	0.8768	0	0	0
$dt = \frac{1}{1000}$	S	1.1928	1.1799	1.0514	1.0294	1.0277	1.0206	0.1700
		-	1.1799	1.0514	1.0294	1.0277	1.0206	0.1700
	F	1.1541	1.1453	1.0389	1.0361	1.0169	0.9336	0.0890
		-	1.1453	1.0389	1.0361	1.0169	0.9336	0.0890

Normalized IV estimates are reported for price processes generated at increasingly high frequencies while the maximum observed frequency is fixed at 1 second. Where $\{\varepsilon_t^{(S)}\}$ and $\{\varepsilon_t^{(F)}\}$ are $MA(100)$ processes and the noise variances are respectively: $\sigma^2_{\varepsilon(S)} = 3.3985e^{-7}$, $\sigma^2_{\varepsilon(F)} = 4.3215e^{-7}$.

Table 10: Residual noise variance - ratio $\frac{\sigma_\varepsilon^2(Z)}{\sigma_\varepsilon^2(S)}$

		All	$\lambda = 23400$	$\lambda = 7800$	$\lambda = 1560$	$\lambda = 390$	$\lambda = 78$	$\lambda = 39$
$dt = 1$	$Z(\alpha^\dagger)$	0.1766	0.1773	0.3720	0.7574	1.9768	7.9081	15.7479
		-	0.1766	0.2751	0.4276	0.6685	1.8129	3.2190
	$Z(\hat\alpha)$	0.1768	0.1776	0.3724	0.7575	1.9741	7.8593	15.5946
		-	0.1768	0.2753	0.4279	0.6714	1.9502	3.8969
$dt = \frac{1}{10}$	$Z(\alpha^\dagger)$	0.2647	0.6342	1.1863	4.3787	16.4470	78.2039	148.2509
		-	0.4154	0.5324	1.0877	3.0521	14.2024	26.4866
	$Z(\hat\alpha)$	0.3369	0.8065	1.5084	5.5638	20.8765	99.7276	191.3655
		-	0.5282	0.6773	1.3888	3.9369	19.1497	37.2321
$dt = \frac{1}{100}$	$Z(\alpha^\dagger)$	0.2647	3.0710	8.4098	40.0425	158.7894	779.3936	1475.2913
		-	0.8635	1.7672	7.1023	27.5569	134.5665	254.3645
	$Z(\hat\alpha)$	0.2650	3.0682	8.3962	40.0020	776.4605	1469.5368	5300.0884
		-	0.8644	1.7634	7.0729	27.8824	153.4087	295.4018
$dt = \frac{1}{1000}$	$Z(\alpha^\dagger)$	0.2647	3.0817	27.4384	82.3834	394.8216	1552.2966	4797.0714
		-	0.8627	1.8243	13.8977	69.7436	280.9457	619.4104
	$Z(\hat\alpha)$	0.2650	3.0780	27.3679	82.6029	391.4762	1552.1506	5300.0884
		-	0.8622	1.8743	14.0745	74.8253	390.8942	1217.6001

We report the residual noise variance of the reduced series expressed as a ratio over the noise variance of the objective series. For each noise reduced series we present the results obtained under the previous tick and refresh time synchronization schemes respectively in the top and bottom row.

Overall the results presented in this section hint that in a situation of extreme illiquidity the joint use of the standard noise covariance estimators (as presented in (19) and (20)) and the previous tick synchronization methodology leads to unreliable estimates as the correlation between the two assets irremediably fades away and so does the true relationship between the two noise processes. Hence, covariance estimators robust to asynchronicity (see, *e.g.* Christensen et al., 2010a; Hayashi and Yoshida, 2005) need to be implemented or a different synchronization scheme (as the refresh time methodology) must be used. However, as evidenced by the results obtained for $Z(\alpha^\dagger)$, even knowing the true noise correlation, if the asynchronicity becomes too marked the construction of a noise reduced series simply will not be feasible.

Table 11: The Epps effect

		All	$\lambda = 23400$	$\lambda = 7800$	$\lambda = 1560$	$\lambda = 390$	$\lambda = 78$	$\lambda = 39$
$dt = 1$	$\rho_{S,F}$	1.0000	0.9955	0.2001	0.0344	0.0080	0.0019	$8.5490e^{-4}$
	$RT - \rho_{S,F}$	1.0000	1.0000	0.7705	0.6845	0.6696	0.6569	0.6648
	Obs. distance	0	2	71	200	380	904	1.190
$dt = \frac{1}{10}$	$\rho_{S,F}$	1.0000	0.0525	0.0171	0.0032	0.0010	2.9154^{-5}	0
	$RT - \rho_{S,F}$	1.0000	0.6958	0.6757	0.6675	0.6643	0.6680	0.6570
	Obs. distance	0	400	826	2.446	4.179	6.688	15.327
$dt = \frac{1}{100}$	$\rho_{S,F}$	1.0000	0.0052	0.0017	$2.8290e^{-4}$	$5.9926e^{-5}$	0	0
	$RT - \rho_{S,F}$	1.0000	0.6697	0.6674	0.6654	0.6580	0.6649	0.6668
	Obs. distance	0	5.066	8.384	19.223	41.115	100.231	135.768
$dt = \frac{1}{1000}$	$\rho_{S,F}$	1.0000	5.0875^{-4}	$1.1484e^{-4}$	$2.5329e^{-6}$	$-1.6759e^{-7}$	0	0
	$RT - \rho_{S,F}$	1.0000	0.6669	0.6662	0.6677	0.6552	0.6528	0.6543
	Obs. distance	0	49,441	69,384	231,372	247,790	847,460	1,075,818

The estimated correlation between the two noise free price processes is reported for different level of trading intensity and for processes generated at increasingly high frequencies together with the average realization's distance (before synchronization).

Figure 6: Realized Covariance signature plot for varying trading intensities
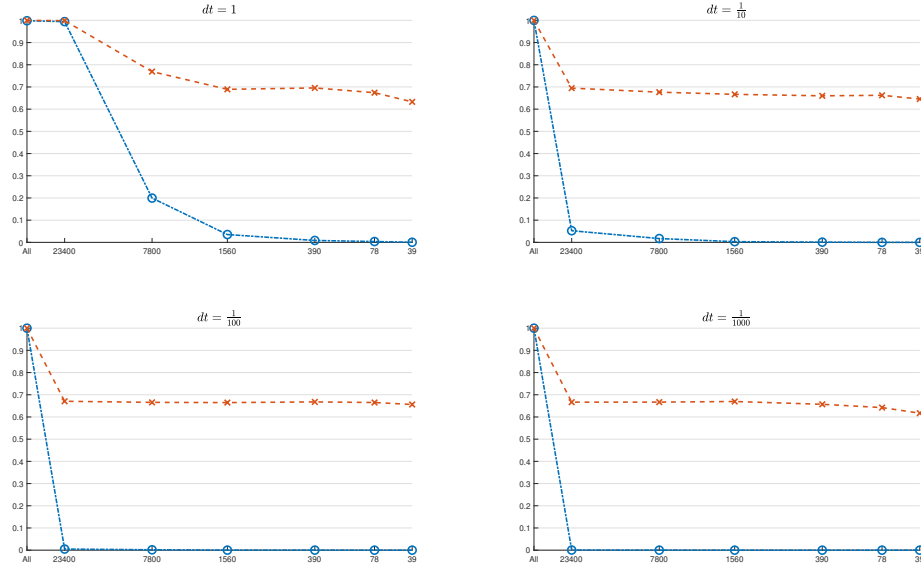
Signature plot for previous tick $RC_{S,F}$ (○--○--○) and refresh time $RC_{S,F}$ (×--×--×) with varying trading intensities (λ).

4 Empirical

The objective of this section is to empirically test the accuracy of our noise reduction methodology. We will consider a pair of assets; a futures and its underlying stock with the intention to measure the daily quadratic variation of the underlying as accurately as

possible. We will adopt the noise reduction methodology introduced in Section 2.3 to show the efficiency gain of the realized measures employed on the noise reduced series Z. The empirical performance of the procedure is then assessed through visual inspection of the stabilized signature plot and the proximity of the simple realized variance computed at very high frequencies to different noise robust estimators.

4.1   Dataset

Our dataset consists of transaction data recorded over 13 years; from January 2001 to December 2013 for a total of $3,226$ trading days. The assets examined are the Spyder (SPY), an ETF tracking the S&P 500 index, and the E-mini S&P 500 futures (ES). E-mini futures contracts represent a fraction of a standard futures contract with their price reflecting this proportion. The small size and low price together with the low margin rates of these contracts, make them extremely appealing to investors and hedgers. As a consequence, the dataset is exceptionally liquid; a particularly desirable characteristic given the purpose of our study and the discussion of Section 2.5.

However, futures contracts present some peculiar features that require a further discussion on how to handle the data. Unlike normal stocks, futures contracts are active and can be traded only for a specific amount of time as they have a predefined maturity (or expiration) date after which the instrument cannot be traded anymore. Further, several futures written on the same underlying but with different maturities are simultaneously traded. Generally, the futures closer to maturity (F_{T_1}) is more actively traded than futures with longer maturities (say, F_{T_2} and F_{T_3}). This is true until we approach the maturity date (T_1); at that point a swift drop in the number of transactions for contract (F_{T_1}) and an opposite, sharp, increase in the liquidity of the following contract in order of maturity (F_{T_2}) is simultaneously recorded. This is known as rollover period as investors tend to move their positions from one contract to the other. We define an ensamble of rules (referred to as "rollover scheme") in order to create a continuous time series of prices. The main idea is to consider, at each point in time, the contract with highest liquidity (as previously said, this is generally the futures with closest maturity, T_1). Hence, our time series will always consider the contract closest to expiration up until two weeks to the maturity date. The two weeks (10 working days) preceding T_1 are our "rollover window". Within this window we measure the daily liquidity of the two contracts with closest maturity $(F_{T_1}$ and $F_{T_2})$ and, as soon as the liquidity of the contract expiring at T_2 exceeds that of the contract with maturity T_1, we rollover to the former.

Both time series have further been filtered for outliers. After removing all the transactions

recorded outside of normal market hours (09:30 - 16:00, New York local time) and obviously incorrect transactions[11] we implemented a statistical filter to verify the likelihood that each observation has to be an outlier. The filtering is closely based on the methodology proposed by Brownlees and Gallo (2006) and we refer readers to the aforementioned article for details.

Table 12: Summary statistics

	Tick		Transaction	
	ES	SPY	ES	SPY
Avg. $\widehat{\sigma}_{\varepsilon}^2$ (e^{-5})	11.5674	8.7108	13.1047	7.4317
Avg. $\hat{\gamma}$	1.5302	1.1176	1.4508	1.1727
Avg. $\hat{\lambda}$ (e^{-3})	0.1119	0.0848	0.1179	0.0667
Max. # of obs.	145,021	535,510	404,392	2,245,367
Min. # of obs.	1,619	1,034	780	1,299
Avg. # of obs.	13,355	46,922	35,934	259,073
Time between trades (sec.)	1.7522	0.4987	0.6512	0.0903

The reported noise variances have been computed using percentage returns and without further subsampling (i.e. using unequally spaced observations).

We will conduct our analysis using two different sampling schemes; namely *tick* time and *transaction* time. When working in *transaction* time all the available observations for a given asset on a given trading day are considered while in the *tick* sampling scheme stale prices are disregarded and only if a price change is registered the observation is retained. Table 12 reports basic summary statistics for the assets considered in both sampling schemes. Following Hansen and Lunde (2006) we report the average noise-to-signal ($\hat{\lambda}$) constructed as $\hat{\lambda} = \bar{\sigma}_{\varepsilon}^2 / \bar{IV}$ with $\bar{\sigma}_{\varepsilon}^2 = n^{-1} \sum_{t=1}^{n} \sigma_{\varepsilon,t}^2$ and $\bar{IV} = n^{-1} \sum_{t=1}^{n} IV_t$ being the sample averages of the noise variance and the estimated integrated variance. Both the noise-to-signal and the noise ratio (γ) hint to a low noise intensity in both series with the E-mini futures being slightly noisier than the SPY. A relevant difference in liquidity can easily be evinced by looking at the average and the max number of observations (which, in turn, reflects into different time between trades). While significant, this difference is not problematic as the overall liquidity of each asset is still extremely high with the less liquid case enjoying, on average, more than one observation every two seconds.

4.2   Empirical Results

Following the format implemented in our simulation, we investigate the results for different realized measures and we focus particularly on the signature plot of the simple RV to

---

[11]For example observations with a recorded bid, ask or price equal to zero. A detailed step-by-step procedure can be found in Barndorff-Nielsen et al. (2009).

highlight the stabilization performance obtained via cross-sectional noise reduction.

Table 13 presents a summary of the key noise-reduction parameters employed for some of the sampling frequencies considered. We can see that for each different frequency, $\hat{\alpha}$ adapts to the different correlation and noise variance estimates. However, it should be noted that, as we decrease the sampling frequency, the noise estimates become less reliable as the microstructure bias tends to disappear quite fast.[12] For this reason we construct the noise reduced series $Z(\hat{\alpha}_{hf})$ obtained estimating the characteristics of the noise, and thus $\hat{\alpha}$ at the highest possible frequency and keeping the obtained value fixed also when sampling at lower frequencies. The idea is that at the highest frequency we expect, in principle, to obtain a more reliable estimate of the noise component which, in turn, should produce values of $\hat{\alpha}$ closer to the optimal $\alpha^{\dagger}$ and positively affect the construction of the noise reduced series.

Table 13: Noise Reduction key figures

		1	5	10	30	60	300	600	1800	Ref.Time
	$\hat{\alpha}$	0.6155	0.6510	0.6937	0.7662	0.7777	0.7692	0.8696	1.2936	0.6926
	$\hat{\rho}_{\varepsilon^S,\varepsilon^F}$	0.1416	0.4387	0.5765	0.7530	0.8168	0.9442	0.9590	0.9488	0.2008
	$\widehat{RC}_{S,F}$	0.3799	0.7200	0.7996	0.9020	0.9497	1.0297	1.0177	0.9591	0.4440
Tick	$\hat{\sigma}^2_{\varepsilon^S}\ (e^{-3})$	0.0601	0.1832	0.2990	0.7610	1.4924	6.9495	13.4631	38.6353	0.2259
	$\hat{\sigma}^2_{\varepsilon^F}\ (e^{-3})$	0.0547	0.1678	0.2937	0.7748	1.4889	7.0319	13.7478	39.1300	0.3043
	$\hat{\sigma}^2_{\varepsilon^Z}\ (e^{-3})$	0.0261	0.1140	0.2200	0.6529	1.3246	6.5859	12.5954	33.3346	0.2719
	$\hat{\alpha}$	0.6155	0.6511	0.6942	0.7688	0.7853	0.8441	1.0205	1.6808	0.6265
	$\hat{\rho}_{\varepsilon^S,\varepsilon^F}$	0.1053	0.3782	0.5183	0.7230	0.8098	0.9430	0.9565	0.9427	0.1271
	$\widehat{RC}_{S,F}$	0.3800	0.7201	0.8000	0.9028	0.9509	1.0341	1.0257	0.9897	0.3309
Transaction	$\hat{\sigma}^2_{\varepsilon^S}\ (e^{-3})$	0.0601	0.1832	0.2990	0.7610	1.4924	6.9495	13.4631	38.6353	0.1628
	$\hat{\sigma}^2_{\varepsilon^F}\ (e^{-3})$	0.0547	0.1679	0.2938	0.7756	1.4911	7.0724	13.8920	40.6640	0.1721
	$\hat{\sigma}^2_{\varepsilon^Z}\ (e^{-3})$	0.0261	0.1140	0.2200	0.6530	1.3249	6.5757	12.5460	33.4583	0.2141

All the reported values represent averages over the considered period. The estimated realized covariance has been computed on the two series synchronized with the previous tick methodology.

Table 14 presents the average estimates of each realized measure and each series. Here, $Z(\hat{\alpha})_{\text{RT}}$ identifies the noise reduced series constructed using refresh time synchronization while the remaining series have been sampled at one second frequency using previous tick interpolation. For each series, the last column reports the estimated noise variance at one second frequency as ratio over the noise variance of the SPY $(\sigma^2_{\varepsilon}(S))$.[13] As previously mentioned, $Z(\hat{\alpha}_{hf})$, refers to the noise reduced series obtained estimating the characteristics of the noise process at the highest possible frequency (i.e., the noise variance of S and F is computed before synchronizing the two series).

Unfortunately, as the integrated variance is a latent variable, its true value remains unobserv-

---

[12] Indeed, for low sampling frequencies we obtain (scaled) consistent estimates of the realization of the true process rather than the microstructure noise characteristics.

[13] Since we defined IV of the underlying stock to be our object of interest.

Table 14: QV Estimates

		PA	RK	TSRV	MSRV	RV^*	RV5	RV	σ_ε^2
Tick	$Z(\hat{\alpha})_{\text{RT}}$	1.0814	1.0818	0.5118	1.0640	1.0725	1.0561	0.9755	0.5738
	$Z(\hat{\alpha})$	1.0855	1.0908	0.5733	1.0280	1.0684	1.0274	1.2211	0.4343
	$Z(\hat{\alpha}_{hf})$	1.0710	1.0771	0.5577	1.0031	1.0560	1.0156	1.1867	0.4073
	S	1.1106	1.1143	0.9731	1.2334	1.1168	1.0841	2.8119	1
	F	1.1098	1.1118	0.8579	1.2300	1.0967	1.0970	2.5605	0.9106
Transaction	$Z(\hat{\alpha})_{\text{RT}}$	1.0836	1.0852	0.5244	1.0553	1.0734	1.0592	1.0408	0.5756
	$Z(\hat{\alpha})$	1.0856	1.0909	0.5733	1.0280	1.0677	1.0258	1.2212	0.4342
	$Z(\hat{\alpha}_{hf})$	1.0713	1.0774	0.5579	1.0033	1.0555	1.0580	1.1871	0.4074
	S	1.1106	1.1143	0.9731	1.2334	1.1168	1.0841	2.8119	1
	F	1.1098	1.1118	0.8584	1.2300	1.0957	1.1033	2.5611	0.9580

All the reported values represent averages over the considered period. The two price series S and F have been equally spaced at one second interval using previous tick interpolation. $Z(\hat{\alpha})_{\text{RT}}$ reports the values for the noise reduced series constructed using refresh time synchronization. The Noise variance is expressed as ratio over the ETF (subsampled at 1 second frequency) noise variance. The realized measures are constructed using percentage returns.

able even ex-post; thus making a performance comparison a non-trivial task. Nonetheless, we can get an idea of the good performance of the noise reduction methodology by jointly looking at the proximity of the estimates to realized measures known to be asymptotically robust and inspecting the stable behaviour of RV constructed at ever increasing sampling frequencies. Figure 7 displays the signature plots obtained with tick and transaction time. Very little difference can be seen between the two plots and the remarkable stabilization obtained appear to be indisputable; hinting to a very good performance of the noise reduction methodology. As expected, the stabilization is further improved by $Z(\hat{\alpha}_{hf})$; estimating the noise characteristics and "portfolio" weights ($\hat{\alpha}_{hf}$) at the highest frequency grants an increase in performance. Undoubtedly, however, the noise is not completely eliminated and sampling at the highest frequency still provides biased estimates. Nonetheless, it appears that very high sampling frequencies as ten or even five seconds can be considered virtually noise free allowing practitioner to obtain robust estimates of IV using the trivial RV estimator while retaining an incredible amount of data compared to the commonly used five minutes frequency.

Figure 7: Stabilization of the signature plot
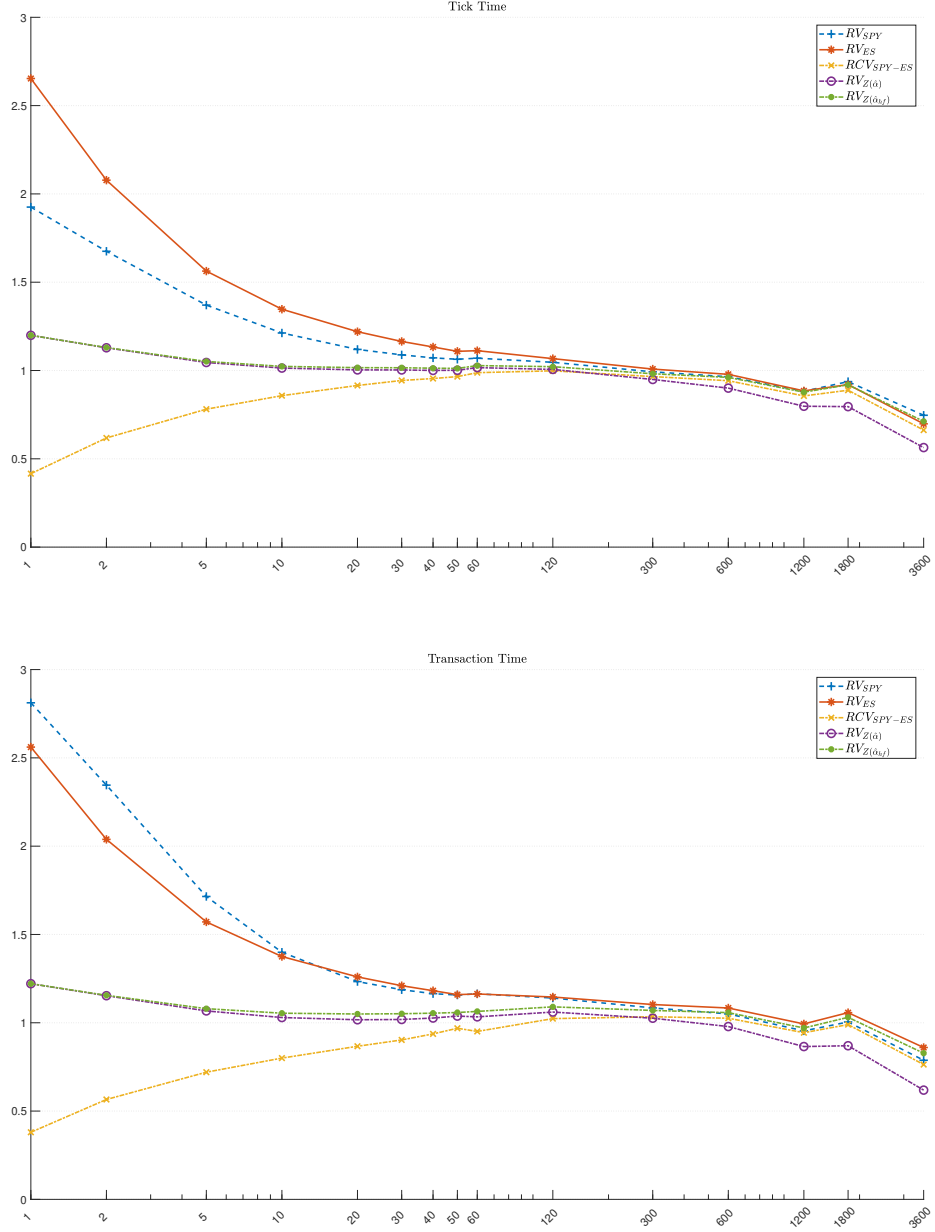
Signature plot for $RV_{Z(\hat{\alpha})}$ ($\bigcirc$--$\bigcirc$--$\bigcirc$), $RV_{Z(\hat{\alpha}_{hf})}$ ($\bullet$--$\bullet$--$\bullet$), RC ($\times$--$\times$--$\times$), RV_S ($+$--$+$--$+$), RV_F ($*$--$*$--$*$) with varying values of $\rho_{\varepsilon^{(S)},\varepsilon^{(F)}}$.

5   Conclusion

In the present paper, a new procedure to reduce the microstructure noise of a series while preserving its quadratic variation is proposed. Being subject to a lower microstructure noise, implementation of any realized measure on the resulting series (Z) grants more precise

estimates of the quadratic variation. This allows to obtain consistent QV estimates using the trivial realized variance estimator at frequencies as high as five seconds. The proposed methodology is based on cross-sectional combination of different price series with the same variance (as a futures and its underlying or an asset traded on different markets). Therefore, due to the nature of the technique, two main factors influence the quality of the noise reduction: the correlation between the noise processes affecting the different time series and the asynchronicity of the assets' realizations. Fortunately, as shown in Section 3, provided that the price series are not excessively illiquid the procedure will only fail to produce any significant noise reduction but will not negatively affect the estimators performance. We investigate the behavior of the noise reduction methodology on a dataset spanning over 13 years and document exceptionally good results, evidenced both by the proximity of the realized measures computed on the noise-reduced series Z to noise-robust estimators (as the realized kernel or the pre-averaged realized variance) and by the remarkable stabilization displayed by the volatility signature plot of the noise reduced series. This novel approach opens to extensions in several directions; for instance, more refined synchronization procedures might increase the robustness to asynchronicity of the proposed methodology, making it less susceptible to problem caused by illiquid series.

References

Aït-Sahalia, Y. and Yu, J. High frequency market microstructure noise estimates and liquidity measures. *The Annals of Applied Statistics*, 3(1):422 – 457, 2009.

Andersen, T., Tim, B., Diebold, F., and Labys, P. Great realizations. *Risk*, 2000.

Bandi, F. M. and Russell, J. R. Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies*, 75(2):339–369, 2008.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536, 2008.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, 12(3):C1–C32, 2009.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics*, 162(2):149 – 169, 2011. ISSN 0304-4076. doi: http://dx.doi.org/10.1016/j.jeconom.2010.07.009. URL http://www.sciencedirect.com/science/article/pii/S0304407611000029.

Björk, T. *Arbitrage theory in continuous time*. Oxford university press, 2009.

Brownlees, C. T. and Gallo, G. M. Financial econometric analysis at ultra-high frequency: Data handling concerns. *Computational Statistics & Data Analysis*, 51:2232–2245, 2006. doi: 10.1016/j.csda.2006.09.030.

Christensen, K., Kinnebrock, S., and Podolskij, M. Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159(1):116 – 133, 2010a. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2010.05.001. URL http://www.sciencedirect.com/science/article/pii/S0304407610001260.

Christensen, K., Oomen, R., and Podolskij, M. Realised quantile-based estimation of the integrated variance. *Journal of Econometrics*, 159(1):74 – 98, 2010b. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2010.04.008. URL http://www.sciencedirect.com/science/article/pii/S0304407610001144.

Christensen, K., Oomen, R. C., and Podolskij, M. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576 – 599, 2014. ISSN 0304-405X. doi: http://dx.doi.org/10.1016/j.jfineco.2014.07.007. URL http://www.sciencedirect.com/science/article/pii/S0304405X14001548.

Cox, J. C., Ingersoll Jr, J. E., and Ross, S. A. A theory of the term structure of interest rates. *Econometrica*, pages 385–407, 1985.

Delbaen, F. and Schachermayer, W. A general version of the fundamental theorem of asset pricing. *Mathematische annalen*, 300(1):463–520, 1994.

Epps, T. W. Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74(366):291–298, 1979. ISSN 01621459. URL http://www.jstor.org/stable/2286325.

Hansen, P. R. and Lunde, A. Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161, 2006.

Hayashi, T. and Yoshida, N. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379, 04 2005. doi: 10.3150/bj/1116340299. URL https://doi.org/10.3150/bj/1116340299.

Jacod, J. and Protter, P. E. *Discretization of processes*, volume 67. Springer Science & Business Media, 2011.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications*, 119(7):2249–2276, 2009.

Jacod, J., Li, Y., and Zheng, X. Statistical properties of microstructure noise. 2016.

Lunde, A., Mirone, G., and Zeng, Y. Incorporating overnight futures data in daytime stock volatility forecasting. CREATES Working Paper, Aarhus University, 2017.

Mirone, G. Inference from the futures: Ranking the noise cancelling accuracy of realized measures. CREATES Working Paper, Aarhus University, 2017.

Munk, C. *Fixed Income Modelling*. Oxford University Press, USA, 2011. ISBN 0199575088,9780199575084.

Ramaswamy, K. and Sundaresan, S. M. The valuation of options on futures contracts. *The Journal of Finance*, 40(5):1319–1340, 1985.

Renò, R. A closer look at the epps effect. *International Journal of Theoretical and Applied Finance*, 06(01):87–102, 2003. doi: 10.1142/S0219024903001839. URL http://www.worldscientific.com/doi/abs/10.1142/S0219024903001839.

Rossi, E. and Santucci de Magistris, P. A no-arbitrage fractional cointegration model for futures and spot daily ranges. *Journal of Futures Markets*, 33(1):77–102, 2013. ISSN 1096-9934. doi: 10.1002/fut.20556. URL http://dx.doi.org/10.1002/fut.20556.

Zhang, L. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12(6):1019–1043, 2006.

Zhang, L. Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33 – 47, 2011. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2010.03.012. URL http://www.sciencedirect.com/science/article/pii/S0304407610000540. Realized Volatility.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.

Research Papers
2018

CREATES
Center for Research in Econometric
Analysis of Time Series

2018-01:	Emilio Zanetti Chini: Forecaster's utility and forecasts coherence
2018-02:	Torben G. Andersen, Nicola Fusari and Viktor Todorov: The Pricing of Tail Risk and the Equity Premium: Evidence from International Option Markets
2018-03:	Torben G. Andersen, Nicola Fusari, Viktor Todorov and Rasmus T. Varneskov: Unified Inference for Nonlinear Factor Models from Panels with Fixed and Large Time Span
2018-04:	Torben G. Andersen, Nicola Fusari, Viktor Todorov and Rasmus T. Varneskov: Option Panels in Pure-Jump Settings
2018-05:	Torben G. Andersen, Martin Thyrsgaard and Viktor Todorov: Time-Varying Periodicity in Intraday Volatility
2018-06:	Niels Haldrup and Carsten P. T. Rosenskjold: A Parametric Factor Model of the Term Structure of Mortality
2018-07:	Torben G. Andersen, Nicola Fusari and Viktor Todorov: The Risk Premia Embedded in Index Options
2018-08:	Torben G. Andersen, Nicola Fusari and Viktor Todorov: Short-Term Market Risks Implied by Weekly Options
2018-09:	Torben G. Andersen and Rasmus T. Varneskov: Consistent Inference for Predictive Regressions in Persistent VAR Economies
2018-10:	Isabel Casas, Xiuping Mao and Helena Veiga: Reexamining financial and economic predictability with new estimators of realized variance and variance risk premium
2018-11:	Yunus Emre Ergemen and Carlos Velasco: Persistence Heterogeneity Testing in Panels with Interactive Fixed Effects
2018-12:	Hossein Asgharian, Charlotte Christiansen and Ai Jun Hou: Economic Policy Uncertainty and Long-Run Stock Market Volatility and Correlation
2018-13:	Emilio Zanetti Chini: Forecasting dynamically asymmetric fluctuations of the U.S. business cycle
2018-14:	Cristina Amado, Annastiina Silvennoinen and Timo Teräsvirta: Models with Multiplicative Decomposition of Conditional Variances and Correlations
2018-15:	Changli He, Jian Kang, Timo Teräsvirta and Shuhua Zhang: The Shifting Seasonal Mean Autoregressive Model and Seasonality in the Central England Monthly Temperature Series, 1772-2016
2018-16:	Ulrich Hounyo and Rasmus T. Varneskov: Inference for Local Distributions at High Sampling Frequencies: A Bootstrap Approach
2018-17:	Søren Johansen and Morten Ørregaard Nielsen: Nonstationary cointegration in the fractionally cointegrated VAR model
2018-18:	Giorgio Mirone: Cross-sectional noise reduction and more efficient estimation of Integrated Variance