

Realized Drift*

Sébastien Laurent[‡] Roberto Renò[§] Shuping Shi[¶]

April 15, 2022

Abstract

Drift and volatility are two mainsprings of asset price dynamics. While volatilities have been studied extensively in the literature, drifts are commonly believed to be impossible to estimate and largely ignored in the literature. This paper shows how to detect drift using *realized autocovariance* implemented on high-frequency data. We use a theoretical treatment in which the classical model for the efficient price, an Itô semimartingale possibly contaminated by microstructure noise, is enriched with drift and volatility explosions. Our theory advocates a novel decomposition for realized variance into a drift and a volatility component, which leads to significant improvements in volatility forecasting.

JEL classification: C58, C12, C14;

Keywords: Drift, High-frequency Data, Serial Covariance, Volatility Forecasting

*We are grateful to Francesco Benvenuti, Giuseppe Buccheri, Aleksey Kolokolov, Jia Li, Peter C.B. Phillips, and Jun Yu for constructive comments. The authors would also like to thank the seminar participants at the Singapore Management University, Nottingham University, CREST Paris, and Melbourne University and the participants at the 2021 Asian Meeting of Econometric Society, 2021 International Association of Applied Econometrics, and 2021 Society of Financial Econometrics for useful comments and discussions. Sébastien acknowledges the research support of the French National Research Agency Grants ANR-17-EURE-0020 and ANR-21-CE26-0007-01. Roberto acknowledges support by the PRIN project HiDEA, MIUR Grant 2017RSMPZZ. Shuping acknowledges research support from the Australian Research Council under project No. DE190100840.

[‡]Aix-Marseille University, CNRS & EHESS, Aix-Marseille Graduate School of Management – IAE, e-mail: SEBASTIEN.LAURENT@UNIV-AMU.FR

[§]University of Verona, Dipartimento di Scienze Economiche, e-mail: ROBERTO.RENO@UNIVR.IT

[¶]Macquarie University, Department of Economics, e-mail: SHUPING.SHI@MQ.EDU.AU

1 Introduction

The availability of intraday high-frequency data has fostered the rapid development of financial econometrics over the past two decades. In the high-frequency financial econometrics literature, asset prices are typically modelled as an Itô semi-martingale process with two main components: drift and diffusion. While the diffusion component has played a central role, the drift term is largely ignored in the literature, as under standard model assumptions the drift term is asymptotically dominated by the diffusion component.¹

There is, however, substantial empirical evidence documenting the presence of non-negligible drifts in asset prices in both low frequency (daily, weekly, or monthly) and high-frequency settings. The non-negligible drift can take various forms in discrete and continuous time models. A special case of the Itô semi-martingale process is the well-known Ornstein-Uhlenbeck (OU) process, whose exact discrete-time solution is an AR(1) process (Arnold, 1974). The non-negligible drift can appear as either a large intercept or the autoregressive coefficient deviating from unity in the AR(1) model. In the low-frequency setting, Phillips and Shi (2019) find the presence of a large random intercept in the AR(1) model for log prices and bond yields during crisis periods. There is also a large literature showing temporary deviations from the random walk of log prices in the low-frequency data setting² and more recently in the high-frequency setting by Laurent and Shi (2022). Finally, Christensen et al. (2022) find episodes of drift bursting across equities, fixed income, currencies, and commodities.

Drift is an elusive quantity, a fact which is well known since, at least, Merton (1980). Existing methods for estimating the drift of asset returns require a long span of data (see, e.g., Bandi and Phillips, 2003), and even in this case, the estimate is noisy. Estimating drift locally, using high-frequency data, is just im-

¹See, for example, Andersen and Bollerslev (1998); Barndorff-Nielsen and Shephard (2002b); Lee and Mykland (2008); Andersen et al. (2012). There are a few exceptions. For example, to reveal the impact of drifts on their respective estimators, Bollerslev et al. (2020) investigate the second-order asymptotics and Laurent and Shi (2020) study their finite sample theory. Barndorff-Nielsen et al. (2010) show that drift induces biases for realized semivariance.

²See, for example, Fama and French (1988); Bekaert and Hodrick (1992); Bessembinder and Chan (1992); Campbell and Ammer (1993); Campbell and Hamao (1992); Phillips et al. (2011); Phillips and Yu (2011); Phillips et al. (2015); Shi and Song (2016).

possible (see, e.g., [Bandi, 2002](#); [Kristensen, 2010](#)) since the variance of the local drift estimator diverges to infinity. In portfolio management, people simply give up at estimating a useless drift from historical data ([Michaud, 1989](#)) and either resort to sophisticated techniques to grasp some signal in the conditional mean or even prefer to use subjective views à la Black-Litterman.

This paper shows that drift can be quantified using high-frequency data, even when the sampling frequency shrinks indefinitely in a fixed time span, and its significant presence is massively revealed by financial price data. As a relevant application of our findings, we show that drift identification yields substantial improvements in volatility forecasting.

The point that drift is not invisible to high-frequency in-fill sampling is more subtle and less intuitive and represents the central contribution of this paper. Drift, of course, appears as a small sample correction in all quantities estimated using financial return. This correction vanishes asymptotically, unless the drift gets very large, an occurrence which happens undeniably often in financial data ([Laurent and Shi, 2020, 2022](#)) and that can be formally tested ([Christensen et al., 2022](#)). In this paper, we show the *realized autocovariance* is sufficient to isolate the drift component. Borrowing from [Christensen et al. \(2022\)](#), we build a model specification in continuous time (encompassing traditional financial models and enriching them with drift and volatility explosions) under which the drift is proportional to the first-order term in the asymptotically vanishing serial covariance. Our theoretical contribution is thus to provide the asymptotic distribution of serial covariance under all possible settings of drift/volatility explosion, assessing under which circumstances the drift is the dominating asymptotic term.

Relating our work to existing works on realized volatility estimation, we show indeed that the serial covariance is just the difference between realized variance and the Rice estimator ([Rice, 1984](#)), which is specifically designed to remove drift (see, e.g., [Andersen et al., 2021](#), who use a similar idea in concurrent work). We show that both realized variance and the Rice estimator are biased in the presence of an exploding drift, but the bias of the Rice estimator is smaller. So, it makes perfect sense that their difference measures drift, as we indeed show.

Importantly, our theoretical results imply that the leading term in the serial co-

variance as the sampling frequency increases should be positive when the drift is large, while it should be centered around zero otherwise. A realistic simulation study which includes standard features of the price process (stochastic volatility, intraday volatility effects, and market microstructure noise) reveals that, when computing serial covariance at a frequency large enough, a constant drift or an OU process (as in [Laurent and Shi, 2020](#)) inducing a daily price change of 5% can easily be revealed. Interestingly, a much smaller price change is needed to reveal drift when the drift explodes, as in [Christensen et al. \(2022\)](#).

As mentioned, our paper is close to [Andersen et al. \(2021\)](#), which studies the Rice estimator as a drift-robust replacement of realized variance under drift explosion. A crucial difference in our paper is that we do not truncate returns using the [Mancini \(2009\)](#) technique. Truncation, which was originally designed to remove jumps, annihilates all returns which vanish at a slower rate than a threshold (with a slightly higher order than the Brownian component). Thus, its effect is to eliminate both returns associated with explosive drift and returns associated with explosive volatility, since both vanish at a slower rate than the threshold. Thus, if the interest is in estimating drift, we should not truncate. Without truncation, the bias term due to the drift has the same vanishing rate in realized volatility, the Rice estimator, and the autocovariance. However, drift is the prevailing term in the autocovariance only. Moreover, in the autocovariance, jumps are anyway removed asymptotically by a mechanism identical to that of bipower variation ([Barndorff-Nielsen and Shephard, 2006](#)), so there is no need of truncating to get rid of them. Another important difference with their paper is that we add an explicit volatility explosion term. This theoretical addition is compelling from an economic standpoint, since an exploding volatility is a necessary ingredient to soften the violation of no-arbitrage principle under drift explosion (see the discussion in [Christensen et al., 2022](#)). The joint presence of a drift and a volatility explosion allows us to study accurately the interplay between the two explosion rates in determining the central limit theorems for realized variance, Rice estimator, and the autocovariance.

Our asymptotic theory has relevant consequences for volatility forecasting. We decompose realized variance into a volatility and a drift component. If the two

components have different persistence, this implies that realized autocovariance (estimating drift) should forecast future realized variances, even in the absence of interaction effects between the two quantities. Moreover, following the reasoning of [Bollerslev et al. \(2016\)](#) for the quarticity expanded heterogeneous autoregression model (HARQ), a new interaction term in which realized autocovariance multiplies a realized quarticity estimator is predicted to have forecasting power on top of the interaction term of the HARQ model. These two novel explanatory variables for future realized volatility are unexplored in the vast volatility forecasting literature. In our empirical application, we prove that the two predicted effects are significant and that they lead to significant improvements in the quality of the volatility forecast, corroborating our theoretical analysis. Moreover, we show that serial covariance is often significantly positive and never significantly negative, in line with our theory. Our empirical analysis thus reveals the massive presence of price drifts in the data, especially in periods with visible trends, like the dotcom bubble in the late 1990s and early 2000s and the recent COVID-19 pandemic.

The rest of the paper is organized as follows. Section 2 introduces our model specification, which allows for both drift and volatility bursting and incorporates the standard Itô semi-martingale process as a special case. Section 3 presents the three estimators (realized volatility, Rice estimator, and realized autocovariance) and provides their asymptotic properties under the general model specification. A preliminary analysis of a noise-robust version of the realized autocovariance is also conducted here. Section 4 discusses the implications of our theoretical results for volatility forecasting. The simulation study is detailed in Section 5. Section 6 presents the realized autocovariance estimates and the in-sample estimation and out-of-sample forecasting results using the new volatility forecasting models (along with the most relevant benchmark models) for the Nasdaq Composite index. Main proofs are collected in Appendix [A](#). Additional mathematical results are presented in the Online Appendices [B](#) and [C](#).

2 The model

We start with preliminary considerations. Denoting by X_t the logarithmic price, a standard way to model (discrete) returns at a given frequency Δ_n is to write

$$X_{t+\Delta_n} - X_t = \bar{\mu}_t \Delta_n + \bar{\sigma}_t \sqrt{\Delta_n} \epsilon_t, \quad (1)$$

where ϵ_t is iid noise with mean zero and unit variance. We wish to measure the extent of $\bar{\mu}_t$ (the drift) over a window relative to $\bar{\sigma}_t$ (the volatility). In the fixed Δ_n case, we cannot identify both terms (the drift $\bar{\mu}_t$ and the volatility $\bar{\sigma}_t$) in the above equation unless we dispose of an infinite number of returns over a window of an infinite size. As it is well known, we can instead identify the volatility *only* using *infill asymptotics*, that is a setting in which

- i) $\Delta_n \rightarrow 0$;
- ii) Eq. (1) becomes an It \bar{o} semimartingale (possibly, with jumps);
- iii) $\bar{\mu}_t$ is a locally bounded stochastic process.

In this setting, we can estimate spot volatility ([Jacod and Protter, 2011](#)) and the integrated variance $\int_0^T \bar{\sigma}_s^2 ds$ consistently, while the drift becomes invisible since $\Delta_n \ll \sqrt{\Delta_n}$ in the limit.

The above discussion is classical, but is useful in pointing at a direction in which drift becomes visible even when $\Delta_n \rightarrow 0$: allowing for unbounded drift in the data generating process, which is what we do in this paper. With this additional flexibility with respect to the traditional model, the asymptotic ranking between drift and volatility inverts and we are able to identify the drift part. In particular, the point of this paper is to show that when the drift is unbounded, realized volatility remains a consistent estimator of the volatility, and the integrated square drift is proportional to realized autocovariance. Using the two quantities thus allows to estimate drift and volatility separately in small samples.

Following the above intuition, the model we study in this paper deviates from the classical It \bar{o} semi-martingale by allowing both volatility and drift to explode locally, still preserving the continuity of price paths. Since our interest is in understanding the contribution of drift to quadratic variation and autocovariation, we

exclude jumps in the primitive assumptions. This is completely harmless. Jumps can be inserted in our picture at any moment without affecting main intuitions, since the estimators we are going to consider preserve their consistency properties even in their presence. Jumps would just change the estimation target and the asymptotic variance of the considered estimators in a straightforward manner.

Let $X = (X_t)_{0 \leq t \leq 1}$ denote the log-price of a traded security. We assume the following.

Assumption 2.1. *X is defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathcal{P})$ and assumed to be an Itô semimartingale described by the dynamics:*

$$X_t = \int_0^t \mu_s \left(1 - \frac{s}{\tau}\right)^{-\alpha} ds + \int_0^t \sigma_s \left(1 - \frac{s}{\tau}\right)^{-\beta} dW_s, \quad (2)$$

where $0 \leq \alpha < 1$, $0 \leq \beta < \frac{1}{2}$, $\tau > 0$, X_0 is \mathcal{F}_0 -measurable, $\mu = (\mu_t)_{0 \leq t \leq 1}$ is a locally bounded, predictable drift and $|\mu_s|$ is strictly positive (almost surely), $\sigma = (\sigma_t)_{0 \leq t \leq 1}$ is an adapted, càdlàg and strictly positive (almost surely) volatility, $W = (W_t)_{0 \leq t \leq 1}$ is a standard Brownian motion. The coefficients μ_s and σ_s are such that, for a suitable $\Gamma > 2\alpha + \min(1/2, 1 - 2\beta) - 2$ and $C > 0$, and for $|u - s|$ small enough, we have

$$E_{u \wedge s} [|\mu_u - \mu_s|^2 + |\sigma_u - \sigma_s|^2] \leq C|u - s|^\Gamma, \quad (3)$$

where $E_t[\cdot] = E[\cdot | \mathcal{F}_t]$.

Assumption 2.1 is mild. We just assume that the coefficients μ_s and σ_s are stochastically continuous and locally bounded, an assumption which encompasses virtually all continuous-time specifications in financial economics. The condition $\Gamma > 2\alpha + \min(1/2, 1 - 2\beta) - 2$ is imposed to avoid that the stochasticity of the bias term dominates the variance terms, and is also very mild (since, in the worst possible case, $\alpha = 1 - \epsilon$ and $\beta < 1/4$ and in this case we have $\Gamma > 1/2 - 2\epsilon$).

This model allows for drift explosion (when $\alpha > 0$) and for volatility explosion (when $\beta > 0$), capturing both flash crashes and longer term market deviations. Without loss of generality, both the drift and the volatility components burst at $t = \tau$ and from now on we assume $\tau = 1$.

Remark 2.1. *The rationale of the model is the following. In the classical case*

without explosions, when $\alpha = \beta = 0$, drift and volatility coefficients of the Itô semi-martingale are bounded stochastic processes. In this case, the drift is just invisible to high-frequency data. The reason is simply that, when $\Delta_n \leq t \leq 1$ and for a small Δ_n ,

$$\alpha = 0 \Rightarrow \int_{t-\Delta_n}^t \mu_s ds = O_p(\Delta_n) \quad (4)$$

and

$$\beta = 0 \Rightarrow \int_{t-\Delta_n}^t \sigma_s dW_s = O_p(\Delta_n^{1/2}), \quad (5)$$

so that the volatility always dominates the drift term in high-frequency returns. The only way to make the drift prevailing over volatility is to have it exploding, that is $\alpha > 0$. When $\alpha > 0$ and $\beta > 0$, the order of magnitude of the drift and diffusion components over the interval $[t - \Delta_n, t]$ are, respectively,

$$\int_{t-\Delta_n}^t \mu_s (1-s)^{-\alpha} ds = O_p(\Delta_n^{1-\alpha}) \quad \text{and} \quad \int_{t-\Delta_n}^t \sigma_s (1-s)^{-\beta} dW_s = O_p(\Delta_n^{1/2-\beta}),$$

so that the drift can dominate volatility when $\alpha - \beta > 1/2$.

Remark 2.2. As for our preliminary discussion, our modeling choice should be considered as a technical way to represent the fact that, in some situations, the drift might be as large as to dominate the volatility component. Our specification is thus meant to capture drift explosions, as in [Christensen et al. \(2022\)](#), drifts following an Ornstein-Uhlenbeck (OU) model, as in [Laurent and Shi \(2020, 2022\)](#); [Phillips et al. \(2015\)](#); [Phillips and Shi \(2019\)](#), or just large constant drifts, as in one of our simulation settings below.

Remark 2.3. The no-arbitrage condition requires $\alpha - \beta < 1/2$, so that the so-called structural condition:

$$\int_0^1 \left(\frac{\mu_s (1-s)^{-\alpha}}{\sigma_s (1-s)^{-\beta}} \right)^2 ds = \int_0^1 \frac{\mu_s^2}{\sigma_s^2} (1-s)^{-2(\alpha-\beta)} ds < \infty$$

is satisfied. This means that for the drift strictly prevailing over volatility in high-frequency returns, absence of arbitrage has to be violated locally. Of course, this is only an asymptotic condition. The actual possibility of arbitrage violation is more nuanced in small samples.

3 The estimators

We assume to observe X (the logarithmic price) at $n + 1$ equally spaced points $0, \Delta_n, 2\Delta_n, \dots, n\Delta_n = 1$ in $[0, 1]$, with $\Delta_n = 1/n$. We write $t_i = i\Delta_n$ with $i = 0, \dots, n$. The high-frequency return is denoted by $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$.

3.1 Realized volatility

The first estimator we consider is the well-known realized volatility, defined as:

$$\text{RV} = \sum_{i=1}^n (\Delta_i^n X)^2. \quad (6)$$

We first show that RV remains a consistent estimator of the integrated variance

$$\text{IV} = \int_0^1 \sigma_s^2 (1-s)^{-2\beta} ds, \quad (7)$$

even under model (2) with drift and volatility explosions.

Lemma 3.1. *Under Assumption 2.1, as $n \rightarrow \infty$, we have $\text{RV} \xrightarrow{p} \text{IV}$.*

The intuition for the above result is the following. Assume $\mu_t = 1$ and $\sigma_t = 0$. We then have:

$$\text{RV} = \sum_{i=1}^n \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\alpha} ds \right)^2 = \frac{\Delta_n^{2-2\alpha}}{(1-\alpha)^2} \underbrace{\sum_{j=1}^n (j^{1-\alpha} - (j-1)^{1-\alpha})^2}_{\text{convergent when } \alpha > 1/2}. \quad (8)$$

Thus, even an explosive drift vanishes asymptotically in realized variance (in the case $\alpha \leq 1/2$, convergence to zero is even faster) at rate $\Delta_n^{2-2\alpha}$. This calculation is generalized in the proof of Theorem 3.1 to the case with the coefficients μ_t and σ_t as in Assumption 2.1. However, Eq. (8) also suggests that the bias due to the drift term can be the leading term in the asymptotic error of RV for sufficiently large α , since the vanishing rate can get arbitrarily slow. For example, when $\beta = 0$, the order of the bias term is $\Delta_n^{2-2\alpha}$ while the order of the variance term is $\Delta_n^{1/2}$, so that the bias dominates when $\alpha > 3/4$. If volatility explodes too, convergence of RV to IV is also slower. This intuition is proved formally, in the general case, in Theorem 3.1 below.

For its statement, we need to introduce two quantities which are related, in the case of constant coefficients μ_t and σ_t , to the Riemann ζ , a function which comes out naturally from the limit of the series in (8). To unify the notation across theorems, we define the quantity $\zeta_{(c_1, d_1, k_1), \dots, (c_M, d_M, k_M)}^f$ for a stochastic process f using M triplets (c_m, d_m, k_m) with $m = 1, \dots, M$, and by $K = \max_{m=1, \dots, M}(k_m)$, as

$$\zeta_{(c_1, d_1, k_1), \dots, (c_M, d_M, k_M)}^f := \lim_{n \rightarrow \infty} \sum_{j=1}^{n-K} f_{(n-j)\Delta_n} \prod_{m=1}^M \frac{[(j+k_m)^{1-c_m} - (j+k_m-1)^{1-c_m}]^{d_m}}{(1-c_m)^{d_m}}. \quad (9)$$

Theorem 3.1. *Under Assumption 2.1, as $n \rightarrow \infty$, the limiting distribution of $\text{RV} - \text{IV}$ is as follows.*

(1) *When $\beta < 1/4$ and $\alpha < 3/4$:*

$$n^{1/2} [\text{RV} - \text{IV}] \xrightarrow{d} \mathcal{N} \left(0, 2 \int_0^1 \sigma_s^4 (1-s)^{-4\beta} ds \right).$$

(2) *When $\beta > 1/4$ and $\alpha - \beta < 1/2$:*

$$n^{1-2\beta} [\text{RV} - \text{IV}] \xrightarrow{d} \mathcal{N} \left(0, 2\zeta_{(2\beta, 2, 0)}^{\sigma^4} \right).$$

(3) *When $\beta < 1/4$, $\alpha > 3/4$ and $\alpha + \beta < 1$:*

$$n^{1/2} \left[\text{RV} - \text{IV} - \Delta_n^{2-2\alpha} \zeta_{(\alpha, 2, 0)}^{\mu^2} \right] \xrightarrow{d} \mathcal{N} \left(0, 2 \int_0^1 \sigma_s^4 (1-s)^{-4\beta} ds \right).$$

(4) *When $\alpha + \beta > 1$ and $\alpha - \beta > 1/2$:*

$$n^{3/2-\alpha-\beta} \left[\text{RV} - \text{IV} - \Delta_n^{2-2\alpha} \zeta_{(\alpha, 2, 0)}^{\mu^2} \right] \xrightarrow{d} \mathcal{N} \left(0, 4\zeta_{(\alpha, 2, 0), (2\beta, 1, 0)}^{\mu^2 \sigma^2} \right).$$

The above convergences are stable in law. See Corollary B.1 in the Online Appendix B for border cases.

Theorem 3.1 shows that the asymptotic distribution of RV depends on the values of the drift explosion rate (α) and the volatility explosion rate (β). We identify four main cases in the (α, β) space, illustrated in Figure 1. In case (1), in which both explosions in drift and in volatility are “moderate”, we recover the standard

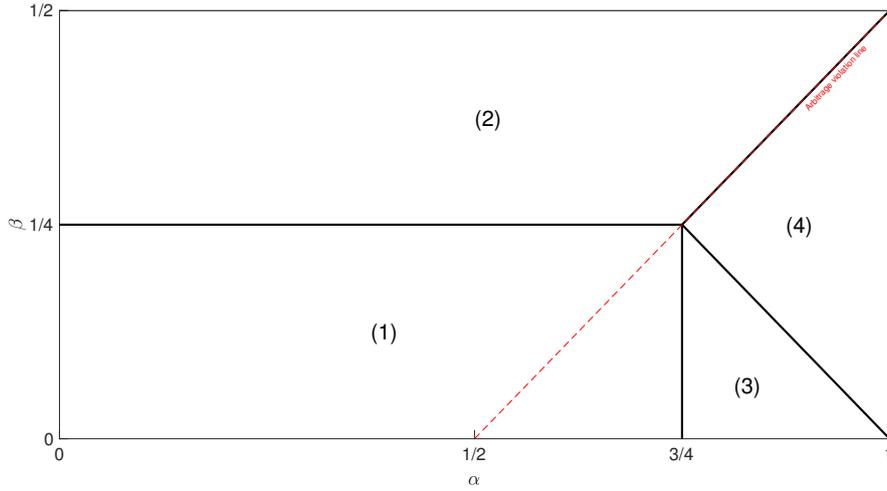


Figure 1: The different regions in the (α, β) plain for convergence in Theorems 3.1, 3.2, 3.3.

limit for RV. Case (1) is local to the standard model in which $\alpha = \beta = 0$. Even in this case the arbitrage condition may be violated when $\alpha - \beta > 1/2$. The variance of RV can get arbitrarily large as $\beta \rightarrow 1/4$ from the left, since $\int_0^1 (1-s)^{-4\beta} ds = 1/(1-4\beta)$. In case (2), volatility explosion dominates drift explosion. In this case the convergence rate of RV to IV is slowed down with respect to case (1). In cases (3) and (4) the bias term due to drift explosion dominates the variance term (and arbitrage is always violated). In all cases, consistency is however preserved since the bias term is of order $\Delta_n^{2-2\alpha} \rightarrow 0$ with $\alpha < 1$, so it is always vanishing, even if the vanishing rate can be arbitrarily slow as α approaches 1. The difference between case (3) and case (4) is in the dominating term in the variance, and the corresponding convergence rate which is slower in (4) than in (3).

3.2 Rice estimator

To attenuate the bias due to the drift we borrow from the difference estimator proposed by Rice (1984) (see Hans-Georg and Stadtmüller (1988) and Hall et al. (1990) for extensions and refinements, and Von Neumann et al. (1941) for an early treatment and an historical discussion of the estimator), defined, for a given

integer $k > 0$, as

$$\text{RiceV}(k) = \frac{1}{2} \sum_{i=k+1}^n (\Delta_i^n X - \Delta_{i-k}^n X)^2. \quad (10)$$

The estimator (10) is still consistent, as RV, for integrated variance. The idea of this estimator is simply to eliminate the drift by taking first differences (in our case, of returns). The following theorem shows that the estimator based on this commonly used technique in nonparametric regression is still consistent, but it cannot eliminate the asymptotic bias. However, it can strongly attenuate it, at the cost of an inflated variance.

Theorem 3.2. *Under Assumption 2.1, as $n \rightarrow \infty$, $\text{RiceV}(k) \xrightarrow{p} \int_0^1 \sigma_s^2 (1-s)^{-2\beta} ds$. The limiting distribution of $\text{RiceV}(k) - \text{IV}$ is as follows.*

(1) When $\beta < 1/4$ and $\alpha < 3/4$:

$$n^{1/2} [\text{RiceV}(k) - \text{IV}] \xrightarrow{d} \mathcal{N} \left(0, 3 \int_0^1 \sigma_s^4 (1-s)^{-4\beta} ds \right).$$

(2) When $\beta > 1/4$ and $\alpha - \beta < 1/2$:

$$n^{1-2\beta} [\text{RiceV}(k) - \text{IV}] \xrightarrow{d} \mathcal{N} \left(0, 2\zeta_{(2\beta,2,0)}^{\sigma^4} + \zeta_{(2\beta,1,k),(2\beta,1,0)}^{\sigma^4} \right).$$

(3) When $\beta < 1/4$, $\alpha > 3/4$ and $\alpha + \beta < 1$:

$$n^{1/2} \left[\text{RiceV}(k) - \text{IV} - \Delta_n^{2-2\alpha} \left(\zeta_{(\alpha,2,0)}^{\mu^2} - \zeta_{(\alpha,1,0),(\alpha,1,k)}^{\mu^2} \right) \right] \xrightarrow{d} \mathcal{N} \left(0, 3 \int_0^1 \sigma_s^4 (1-s)^{-4\beta} ds \right).$$

(4) When $\alpha + \beta > 1$ and $\alpha - \beta > 1/2$:

$$n^{3/2-\alpha-\beta} \left[\text{RiceV}(k) - \text{IV} - \Delta_n^{2-2\alpha} \left(\zeta_{(\alpha,2,0)}^{\mu^2} - \zeta_{(\alpha,1,0),(\alpha,1,k)}^{\mu^2} \right) \right] \xrightarrow{d} \mathcal{N} \left(0, V_4^{\text{RiceV}}(\alpha, \beta) \right),$$

where

$$\begin{aligned} V_4^{\text{RiceV}}(\alpha, \beta) = & 4\zeta_{(\alpha,2,0),(2\beta,1,0)}^{\mu^2\sigma^2} + \zeta_{(2\beta,1,k),(\alpha,2,0)}^{\mu^2\sigma^2} + \zeta_{(2\beta,1,0),(\alpha,2,k)}^{\mu^2\sigma^2} \\ & - 4\zeta_{(\alpha,1,k),(2\beta,1,k),(\alpha,1,0)}^{\mu^2\sigma^2} - 4\zeta_{(\alpha,1,k),(2\beta,1,0),(\alpha,1,0)}^{\mu^2\sigma^2} \\ & + 2\zeta_{(\alpha,1,2k),(2\beta,1,k),(\alpha,1,0)}^{\mu^2\sigma^2}. \end{aligned} \quad (11)$$

The above convergences are stable in law. See Corollary B.2 in the Online Ap-

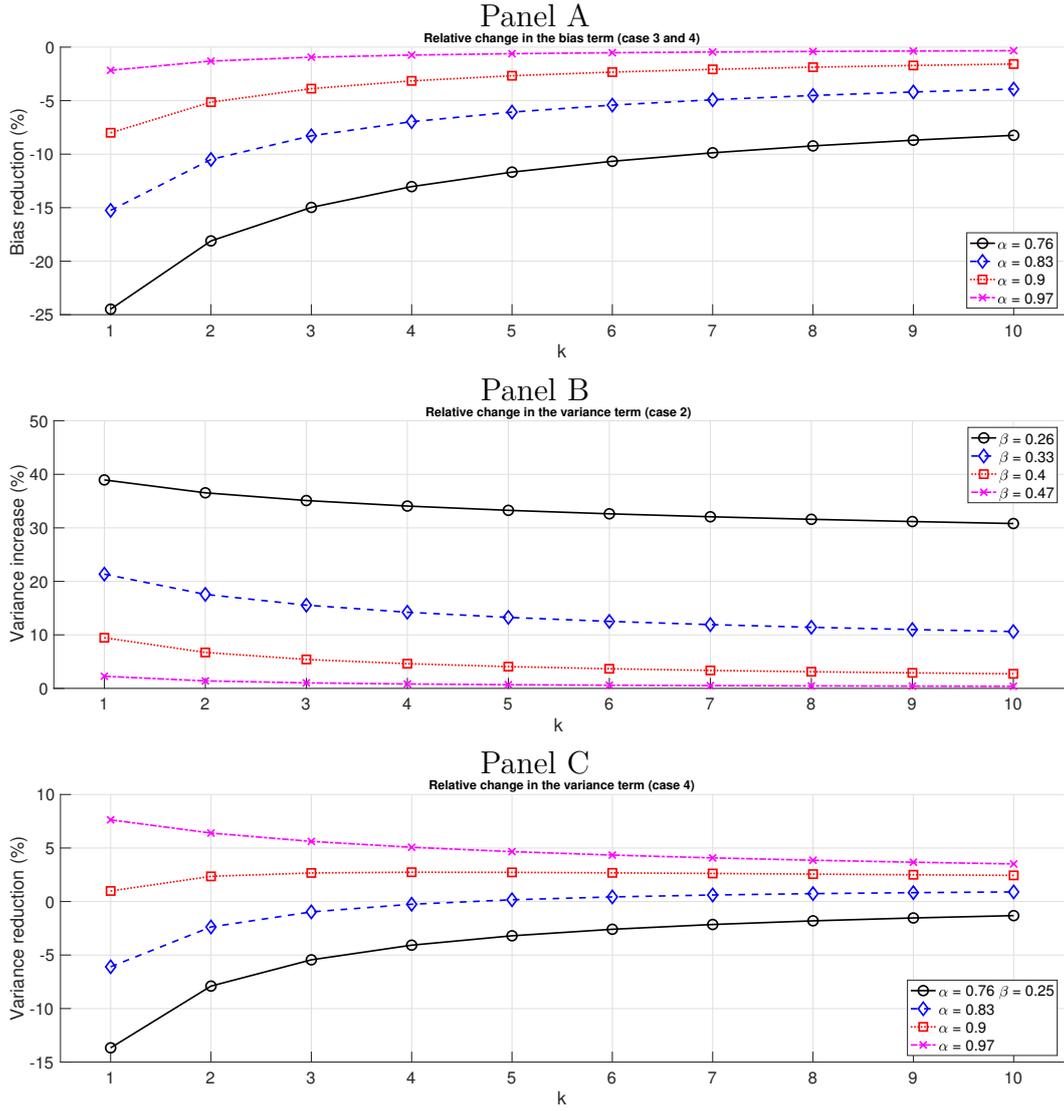


Figure 2: The attenuation of the bias and change in the variance of the $\text{RiceV}(k)$ estimator, as a function of k , relative to the RV estimator, in the case with constant μ_t and σ_t in model (2).

pendix B for border cases.

The structure of Theorem 3.2 is similar to that of Theorem 3.1, since the same four areas in the $\alpha - \beta$ region in Figure 1 are identified. In case (1), the only effect of the $\text{RiceV}(k)$ estimator is to inflate the asymptotic variance of 50%. The variance is also inflated in case (2), where the inflation depends on β and k . Figure 2 Panel B illustrates the inflation rate in the case of constant σ_t for $\text{RiceV}(k)$, showing that it decreases with k and with the explosion rate β . In particular, with exploding

volatility the advantage in precision of RV versus RiceV(k) attenuates strongly. In case (4), as shown in Panel C of Figure 2, the variance of the RiceV(k) estimator can even be *smaller* than that of RV. Panel A of Figure 2 shows the bias reduction as a function of k in the case of constant drift coefficient μ_t and when the drift is dominating, that is in case (3) and (4). The size of the reduction is larger with smaller α and it is largest for $k = 1$. As Figure 2 shows, we expect a bias reduction up to 25% for sensible values of α .

The estimator (10) has also been studied by Andersen et al. (2021). They also find a reduction in the bias with respect to RV. In our setting, there are two main differences. The first is that we have volatility explosions, and we study the interplay between the drift and volatility explosion rates in determining the asymptotic limit (Figure 1). The second difference is that they truncate returns whose absolute value is above the threshold $c\Delta_n^\omega$, with c and ω being constants. As a result, they also find a bias reduction driven by a faster rate of convergence of the bias to zero. The fact that their rate of convergence of the bias is faster comes from truncation. As discussed in the introduction, truncating will eliminate the largest returns due to drift explosion. In our case, in which we do not truncate, the rate of convergence of the bias to zero is the same as that for RV ($\Delta_n^{2-2\alpha}$), and the reduction in the bias comes from a smaller constant in front of the bias term. An important motivation for not truncating comes from the analysis of the next estimator.

3.3 Realized autocovariance and realized drift

We finally come to our definition of *realized drift*. It is clear now that the difference RiceV(k) – RV can be dominated by the bias (at the first order) and therefore proportional to the squared drift when the drift is large enough. Simple math however shows that this difference is, apart from a negligible end-effect, the realized autocovariance estimator:

$$\text{RAC}(k) = \sum_{i=k+1}^n \Delta_i^n X \Delta_{i-k}^n X. \quad (12)$$

From the proof of Theorem 3.2, we indeed have:

$$\text{RAC}(k) = \text{RV} - \text{RiceV}(k) + o_p(1),$$

where the $o_p(1)$ term is the end effect. This means that, in model (2), the autocovariance estimator is actually estimating drift. This intuition is formalized in the following Theorem.

Theorem 3.3. *Under Assumption 2.1, as $n \rightarrow \infty$, $\text{RAC}(k) \xrightarrow{p} 0$. The limiting distribution of $\text{RAC}(k)$ is as follows.*

(1) *When $\beta < 1/4$ and $\alpha < 3/4$:*

$$n^{1/2} \text{RAC}(k) \xrightarrow{d} \mathcal{N} \left(0, \int_0^1 \sigma_s^4 (1-s)^{-4\beta} ds \right).$$

(2) *When $\beta > 1/4$ and $\alpha - \beta < 1/2$:*

$$n^{1-2\beta} \text{RAC}(k) \xrightarrow{d} \mathcal{N} \left(0, \zeta_{(2\beta,1,k),(2\beta,1,0)}^{\sigma^4} \right).$$

(3) *When $\beta < 1/4$, $\alpha > 3/4$ and $\alpha + \beta < 1$:*

$$n^{1/2} \left[\text{RAC}(k) - \Delta_n^{2-2\alpha} \zeta_{(\alpha,1,0),(\alpha,1,k)}^{\mu^2} \right] \xrightarrow{d} \mathcal{N} \left(0, \int_0^1 \sigma_s^4 (1-s)^{-4\beta} ds \right).$$

(4) *When $\alpha + \beta < 1$ and $\alpha - \beta > 1/2$:*

$$n^{3/2-\alpha-\beta} \left[\text{RAC}(k) - \Delta_n^{2-2\alpha} \zeta_{(\alpha,1,0),(\alpha,1,k)}^{\mu^2} \right] \xrightarrow{d} \mathcal{N} \left(0, V_4^{\text{RAC}}(\alpha, \beta) \right),$$

where

$$V_4^{\text{RAC}}(\alpha, \beta) = \zeta_{(2\beta,1,k),(\alpha,2,0)}^{\mu^2\sigma^2} + \zeta_{(2\beta,1,0),(\alpha,2,k)}^{\mu^2\sigma^2} + \zeta_{(2\beta,1,k),(\alpha,1,0),(\alpha,1,2k)}^{\mu^2\sigma^2}. \quad (13)$$

The above convergences are stable in law. See Corollary B.3 in the Online Appendix B for border cases.

We see from Theorem 3.3 that, even if RAC vanishes asymptotically, it does so with a different rate in the four different cases. In particular, in cases (3) and (4), the dominating order of RAC is proportional to the squared drift. Our notion of

realized drift thus comes from the results in cases (3) and (4). When we are in these situations, in which drift explodes at a sufficiently high rate, we have:

$$\text{RDrift}(k) := \frac{1}{\Delta_n^{2-2\alpha}} \text{RAC}(k) \xrightarrow{p} \zeta_{(\alpha,1,0),(\alpha,1,k)}^{\mu^2}.$$

When $\mu_t = \mu$, that is when the drift coefficient is constant, we have by definition (9)

$$\zeta_{(\alpha,1,0),(\alpha,1,k)}^{\mu^2} = \mu^2 \frac{\bar{\zeta}_{k,\alpha}}{(1-\alpha)^2},$$

where

$$\bar{\zeta}_{k,\alpha} = \sum_{j=1}^{+\infty} (j^{1-\alpha} - (j-1)^{1-\alpha})((j+k)^{1-\alpha} - (j+k-1)^{1-\alpha})$$

is a constant. For example, with constant $\mu_t = \mu$ in model (2) (such that the total drift is $\mu(1-t)^{-\alpha}$) and $\alpha = 0.9$, we have:

$$\frac{1}{\Delta_n^{0.2}} \text{RAC}(1) \xrightarrow{p} 8.10\mu^2,$$

and

$$\frac{1}{\Delta_n^{0.2}} \text{RAC}(2) \xrightarrow{p} 5.20\mu^2.$$

Theorem 3.3 allows us to make the following predictions:

1. If drift is “small”, RAC is centered around zero.
2. If drift is “large”, RAC is centered on a positive value which grows with μ_t^2 and declines with increasing Δ_n .

Remark 3.1. *In our main model (2), there is a single explosion (of drift and volatility) at the end-point. We could have an explosion at the beginning, or multiple explosions (as in a V-shape, see [Flora and Renò, 2021](#)). In this case, the structure of Theorems 3.1, 3.2, 3.3 would remain unchanged with the same four cases and the same rates of convergence for the bias and the variance term in each case. What would change is the coefficient (straightforwardly) in cases (2), (3) and (4).*

Remark 3.2. *We can safely introduce jumps in the primitive process (2) without altering the intuition of Theorem 3.3. Jumps indeed will be asymptotically eliminated via the same mechanism which eliminates the bias in bipower variaton, that*

is the product of two consecutive returns, see e.g. [Barndorff-Nielsen and Shephard \(2006\)](#). Jump will just impact the asymptotic variance in a straightforward way in all four cases.

3.4 Market microstructure noise

Market microstructure noise is going to play a role in RAC, actually dominating it completely in the limit. In this sense, our results can be seen as complementary to those of the celebrated illiquidity estimator of [Roll \(1984\)](#). In his model, the bid-ask spread can be recovered from the serial autocovariance, but only when it is negative (since that would be the signature of the bid-ask bounce). In our model, in which frictions are absent, a positive serial autocovariance is a signature of the presence of large drift. If the microstructure noise is present, Δ_n should be small (to get closer to the asymptotic limit) but at the same time large enough to avoid distortions from market microstructure noise.

Recently, new approaches have been proposed to estimate the serial autocovariance in realistic frictional models, see, for example, [Jacod et al. \(2017\)](#) and [Li and Linton \(2022\)](#). Consider the noise contaminated data

$$X_{i\Delta_n}^o = X_{i\Delta_n} + \omega\varepsilon_i, \quad (14)$$

where $X_{i\Delta_n}$ is the log prices of the underlying asset with its dynamic given in [Assumption 2.1](#), $\omega > 0$, and ε_i is the noise, as defined below.

Assumption 3.4. *Assume that the noise component $\{\varepsilon_i\}_{i=1}^n$ is independent of X , stationary, mean zero, variance one, and with finite moments of all orders. Denote by $\gamma_s = \mathbb{E}(\varepsilon_i, \varepsilon_{i+s})$ for any integer $s \geq 0$. We further assume that for some $v > 1$ and $K > 0$,*

$$|\gamma_k| \leq \frac{K}{s^v}.$$

The realized autocovariance can be computed from pre-averaged noise-contaminated returns $\bar{r}_{i\Delta_n}^o$ in the spirit of [Jacod et al. \(2009\)](#). That is,

$$\text{RAC}^o(k_n) = \sum_{i=k_n+1}^n \bar{r}_{i\Delta_n}^o \bar{r}_{(i-k_n)\Delta_n}^o, \quad (15)$$

where the pre-averaged return

$$\bar{r}_{i\Delta_n}^o = \sum_{j=1}^{l_n} g\left(\frac{j}{l_n}\right) (X_{(i+j)\Delta_n}^o - X_{(i+j-1)\Delta_n}^o)$$

with $g(\cdot)$ being the weight function (see Assumption A.1 in the Appendix for its precise definition) and l_n being the pre-averaging window. We assume that the pre-averaging window is smaller than autocovariance lag order (i.e., $l_n < k_n$) and l_n diverges to infinity as $n \rightarrow \infty$.

The pre-averaged noise contaminated return can be rewritten as

$$\bar{r}_{i\Delta_n}^o = - \sum_{j=0}^{l_n-1} h_j^n X_{(i+j)\Delta_n}^o = - \sum_{j=0}^{l_n-1} h_j^n [X_{(i+j)\Delta_n} + \omega \varepsilon_{(i+j)\Delta_n}] = \bar{r}_{i\Delta_n} - \omega \sum_{j=0}^{l_n-1} h_j^n \varepsilon_{(i+j)\Delta_n},$$

where $h_j^n = g\left(\frac{j+1}{l_n}\right) - g\left(\frac{j}{l_n}\right)$ and $\bar{r}_{i\Delta_n} = - \sum_{j=0}^{l_n-1} h_j^n X_{(i+j)\Delta_n}$. The realized autocovariance consists of four components:

$$\begin{aligned} \text{RAC}^o(k_n) = & \sum_{i=k_n+1}^n \left[\bar{r}_{i\Delta_n} \bar{r}_{(i-k_n)\Delta_n} - \omega \bar{r}_{i\Delta_n} \sum_{j=0}^{l_n-1} h_j^n \varepsilon_{(i-k_n+j)\Delta_n} \right. \\ & \left. - \omega \bar{r}_{(i-k_n)\Delta_n} \sum_{j=0}^{l_n-1} h_j^n \varepsilon_{(i+j)\Delta_n} + \omega^2 \sum_{j_1=0}^{l_n-1} h_{j_1}^n \varepsilon_{(i-k_n+j_1)\Delta_n} \sum_{j_2=0}^{l_n-1} h_{j_2}^n \varepsilon_{(i+j_2)\Delta_n} \right]. \end{aligned} \quad (16)$$

The first term of (16) is the signal component, while the remaining three terms are brought by the noise. We show in the theorem below that RAC^o is dominated by the signal term when the pre-averaging window l_n satisfies suitable conditions.

Theorem 3.5. *Under the model specification of (14) and Assumption 3.4 and A.1, as $n \rightarrow \infty$, the realized autocovariance estimator*

$$\text{RAC}^o(k_n) = \sum_{i=k_n+1}^n \bar{r}_{i\Delta_n} \bar{r}_{(i-k_n)\Delta_n} + o_p(1) \quad (17)$$

conditional on $\Delta_n^{2(1-\alpha)} l_n^3 \rightarrow \infty$, where $l_n < k_n$ and $l_n, k_n \rightarrow \infty$.

The above Theorem (whose proof can be found in the Online Appendix C) shows that pre-averaging data before applying the RAC estimator delivers an estimator that is equivalent to RAC applied to uncontaminated data. To get this, we need to pre-average on enough data such that we can “clean” market microstructure noise without smoothing the exploding drift (according to the condition $\Delta_n^{2(1-\alpha)} l_n^3 \rightarrow \infty$).

A comprehensive investigation of the limiting properties of $\text{RAC}^o(k_n)$ is left for future research.

4 Implications for volatility forecasting

Our new asymptotic theory for the more flexible data generating process (2) has relevant direct implications for volatility forecasting. This is a direct consequence of the decomposition:

$$\text{RV}_t \simeq \text{RiceV}_t + \text{RAC}_t, \quad (18)$$

where t denotes the measurement day, and, here and in what follows, we write $\text{RiceV}_t = \text{RiceV}_t(1)$ and $\text{RAC}_t = \text{RAC}_t(1)$. RiceV_t is designed to capture volatility, while RAC_t to capture drift. If the two components have independent dynamics (a conjecture which is corroborated by our empirical analysis below), that should improve the forecasting of realized volatility, a topic which has been the subject of an extensive empirical literature. Moreover, the reasoning put forward in [Bollerslev et al. \(2016\)](#) implies that additional interaction terms should be added to improve forecasts.

To clarify and formalize this intuition define ‘Integrated Variance’ as $\text{IV}_t = \int_{t-1}^t \sigma_s^2 (1-s)^{-2\beta} ds$ and ‘Integrated Drift’ as $\text{ID}_t = \Delta_n^{2-2\alpha} \zeta_{(\alpha,1,0)(\alpha,1,1)}^{\mu^2}$. Then, based on Theorems 3.2 and 3.3, we write³

$$\begin{aligned} \text{RiceV}_t &\simeq \text{IV}_t + v_t, \\ \text{RAC}_t &\simeq \text{ID}_t + \omega_t, \end{aligned} \quad (19)$$

where v_t and ω_t are independent mixture normal distributions with mean zero and variance depending on Δ_n , α and β . It follows from Eq. (18) that

$$\text{RV}_t \simeq \text{ID}_t + \text{IV}_t + \omega_t + v_t. \quad (20)$$

Now assume that the true dynamics of the true, unobservable objects is given by

³In Eq. (19) we are neglecting the bias term in RiceV . However, this can be added easily since it is roughly proportional (and exactly proportional when μ_t is a constant) to ID_t (see Theorem 3.2). Results in this section would not change.

AR(1) processes with independent shocks, namely:

$$\begin{aligned} \mathbf{IV}_t &= \phi_0 + \phi_1 \mathbf{IV}_{t-1} + u_{1t}, \\ \mathbf{ID}_t &= \gamma_0 + \gamma_1 \mathbf{ID}_{t-1} + u_{2t}, \end{aligned} \quad (21)$$

where u_{1t} , u_{2t} are independent (among them and with v_t, ω_t) *i.i.d.* noise. It follows immediately that

$$\mathbf{RV}_t \simeq \phi_0 + \gamma_0 + \phi_1 \mathbf{RV}_{t-1} + (\gamma_1 - \phi_1) \mathbf{RAC}_{t-1} + v_t + \omega_t - \phi_1 v_{t-1} - \gamma_1 \omega_{t-1} + u_{1t} + u_{2t}. \quad (22)$$

Thus, our theory implies, when $\gamma_1 \neq \phi_1$, a direct impact of RAC on future realized volatility, which has been unexplored by the forecasting literature so far.

Moreover, assume that we misspecify the realized variance dynamics as:

$$\mathbf{RV}_t = \beta_0 + \beta_1 \mathbf{RV}_{t-1} + \alpha_1 \mathbf{RAC}_{t-1} + \varepsilon_t. \quad (23)$$

That is, we ignore the moving average components in Eq. (22). Following [Bollerslev et al. \(2016\)](#), it is immediate to show that

$$\begin{aligned} \beta_1 &\simeq \phi_1 \left(1 + \frac{\mathbb{V}(v_{t-1})}{\mathbb{V}(\mathbf{IV}_{t-1})} \right)^{-1}, \\ \alpha_1 &\simeq \gamma_1 \left(1 + \frac{\mathbb{V}(\omega_{t-1})}{\mathbb{V}(\mathbf{IV}_{t-1})} \right)^{-1} - \phi_1. \end{aligned}$$

Then, if we use the asymptotic theory for RV and RAC in Case (1) and write

$$\mathbb{V}(v_{t-1}) \simeq 2\Delta_n \mathbf{IQ}_{t-1}, \quad \mathbb{V}(\omega_{t-1}) \simeq \Delta_n \mathbf{IQ}_{t-1},$$

where $\mathbf{IQ}_t = \int_{t-1}^t \sigma_s^4 (1-s)^{-4\beta} ds$, using Taylor expansion we have:

$$\beta_1 \simeq \phi_1 \left(1 - \frac{2\Delta_n \cdot \mathbf{IQ}_{t-1}}{\mathbb{V}(\mathbf{IV}_{t-1})} \right), \quad \alpha_1 \simeq \gamma_1 \left(1 - \frac{\Delta_n \cdot \mathbf{IQ}_{t-1}}{\mathbb{V}(\mathbf{IV}_{t-1})} \right) - \phi_1,$$

which predicts a negative load of $\mathbf{IQ}_{t-1} \mathbf{RV}_{t-1}$ and $\mathbf{IQ}_{t-1} \mathbf{RAC}_{t-1}$ on future realized variance. The first term is the main feature of the HARQ model of [Bollerslev et al. \(2016\)](#). The second term is predicted by our theory, and is also unexplored by the forecasting literature.

The new predicted effects should hold even when there is no dependence between the \mathbf{IV}_t and \mathbf{ID}_t dynamics, as postulated in the model (21). Of course, drift may

predicts future integrated volatility and vice versa. Moreover, drift could impact $\text{Rice}V_t$ through its bias term, which we ignored here for simplicity. In these cases (which are actually found to be relevant in our empirical analysis), the predicted effects on realized variance forecasting are reinforced.

Summarizing, the theoretical treatment in Section 3 implies that, in the classical realized volatility forecasting exercise, two new terms should be considered as explanatory variables: the realized autocovariance, and an interaction term in which realized autocovariance is multiplied by an integrated quarticity estimator. Before testing these implications on real data, we study the properties of the realized autocovariance in the presence of price drift on simulated data.

5 Simulation study

The purpose of this section is to show that drift is not invisible to high-frequency data and that it can be detected, in small samples, from the realized autocovariance estimator even in presence of stochastic volatility, intraday volatility effects and market microstructure noise.

The three models considered in this section are embedded in the following specification for the observed log-prices X_t :

$$X_t = M_t + Z_t + \epsilon_t, \text{ for } t = 0, \dots, 1, \quad (24)$$

where

$$dM_t = \begin{cases} \mu^c dt & \text{Constant drift} \\ \mu^l M_t dt & \text{OU or linear drift} \\ \mu^e (1-t)^{-\alpha} dt & \text{Nonlinear drift} \end{cases} \quad (25)$$

and

$$dZ_t = v_t dW_t \text{ with } v_t = \sigma_t f_t \quad (26)$$

$$d\sigma_t^2 = \kappa(\gamma - \sigma_t^2)dt + \zeta \sigma_t dW_t^\sigma \quad (27)$$

$$\epsilon_t \sim N(0, 0.25\Delta_n v_t^2). \quad (28)$$

The drift is specified through M_t . The first model is a **Constant Drift Model** whose magnitude is governed by μ^c . The second one is an OU process, as in [Laurent and Shi \(2020\)](#). We refer to this model as **Linear Drift Model** as it generates a linear drift when $\mu^l \neq 0$. Finally, the third model, called **Nonlinear Drift Model**, has an explosive drift as in (2). We fix $\alpha = 0.9$ and $\tau = 1$ so that this DGP corresponds to case (3) of [Theorem 3.3](#), where the magnitude of the drift is controlled through the μ^e parameter only. A zero-drift model can be obtained by setting either $\mu^c = 0$, $\mu^l = 0$ or $\mu^e = 0$. Volatility burst is not considered in this simulation, and we use the same volatility trajectory for all models (and all replications) for ease of comparison.

Our theory predicts that for the given parameters the drift is the asymptotic dominating term only in the **Nonlinear Drift Model**, while for the other two the dominating term is the variance. However, it is clear that the bias induced by the drift will also be present in the other two models, despite asymptotically vanishing. The point of our simulation experiments is to study the behavior of the realized autocovariance in small samples, where the bias term could be non-negligible even with non-exploding drift. In this sense, this Section complements the asymptotic theoretical results in the previous sections.

In [Equation \(26\)](#), Z_t is an Itô semimartingale with zero drift and stochastic volatility v_t , while W_t and W_t^σ are two correlated standard Brownian motions with $E(dW_t dW_t^\sigma) = \rho dt$. The spot volatility v_t is the product of the conditional volatility σ_t (with a Heston specification) and a diurnal component f_t . The intraday periodicity is obtained as in [Laurent and Shi \(2020\)](#), matching with that of the Nasdaq composite index. We configure the variance process to match key features of real financial high-frequency data as in [Aït-Sahalia and Kimmel \(2007\)](#) and [Christensen et al. \(2022\)](#). The annualized parameters of the model are $(\kappa, \theta, \zeta, \rho) = (5, 0.0225, 0.4, -\sqrt{0.5})$. A value of $\theta = 0.0225$ implies an unconditional standard deviation of 15% per annum for log-returns. The first spot variance of each simulation is drawn from its stationary law, i.e. $\sigma_0^2 \sim \Gamma(2\kappa\theta\zeta^{-2}, 2\kappa\zeta^{-2})$. The third component of X_t , i.e. ϵ_t , is the market microstructure noise. The noise is conditionally heteroscedastic and positively related to the riskiness of the efficient log-price (e.g., [Bandi and Russell, 2011](#); [Oomen, 2006](#); [Kalnina and Linton, 2008](#))

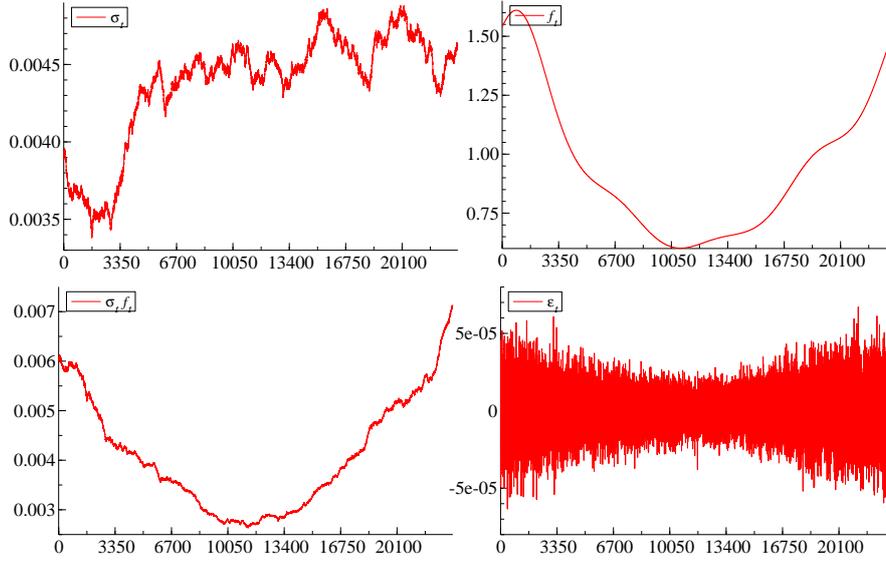


Figure 3: One typical sample path of the simulated Heston volatility σ_t , intraday periodicity f_t , spot volatility $\sigma_t f_t$, and microstructure noise ϵ_t at the one-second frequency over 6.5 hours.

and corresponds to a medium level of contamination (see [Christensen et al., 2014](#)). Figure 3 displays a typical sample path of the simulated 1-second Heston volatility σ_t , intraday periodicity f_t , spot volatility $\sigma_t f_t$, and microstructure noise ϵ_t over 6.5 hours. Typical sample paths of log-prices simulated from the **Constant Drift Model**, the **Linear Drift Model**, and the **Nonlinear Drift Model** (at the one-second frequency, over 6.5 hours, and with an initial logarithmic price of 7) are plotted in Figure 4. The parameters have been chosen so that the daily price change is about 2% for all models (i.e., $\mu^c = 0.2$, $\mu^l = 0.0028$ and $\mu^e = 0.78$). While the patterns of the **Constant Drift Model** and the **Linear Drift Model** are almost identical, the log-prices of the **Nonlinear Drift Model** show a clear exponential trend at the end of the day.

We compute $\text{RAC}(1)$ under various drift specifications. For each model, 101 values of the drift parameters μ^c , μ^l and μ^e are considered. For the **Constant Drift Model** we consider values of μ^c between 0 and 0.1. For the **Linear Drift Model** μ^l varies between 0 and 0.014. Finally, for the **Nonlinear Drift Model** μ^e varies between 0 and 0.0156. The parameters have been chosen so that the daily price change is

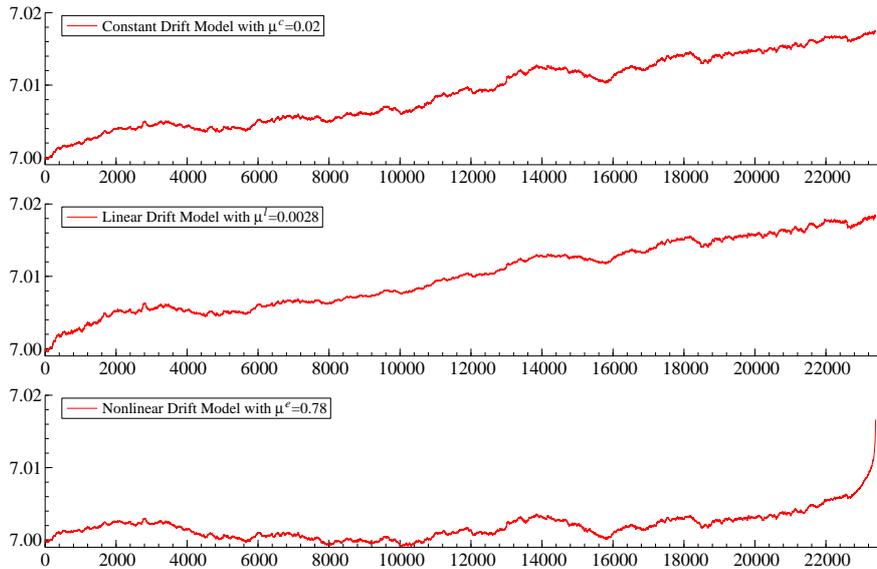


Figure 4: One typical sample path of the simulated log-prices of the Constant Drift Model, the Linear Drift Model, and the Nonlinear Drift Model at the 1-second frequency generating a return of about 2% over 6.5 hours.

maximum 5%. The values of μ^c , μ^l and μ^e are not very informative except that the larger they are the stronger the drift. For ease of comparison, we show on the x-axis of each graph the average daily returns, rather than the values of μ^c , μ^l and μ^e .

Confidence bands of $\text{RAC}(1)$ are computed as using two methods, an infeasible and a feasible one. For the infeasible method, 95% confidence bands are computed as RAC plus the 2.55% and 97.5% quantiles of the empirical distribution of RAC under the null of a zero drift. Feasible 95% confidence bands are computed as

$$[\text{RAC} + q_n(0.025) \text{std}_{\text{RAC}}, \text{RAC} + q_n(0.975) \text{std}_{\text{RAC}}], \quad (29)$$

where $q_n(\alpha)$ is the α quantile of the standard normal distribution.

The std_{RAC} under the null of no drift is proportional to IQ as shown in Theorem 3.2. The RQ estimator

$$\text{RQ} = \frac{1}{3\Delta_n} \sum_{i=1}^n (\Delta_i^n X)^4, \quad (30)$$

has been shown to be a consistent estimator of IQ (Barndorff-Nielsen and Shep-

hard, 2002a) when there is no drift burst. Based on the results for RV in Theorem 3.1, we expect the realized quarticity estimator to be biased in the presence of large drifts. A natural alternative is

$$\text{RiceQ} = \frac{1}{6\Delta_n} \sum_{j=3}^n (\Delta_j^n X - \Delta_{j-1}^n X)^2 (\Delta_{j-1}^n X - \Delta_{j-2}^n X)^2, \quad (31)$$

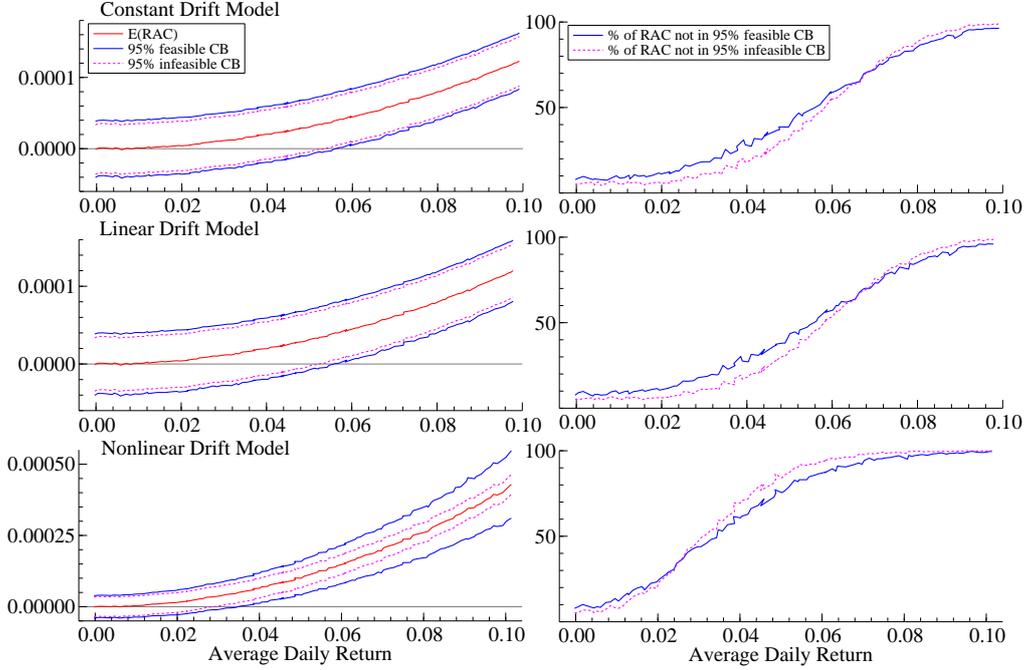
defined in the spirit of the RiceV(1) estimator for IV. Like RiceV, RiceQ is expected to reduce the bias of RQ when there is a large drift but be less efficient than RQ when there is a small or no drift. A drift robust estimator for the standard deviation std_{RAC} is therefore $\sqrt{\text{RiceQ}/n}$.

The former method is infeasible in practice as the true dynamic of Z_t is unknown, while the latter method does not require such knowledge a priori and hence it is feasible. These confidence bands can be used as a test in which the null hypothesis is region (1) (see Figure 1) against all other cases.

Figure 5 presents the average RAC(1) in the first column and the frequencies of rejections (i.e., percentages of RAC estimates outside confidence bands) in the second column. Here, we generate data at the one-second frequency and aggregate them at the 5-minute frequency (i.e., $\Delta_n = 300$). In each of the graphs on the left, the red dashed line $E(\text{RAC})$ is the average of 1,000 RAC(1) computed from the simulated data. The two blue solid (resp. purple dotted) lines correspond to the feasible (resp. infeasible) 95% confidence bands computed from RiceQ (resp. via simulations). The right panel contains three graphs displaying the percentages of RAC estimates outside of the 95% feasible (blue solid line) and infeasible (purple dotted line) confidence bands.

Evidently, RAC(1) increases with the magnitude of the drift. For a given level of daily return, RAC(1) responds more strongly to the non-linear drift than to the constant or linear drift. Indeed, for a daily return of 2% about 40% of the RAC(1) estimates from the **Nonlinear drift Model** are tested to be significantly different from zero, compared to about 15% for the **Constant Drift Model** and the **Linear Drift Model**. The feasible and infeasible confidence bands provide similar result, which is reassuring of the estimation accuracy of RiceQ for the asymptotic variance of RAC. Results also suggest that the feasible confidence bands are a

Figure 5: $RAC(1)$ computed from 5-minute returns as a function of the average daily return for the three models.



little bit too narrow (left panels) and lead to a small size distortion (right panels). This is a finite sample issue that can be resolved using data sampled at a higher frequency. Indeed, unreported results suggest that the size distortion is negligible when $RAC(1)$ and RiceQ are computed on 1 minute data.

6 Empirical Application

The database contains intraday prices of the Nasdaq Composite Index over a period spanning 2.5 decades from January 1996 to December 2020 for a total of 6,087 days. The data are obtained from Refinitive Tick History at the one-second frequency, filtered using standard techniques to remove large errors in the database, and aggregated at the 5-minute frequency to show that our theory has relevant implications for the most popular frequency in volatility forecasting.

Realized Drift Estimates

As it is new to the literature, the dynamic of RAC is itself interesting. The top panel of Figure 6 shows the daily estimates of $\text{RAC}(1)$, which are computed from 5-minute returns and standardized by $\sqrt{\text{RiceQ}/n}$ (consistent with our simulations), where $n = 78$ is the number of 5-minute returns per day. The standardized RAC is labelled as ‘z-stat’ on the graph. We see that the RAC estimator tends to be mostly positive, in keeping with our theory and the constant presence of a drift. In addition to the z-statistic, defined as $\text{RAC}(1)/\sqrt{\text{RiceQ}/n}$, the top panel of Figure 6 displays two solid lines corresponding to the 0.5% and 99.5% quantiles of the standard normal distribution. These values can serve as critical values (at the level 1% for a two sided test and 0.05% for one-sided alternatives) for a test of case (1) again all other cases (see again Figure 1).

The bottom panel of Figure 6 plots the logarithm of significantly positive RAC estimates only (i.e., values above the 99.5% quantile of the standard normal distribution). We can see that the null hypothesis is often rejected. With a two sided test at level 1% (resp. 10%), the null hypothesis is rejected in about 17.2 % (resp. 29.5%) of the days. Interestingly, at the 1% level, all rejections are due to a positive $\text{RAC}(1)$.⁴ These results indicate pervasive presence of drift in the data. Strongest rejections are observed in particular during the dotcom bubble of the late 1990s and early 2000s, the 2008 subprime mortgage crisis period, and at the onset of the coronavirus pandemic in 2020.

In-sample Estimation

We study the implications of our theory for volatility dynamics described in Section 4. In place of the pedagogical dynamics (23), we use the Heterogenous AutoRegressive (HAR) model of Corsi (2009). As in Section 4, we denote by $\text{RiceV}_t := \text{RiceV}_t(1)$ and by $\text{RAC}_t := \text{RAC}_t(1)$. For the interaction terms, which are modeled as in Bollerslev et al. (2016), an estimator for IQ is required. We use RiceQ (resp. RQ) instead of IQ for its interaction with RiceV and RAC (resp. RV). Furthermore, since the signal in the realized drift estimator is always positive, we replace the raw RAC series with $\text{RAC}^+ = \max(0, \text{RAC})$, when used for forecast-

⁴At the level 10%, we observe only 6 rejections with a negative $\text{RAC}(1)$.

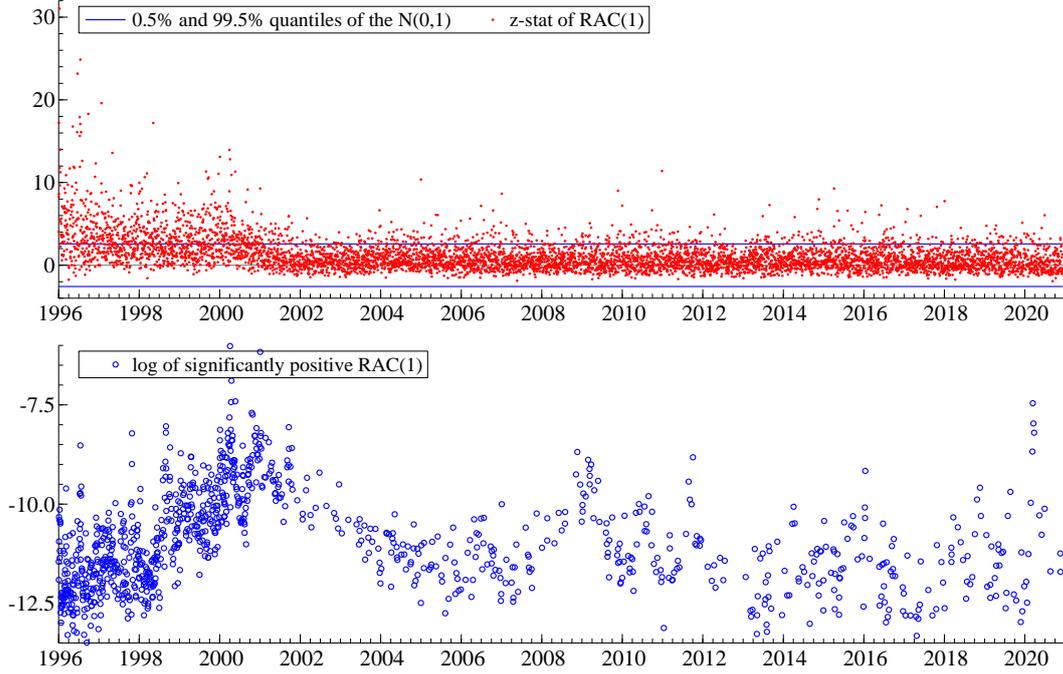


Figure 6: The RAC estimator on Nasdaq Composite Index 5-minute data from 1996 to 2020.

ing RiceV and RV.⁵ As such, we propose the following three general forecasting models:

$$\text{RAC}_t = \delta_0 + (\delta_d + \delta_d^Q \sqrt{\text{RiceQ}_{t-1}}) \text{RAC}_{t-1} + \delta_w \text{RAC}_{t-1:t-5} + \delta_m \text{RAC}_{t-1:t-22} + \varepsilon_{1t}, \quad (32)$$

$$\begin{aligned} \text{RiceV}_t = & \beta_0 + (\beta_d + \beta_d^Q \sqrt{\text{RiceQ}_{t-1}}) \text{RiceV}_{t-1} + \beta_w \text{RiceV}_{t-1:t-5} + \beta_m \text{RiceV}_{t-1:t-22} \\ & + (\gamma_d + \gamma_d^Q \sqrt{\text{RiceQ}_{t-1}}) \text{RAC}_{t-1}^+ + \gamma_w \text{RAC}_{t-1:t-5}^+ + \gamma_m \text{RAC}_{t-1:t-22}^+ + \varepsilon_{2t}, \end{aligned} \quad (33)$$

$$\begin{aligned} \text{RV}_t = & \alpha_0 + (\alpha_d + \alpha_d^Q \sqrt{\text{RQ}_{t-1}}) \text{RV}_{t-1} + \alpha_w \text{RV}_{t-1:t-5} + \alpha_m \text{RV}_{t-1:t-22} \\ & + (\lambda_d + \lambda_d^Q \sqrt{\text{RiceQ}_{t-1}}) \text{RAC}_{t-1}^+ + \lambda_w \text{RAC}_{t-1:t-5}^+ + \lambda_m \text{RAC}_{t-1:t-22}^+ + \varepsilon_t, \end{aligned} \quad (34)$$

referred to as RAC-HARQ, RiceV-HARQ-DQ and RV-HARQ-DQ, respectively, where ‘D’ is for drift and ‘Q’ for quarticity. Note that the model (34) for realized variance nests the HAR and the HARQ model.

Table 1 provides the summary statistics of RV, RiceV, RAC, RQ and RiceQ. Notice

⁵A natural alternative would be to consider only significantly positive values of RAC, e.g., $\text{RAC}^\dagger = \text{RAC}$ when $\text{RAC} > q_n(0.975) \text{std}_{\text{RAC}}$ and $\text{RAC}^\dagger = 0$ otherwise. We tested this specification but found RAC^+ to deliver better results.

Table 1: Summary Statistics on RV, RiceV, RAC, RQ and RiceQ computed on 5-minute returns

	RV $\times 10^2$	RiceV $\times 10^2$	RAC $\times 10^2$	RQ $\times 10^4$	RiceQ $\times 10^4$
Min	0.00024	0.00021	-0.03022	0.00000	0.00000
Mean	0.01123	0.00974	0.00119	0.00087	0.00055
Median	0.00500	0.00440	0.00032	0.00002	0.00002
Var	0.00046	0.00032	0.00003	0.00018	0.00002
Max	0.47138	0.28321	0.24352	0.85896	0.18173
Skewness	1.55677	1.35413	4.73689	10.04570	4.53927
Kurtosis	20.0791	14.2539	174.9571	588.5733	130.1705
ρ_1	0.71574	0.74351	0.22567	0.10821	0.24918
ρ_5	0.54095	0.58472	0.14932	0.05116	0.21694
ρ_{10}	0.46818	0.49955	0.16380	0.02631	0.12310

Note: The table provides summary statistics (i.e., minimum, mean, median, variance, maximum, skewness, kurtosis as well as the autocorrelations at lags 1, 5 and 10) on the key variables used in the empirical application. RV, RiceV and RAC have been multiplied by 10^2 while RQ and RiceQ have been multiplied by 10^4 .

that for readability of the table RV, RiceV and RAC have been multiplied by 10^2 while RQ and RiceQ have been multiplied by 10^4 . The long memory feature of RV and RiceV (to a lesser extent RAC) is obvious from their slow decaying autocorrelation functions. As expected, the RiceV estimator is much smoother than RV with smaller variance and narrower range. Similarly, among the two IQ estimators, RiceQ is smoother than RQ with less observations on the extreme tails (substantially smaller kurtosis).

To model the dynamic of RAC, we also consider a RAC-HAR model obtained by setting $\delta_d^Q = 0$ in (32). For RiceV, in addition to the RiceV-HARQ-DQ in (33), we consider four other models nested in this model, i.e.,

$$\text{RiceV-HAR: when } \gamma_d = \gamma_w = \gamma_m = \gamma_d^Q = \beta_d^Q = 0;$$

$$\text{RiceV-HARQ: when } \gamma_d = \gamma_w = \gamma_m = \gamma_d^Q = 0;$$

$$\text{RiceV-HAR-D: when } \beta_d^Q = \gamma_d^Q = 0;$$

$$\text{RiceV-HARQ-D: when } \gamma_d^Q = 0.$$

Similarly, for RV we rely on the RV-HARQ-DQ (34) as well as the following four specifications:

RV-HAR: when $\lambda_d = \lambda_w = \lambda_m = \lambda_d^Q = \alpha_d^Q = 0$;

RV-HARQ: when $\lambda_d = \lambda_w = \lambda_m = \lambda_d^Q = 0$;

RV-HAR-D: when $\alpha_d^Q = \lambda_d^Q = 0$;

RV-HARQ-D: when $\lambda_d^Q = 0$.

Estimates of the competing models for RAC, RiceV, and RV obtained on the full sample are reported respectively in Tables 2, 3 and 4. Several observations arise from the in-sample estimation results.

Table 2: In-sample estimation results for RAC_t

	RAC-HAR	RAC-HARQ
δ_0 : Constant	0.000 (5.466)	0.000 (4.454)
δ_d : RAC_{t-1}	0.036 (1.104)	0.207 (2.486)
δ_w : $\text{RAC}_{t-1:t-5}$	0.225 (2.201)	0.220 (2.224)
δ_m : $\text{RAC}_{t-1:t-22}$	0.525 (2.920)	0.463 (2.795)
δ_d^Q : $\text{RAC}_{t-1} \sqrt{\text{RiceQ}_{t-1}}$		-1.116 (-2.113)
Log-likelihood	23401.968	23427.394
$\sigma \times 10^4$	51.103	50.893
Adj.R2	0.158	0.165

Note: OLS estimates of the HAR-type models on the full period. Robust (HAC) t-statistics are reported in parenthesis. The last three lines contain respectively the log-likelihood, the standard deviation of the the residuals (multiplied by 10^4) and the adjusted R-squared.

First, the RAC series is persistent. The estimated coefficients of the lag one of the daily, weekly, and monthly average of RAC in Table 2 are all significant and positive at the 5% nominal level in the RAC-HARQ model. Furthermore, RAC responds negatively and significantly to the interaction term. The adjusted R-squared of the RAC-HARQ model is 16.5% which suggests that a large part of RAC remains unpredicted by the model. This is mainly due to the fact that RAC essentially measures noise when the drift is negligible. Persistence in the RAC and its association with price drift may explain the findings of DeMiguel et al. (2014), who show that serial correlation can be used to forecast expected means to improve portfolio allocation.

Second, the realized drift measure RAC is found to be crucial for forecasting RiceV. The RAC_{t-1} term in Table 3 is significant and positive at the 5% nominal level whenever it is included in the model, i.e., for the RiceV-HAR-D, RiceV-HARQ-D

Table 3: In-sample estimation results for RiceV_t

	RiceV-HAR	RiceV-HARQ	RiceV-HAR-D	RiceV-HARQ-D	RiceV-HARQ-DQ
β_0 : Constant	0.001 (3.937)	0.000 (1.107)	0.001 (3.807)	0.000 (1.320)	0.000 (1.159)
β_d : RiceV _{t-1}	0.413 (9.601)	0.633 (8.072)	0.360 (7.660)	0.576 (7.710)	0.526 (7.156)
β_w : RiceV _{t-1:t-5}	0.357 (4.068)	0.310 (3.350)	0.406 (4.566)	0.356 (3.828)	0.358 (3.730)
β_m : RiceV _{t-1:t-22}	0.137 (2.253)	0.080 (1.248)	0.134 (1.952)	0.088 (1.177)	0.101 (1.380)
β_d^Q : RiceV _{t-1} $\sqrt{\text{RiceQ}_{t-1}}$		-0.981 (-3.140)		-0.925 (-3.077)	-0.608 (-1.968)
γ_d : RAC _{t-1}			0.292 (1.783)	0.249 (1.732)	0.726 (3.277)
γ_w : RAC _{t-1:t-5}			-0.211 (-0.773)	-0.190 (-0.803)	-0.218 (-0.958)
γ_m : RAC _{t-1:t-22}			-0.026 (-0.094)	-0.073 (-0.288)	-0.224 (-0.995)
γ_d^Q : RAC _{t-1} $\sqrt{\text{RiceQ}_{t-1}}$					-3.080 (-2.644)
Log-likelihood	18625.158	18701.771	18658.535	18726.148	18755.035
$\sigma \times 10^4$	112.316	110.915	111.727	110.498	109.982
Adj.R2	0.603	0.613	0.607	0.616	0.620

Note: See Table 2.

and RiceV-HARQ-DQ models. The interaction term $\text{RAC}_{t-1} \sqrt{\text{RiceQ}_{t-1}}$ in the RiceV-HARQ-DQ model plays a non-negligible role with a highly significant and negative coefficient. RiceV-HARQ-DQ turns out to be the best specification for modeling RiceV, with an adjusted R-squared of 62%. Furthermore, the restrictions for the RiceV-HAR and RiceV-HARQ models are rejected at any conventional level using either a likelihood ratio test or a Wald test based on a robust estimation the variance covariance matrix (not reported here).

Table 4: In-sample estimation results for RV_t

	RV-HAR	RV-HARQ	RV-HAR-D	RV-HARQ-D	RV-HARQ-DQ
α_0 : Constant	0.001 (4.329)	0.000 (1.882)	0.001 (4.121)	0.000 (1.360)	0.000 (1.514)
α_d : RV _{t-1}	0.409 (8.604)	0.630 (10.249)	0.467 (7.771)	0.600 (10.195)	0.570 (10.191)
α_w : RV _{t-1:t-5}	0.314 (3.599)	0.233 (2.505)	0.362 (3.960)	0.312 (3.199)	0.318 (3.237)
α_m : RV _{t-1:t-22}	0.177 (2.539)	0.129 (2.020)	0.095 (1.214)	0.085 (1.050)	0.086 (1.089)
α_d^Q : RV _{t-1} $\sqrt{\text{RV}_{t-1}}$		-0.567 (-9.632)		-0.690 (-4.986)	-0.414 (-1.892)
λ_d : RAC _{t-1}			-0.218 (-1.061)	0.327 (1.541)	0.541 (2.420)
λ_w : RAC _{t-1:t-5}			-0.428 (-1.192)	-0.398 (-1.345)	-0.416 (-1.408)
λ_m : RAC _{t-1:t-22}			0.478 (0.985)	0.174 (0.445)	0.134 (0.359)
λ_d^Q : RAC _{t-1} $\sqrt{\text{RV}_{t-1}}$					-3.047 (-1.890)
Log-likelihood	17213.253	17314.699	17237.423	17326.936	17335.068
$\sigma \times 10^4$	141.751	139.411	141.222	139.165	138.990
Adj.R2	0.565	0.579	0.568	0.580	0.581

Note: See Table 2.

Similar results are observed in Table 4 for the models fitted to RV. The realized drift measure RAC is found to be significantly positive at the 5% nominal level in the RV-HARQ-DQ model while the interaction term $\text{RAC}_{t-1} \sqrt{\text{RiceQ}_{t-1}}$ is found to be significantly negative (at the 5% nominal level).

Out-of-sample Forecasting

We examine the one-step-ahead out-of-sample forecasting performance of the competing models for RiceV, RAC and RV. All models are estimated on the first 2,000 observations to produce the one-step-ahead forecasts for June 1, 2004. The models are then re-estimated on an expanding window each time a new observation becomes available.

For each model i , the forecasts $\hat{y}_{t+1}^{(i)}$ are compared to the true values y_{t+1} using three different loss functions $L_{i,t}$, i.e.,⁶

$$\begin{aligned} MSFE_{i,t} &= \left(\hat{y}_{t+1}^{(i)} - y_{t+1} \right)^2, \\ MAFE_{i,t} &= \left| \hat{y}_{t+1}^{(i)} - y_{t+1} \right|, \\ QLIKE_{i,t} &= \log y_{t+1} + \frac{\hat{y}_{t+1}^{(i)}}{y_{t+1}}. \end{aligned}$$

The model confidence set (MCS) of Hansen et al. (2011) is employed to rank the models. Let \mathcal{M}_0 be the set of competing models. Their relative performance is measured by $d_{i,j,t} = L_{i,t} - L_{j,t}$ for all $i, j \in \mathcal{M}_0$.

The MCS test is an iterative procedure. For iteration s , it applies a model equivalence test for the null hypothesis of

$$H_{0,\mathcal{M}_s} : E(d_{ij,t}) = 0 \text{ for all } i, j \in \mathcal{M}_s \subset \mathcal{M}_0,$$

against the alternative

$$H_{A,\mathcal{M}_s} : E(d_{ij,t}) \neq 0 \text{ for some } i, j \in \mathcal{M}_s.$$

If H_{0,\mathcal{M}_s} is ‘accepted’ the confidence set $\hat{\mathcal{M}}_{1-\alpha} = \mathcal{M}_s$, otherwise use an elimination rule to remove objects from \mathcal{M}^s and repeat the test. Let $P_{H_{0,\mathcal{M}_s}}$ be the p-value

⁶QLIKE is not considered when forecasting RAC because RAC can take negative values while QLIKE is suited for strictly positive variables only.

associated with the null hypothesis H_{0,\mathcal{M}_s} and $e_{\mathcal{M}_s}$ be the model eliminated from set \mathcal{M}_s when H_{0,\mathcal{M}_s} is rejected. The MCS p-value for model $e_{\mathcal{M}_s}$ is defined by

$$\hat{p}_{e_{\mathcal{M}_s}} = \max_{k \leq s} P_{H_{0,\mathcal{M}_k}},$$

where $\mathcal{M}_1 \supset \mathcal{M}_2 \dots \supset \mathcal{M}_s$. To make sure that our empirical results are not specific to the chosen forecasting period, we apply the MCS test on rolling windows of 500 observations over the full forecasting period. The total number of one-step-ahead forecasts is 4,087 so that the total number of rolling windows of 500 observations is 3,587. For each window, the distribution of the MCS test is obtained from 10,000 bootstrap samples with a block length of 5 observations to account for the possible presence of serial correlation and heteroskedasticity in the loss differences. Table 5 reports the percentages of rolling window MSC p-values (out of 3,587) being above the thresholds of 25%, 10% and 5% for the competing models.

The out-of-sample forecasting results for RAC suggest that RAC–HARQ is a better specification than RAC–HAR, with either MSFE or MAFE. For instance, with the MSFE loss function and a threshold of 10%, the RAC–HARQ (resp. RAC–HAR) belongs to the set of superior models in 97.6% (resp. 66.7%) of the cases.

For the two volatility measures, HARQ performs better than HAR based on QLIKE and MAFE but not with MSFE. Most importantly, our proposed model specifications RiceV-HARQ-DQ and RV-HARQ-DQ are superior to their respective alternative specifications in out-of-sample forecasting regardless of the criterion used. The most striking results are from the MAFE loss function and a threshold of 10% where RiceV-HARQ-DQ and RV-HARQ-DQ belong to the MCS in 100% of the cases while their HARQ versions belong to the MCS in respectively 59.2% and 45.5% of the cases.

The out-of-sample forecasting and in-sample estimation results thus corroborate our theoretical conjecture that explicitly accounting for the drift using the proposed realized autocovariance estimator improves the forecasting quality of both RiceV and RV over the standard HAR and HARQ specifications.

Table 5: Results of the MCS test for the one-step-ahead forecasts of RAC, RiceV, and RV

α	HAR	HARQ	HAR-D	HARQ-D	HARQ-DQ
MSFE for RAC					
25%	35.406	89.796			
10%	66.741	97.658			
5%	75.300	98.411			
MAFE for RAC					
25%	12.434	99.387			
10%	16.839	100.00			
5%	22.080	100.00			
MSFE for RiceV					
25%	96.320	83.607	99.777	85.782	97.212
10%	99.721	86.813	100.00	88.793	99.721
5%	100.00	88.737	100.00	95.344	100.00
MAFE for RiceV					
25%	38.054	36.772	38.472	39.197	100.00
10%	45.219	59.297	52.272	55.227	100.00
5%	49.122	69.083	57.987	60.190	100.00
QLIKE for RiceV					
25%	46.334	72.010	50.739	74.854	88.152
10%	68.414	80.374	70.449	88.124	97.073
5%	74.854	85.336	75.272	95.790	99.610
MSFE for RV					
25%	97.101	91.943	80.485	94.257	98.216
10%	99.638	99.387	87.232	98.996	99.582
5%	99.944	100.00	92.166	100.00	100.00
MAFE for RV					
25%	20.630	44.299	8.5587	78.701	87.538
10%	33.203	45.526	20.296	85.503	100.00
5%	41.149	48.982	26.791	88.096	100.00
QLIKE for RV					
25%	28.938	96.905	30.555	99.861	99.024
10%	37.106	98.996	39.281	100.00	100.00
5%	47.198	99.526	48.313	100.00	100.00

Note: The MCS test is applied on rolling windows of 500 observations. This figures in the table correspond to the percentages of the rolling window MSC p-values being above the thresholds of 25%, 10% and 5% for the competing models over the full forecasting period.

7 Conclusion

We show that realized autocovariance (RAC) reveals realized drift, at first order, when the observation frequency increases, in a model which allows for drift and volatility explosion. When drift is too small, realized autocovariance is instead distributed around zero (with a predictable variance), in line with the efficient

market hypothesis in a frictionless market. The evidence is corroborated by simulations which show that serial covariance can indeed detect drift, and that the presence of additional market frictions does not impair its ability to do so.

The theory also implies a novel decomposition of realized variance into a drift and a volatility component, estimated by realized autocovariance and an estimator based on first differences of high-frequency returns (the Rice estimator) respectively. The decomposition allows for distinct dynamics of the drift and the volatility component.

In our empirical application we show that RAC is largely positive in the data, peaking at times of well known sustained trends and persistent. Finally, we show that past values of RAC help to improve the quality of the in-sample estimation and out-of-sample prediction of the realized variance of the Nasdaq Composite Index over the standard HAR and HARQ specifications.

We thus conclude that realized drift is an important feature of volatility dynamics which has been neglected by the econometric literature so far. We also pose new challenges and questions for future research in financial economics. What is the role of drift in asset pricing? Can drift improve portfolio allocations? Does drift contain new relevant economic information with respect to volatility? These questions are left for future research

References

- Aït-Sahalia, Y. and R. Kimmel (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics* 83(2), 413–452. [22](#)
- Andersen, T. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39(4), 885–905. [2](#)
- Andersen, T., Y. Li, V. Todorov, and B. Zhou (2021). Volatility measurement with pockets of extreme return persistence. *Journal of Econometrics*. [3](#), [4](#), [14](#)
- Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics* 169(1), 75–93. [2](#)
- Arnold, L. (1974). *Stochastic differential equations*. A Wiley-Interscience publication. Wiley. [2](#)
- Bandi, F. M. (2002). Short-term interest rate dynamics: a spatial approach. *Journal of Financial Economics* 65(1), 73–110. [3](#)

- Bandi, F. M. and P. C. Phillips (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71(1), 241–283. [2](#)
- Bandi, F. M. and J. R. Russell (2011). Market microstructure noise, integrated variance estimators, and the accuracy of asymptotic approximations. *Journal of Econometrics* 160(1), 145–159. [22](#)
- Barndorff-Nielsen, O. E., S. Kinnebroek, and N. Shephard (2010). Measuring downside risk: realised semivariance. *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*, (Edited by T. Bollerslev, J. Russell and M. Watson) 117, 136. [2](#)
- Barndorff-Nielsen, O. E. and N. Shephard (2002a). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64(2), 253–280. [24](#)
- Barndorff-Nielsen, O. E. and N. Shephard (2002b). Estimating quadratic variation using realized variance. *Journal of Applied econometrics* 17(5), 457–477. [2](#)
- Barndorff-Nielsen, O. E. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics* 4(1), 1–30. [4](#), [17](#)
- Bekaert, G. and R. J. Hodrick (1992). Characterizing predictable components in excess returns on equity and foreign exchange markets. *The Journal of Finance* 47(2), 467–509. [2](#)
- Bessembinder, H. and K. Chan (1992). Time-varying risk premia and forecastable returns in futures markets. *Journal of Financial Economics* 32(2), 169–193. [2](#)
- Bollerslev, T., J. Li, A. J. Patton, and R. Quaedvlieg (2020). Realized semicovariances. *Econometrica* 88(4), 1515–1551. [2](#)
- Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics* 192(1), 1–18. [5](#), [19](#), [20](#), [27](#)
- Campbell, J. Y. and J. Ammer (1993). What moves the stock and bond markets? a variance decomposition for long-term asset returns. *The Journal of Finance* 48(1), 3–37. [2](#)
- Campbell, J. Y. and Y. Hamao (1992). Predictable stock returns in the united states and japan: A study of long-term capital market integration. *The Journal of Finance* 47(1), 43–69. [2](#)
- Christensen, K., R. C. Oomen, and M. Podolskij (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics* 114(3), 576–599. [23](#)
- Christensen, K., R. C. A. Oomen, and R. Renò (2022). The drift burst hypothesis. *Journal of Econometrics* 227(2), 461–497. [2](#), [3](#), [4](#), [8](#), [22](#)
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics* 7(2), 174–196. [27](#)
- DeMiguel, V., F. J. Nogales, and R. Uppal (2014). Stock return serial dependence and out-of-sample portfolio performance. *The Review of Financial Studies* 27(4), 1031–1073. [30](#)
- Fama, E. F. and K. R. French (1988). Permanent and temporary components of stock prices. *Journal of Political Economy* 96(2), 246–273. [2](#)

- Flora, M. and R. Renò (2021). V-Shapes. Working paper. [16](#)
- Hall, P., J. Kay, and D. Titterton (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* *77*(3), 521–528. [11](#)
- Hans-Georg, M. and U. Stadtmüller (1988). Detecting dependencies in smooth regression models. *Biometrika* *75*(4), 639–650. [11](#)
- Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica* *79*(2), 453–497. [32](#)
- Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications* *119*(7), 2249–2276. [17](#)
- Jacod, J., Y. Li, and X. Zheng (2017). Statistical properties of microstructure noise. *Econometrica* *85*(4), 1133–1174. [17](#)
- Jacod, J. and P. Protter (2011). *Discretization of processes*, Volume 67. Springer Science & Business Media. [6](#), [40](#), [43](#)
- Kalnina, I. and O. Linton (2008). Estimating quadratic variation consistently in the presence of correlated measurement error. *Journal of Econometrics* *147*(1), 47–59. [22](#)
- Kristensen, D. (2010). Nonparametric filtering of the realized spot volatility: A kernel-based approach. *Econometric Theory* *26*(1), 60–93. [3](#)
- Laurent, S. and S. Shi (2020). Volatility estimation and jump detection for drift-diffusion processes. *Journal of Econometrics* *217*(2), 259–290. [2](#), [3](#), [4](#), [8](#), [22](#)
- Laurent, S. and S. Shi (2022). Unit root test with high-frequency data. *Econometric Theory* *38*(1), 113–171. [2](#), [3](#), [8](#)
- Lee, S. and P. A. Mykland (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *The Review of Financial Studies* *21*(6), 2535–2563. [2](#)
- Li, Z. M. and O. Linton (2022). A remedy for microstructure noise. Technical Report 1. [17](#)
- Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* *36*(2), 270–296. [4](#)
- Merton, R. C. (1980). On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics* *8*(4), 323–361. [2](#)
- Michaud, R. O. (1989). The markowitz optimization enigma: Is ‘optimized’ optimal? *Financial analysts journal* *45*(1), 31–42. [3](#)
- Oomen, R. C. A. (2006). Properties of realized variance under alternative sampling schemes. *Journal of Business and Economic Statistics* *24*(2), 219–237. [22](#)
- Phillips, P. C., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? *International Economic Review* *52*(1), 201–226. [2](#)
- Phillips, P. C. and J. Yu (2011). Dating the timeline of financial bubbles during the subprime crisis. *Quantitative Economics* *2*(3), 455–491. [2](#)
- Phillips, P. C. B. and S. Shi (2019). Detecting financial collapse and ballooning sovereign risk. *Oxford Bulletin of Economics and Statistics* *81*(6), 1336–1361.

2, 8

- Phillips, P. C. B., S. Shi, and J. Yu (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review* 56(4), 1043–1078. 2, 8
- Rice, J. (1984, 12). Bandwidth choice for nonparametric regression. *The Annals of Statistics* 12(4), 1215–1230. 3, 11
- Roll, R. (1984). A simple measure of the implicit bid-ask spread in an efficient market. *Journal of Finance* 39(4), 1127–1139. 17
- Shi, S. and Y. Song (2016). Identifying speculative bubbles using an infinite hidden Markov model. *Journal of Financial Econometrics* 14(1), 159–184. 2
- Von Neumann, J., R. Kent, H. Bellinson, and B. Hart (1941). The mean square successive difference. *The Annals of Mathematical Statistics* 12(2), 153–162. 11

A Mathematical annex

In the proofs, we denote by $t_i = i\Delta_n$ for $i = 1, \dots, n$. We also write, for two random variables X and Y , $X \stackrel{p}{\approx} Y$ when $X = Y(1 + o_p(1))$, while for two real series a_n and b_n , we write $a_n \sim b_n$ when $a_n = b_n(1 + o(1))$. The constant C is generic and may change from line to line.

We define

$$\zeta_{(d_1, k_1), \dots, (d_M, k_M)}^{lf} := 2^{d_1 + \dots + d_M} \lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{j=1}^{n-K} f_{(n-j)\Delta_n} \prod_{m=1}^M [(j + k_m)^{1/2} - (j + k_m - 1)^{1/2}]^{d_m}. \quad (35)$$

Proof of Lemma 3.1, Theorem 3.1 and Corollary B.1. By Ito's lemma, we have

$$\begin{aligned} \text{RV} &= \sum_{i=1}^n \left(\int_{t_{i-1}}^{t_i} \mu_s (1-s)^{-\alpha} ds + \int_{t_{i-1}}^{t_i} \sigma_s (1-s)^{-\beta} dW_s \right)^2 \\ &= \int_{t_1}^1 \sigma_s^2 (1-s)^{-2\beta} ds + R_{1,n} + R_{2,n} + R_{3,n}, \end{aligned}$$

where

$$\begin{aligned} R_{1,n} &\equiv \sum_{i=1}^n \left(\int_{t_{i-1}}^{t_i} \mu_u (1-u)^{-\alpha} du \right)^2 \\ R_{2,n} &\equiv 2 \sum_{i=1}^n \left(\int_{t_{i-1}}^{t_i} \mu_s (1-s)^{-\alpha} ds \right) \left(\int_{t_{i-1}}^{t_i} \sigma_s (1-s)^{-\beta} dW_s \right) \\ R_{3,n} &\equiv 2 \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^s \sigma_u (1-u)^{-\beta} dW_u \right) \sigma_s (1-s)^{-\beta} dW_s. \end{aligned}$$

Write $R_{1,n}$, using the stochastic continuity and boundedness properties of μ_s , as

$$\begin{aligned} R_{1,n} &= \sum_{i=1}^n \left(\mu_{t_{i-1}}^2 + O_p(\Delta_n^\Gamma) \right) \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\alpha} ds \right)^2 \\ &= \underbrace{\frac{\Delta_n^{2(1-\alpha)}}{(1-\alpha)^2} \sum_{j=1}^n \mu_{t_{n-j}}^2 (j^{1-\alpha} - (j-1)^{1-\alpha})^2}_{R_{1,a,n}} + \underbrace{O_p(\Delta_n^\Gamma) \Delta_n^{2(1-\alpha)} \sum_{j=1}^n (j^{1-\alpha} - (j-1)^{1-\alpha})^2}_{R_{1,b,n}}. \end{aligned}$$

where we changed index using $j = n - i + 1$. Now consider the sum:

$$S_n = \sum_{j=1}^n (j^{1-\alpha} - (j-1)^{1-\alpha})^2,$$

and notice that this series is convergent when $\alpha > 1/2$ and divergent when $\alpha \leq 1/2$.

- (i) When $\alpha > 1/2$, by the boundedness of μ_t and the convergence of S_n , we have that $\sum_{j=1}^n \mu_{t_{n-j}}^2 (j^{1-\alpha} - (j-1)^{1-\alpha})^2$ is convergent almost surely. Thus, using definition (9),

$$\Delta_n^{2\alpha-2} R_{1,a,n} \xrightarrow{p} \zeta_{\mu^2, \alpha}.$$

- (ii) When $\alpha < 1/2$, we use the following argument. By the mean-value theorem there exists $\xi_j \in]j-1, j[$ (even when $j=1$) such that

$$j^{1-\alpha} - (j-1)^{1-\alpha} = (1-\alpha)\xi_j^{-\alpha}.$$

Thus

$$R_{1,a,n} = \Delta_n^{2(1-\alpha)} \sum_{j=1}^n \mu_{t_{n-j}}^2 \xi_j^{-2\alpha},$$

and by a simple Riemann argument, and changing back to index $i = n - j + 1$, we obtain

$$\Delta_n^{-1} R_{1,a,n} \xrightarrow{a.s.} \int_0^1 \mu_s^2 (1-s)^{-2\alpha} ds \quad (36)$$

since $\int_0^1 (1-s)^{-2\alpha} ds$ is convergent when $\alpha < 1/2$.

- (iii) When $\alpha = 1/2$, neither $\int_0^1 (1-s)^{-1} ds$ nor S_n is convergent. We can, however, find the rate of divergence using properties of the harmonic sum and the mean-value theorem as in point (ii) above. Indeed, using the fact that $(j-1) \leq \xi_j \leq j$, and noticing that $\xi_1 = (1-\alpha)^{1/\alpha}$, we have

$$\begin{aligned} \sum_{j=1}^n \mu_{t_{n-j}}^2 \xi_j^{-1} &\geq C \sum_{j=1}^n \frac{1}{j} \geq C \log(n), \\ \sum_{j=1}^n \mu_{t_{n-j}}^2 \xi_j^{-1} &\leq C \left((1-\alpha)^{1/\alpha} + \sum_{j=2}^n \frac{1}{j-1} \right) \leq C \log(n). \end{aligned}$$

This proves that

$$\lim_{n \rightarrow \infty} \frac{1}{\log n} \sum_{j=1}^n \mu_{t_{n-j}}^2 \xi_j^{-1}$$

exists and is bounded almost surely. Thus, we can conclude that when $\alpha = \frac{1}{2}$, using definition (35),

$$(\Delta_n \log(n))^{-1} R_{1,a,n} \xrightarrow{p} \zeta'_{\mu^2}.$$

Summarizing,

$$\begin{cases} nR_{1,a,n} \xrightarrow{a.s.} \int_0^1 \mu_s^2 (1-s)^{-2\alpha} ds & \text{if } \alpha < 1/2 \\ n^{2(1-\alpha)} R_{1,a,n} \xrightarrow{p} \zeta_{\mu^2, \alpha} & \text{if } \alpha > 1/2 \\ \frac{n}{\log(n)} R_{1,a,n} \xrightarrow{p} \zeta'_{\mu^2} & \text{if } \alpha = 1/2 \end{cases}.$$

For $R_{2,n}$, we have, using similar reasoning,

$$R_{2,n} \stackrel{p}{\approx} 2 \sum_{i=1}^n \mu_{t_{i-1}} \sigma_{t_{i-1}} \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\alpha} ds \right) \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\beta} dW_s \right) := \sum_{i=1}^n v_{2,i},$$

and now we apply Theorem 2.2.14 in [Jacod and Protter \(2011\)](#). The crucial orders are:

$$\sum_{i=1}^n E_{i-1} [v_{2,i}] = 0$$

and

$$\begin{aligned} \sum_{i=1}^n E_{i-1} [v_{2,i}^2] &= 4 \sum_{i=1}^n \mu_{t_{i-1}}^2 \sigma_{t_{i-1}}^2 \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\alpha} ds \right)^2 E_{i-1} \left[\left(\int_{t_{i-1}}^{t_i} (1-s)^{-\beta} dW_s \right)^2 \right] \\ &= 4 \sum_{i=1}^n \mu_{t_{i-1}}^2 \sigma_{t_{i-1}}^2 \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\alpha} ds \right)^2 \int_{t_{i-1}}^{t_i} (1-s)^{-2\beta} ds \\ &= \frac{4\Delta_n^{3-2\alpha-2\beta}}{(1-\alpha)^2 (1-2\beta)} \sum_{j=1}^n \mu_{t_{n-j}}^2 \sigma_{t_{n-j}}^2 [j^{1-\alpha} - (j-1)^{1-\alpha}]^2 [j^{1-2\beta} - (j-1)^{1-2\beta}]. \end{aligned}$$

Thus, we have the following results:

(i) When $\alpha + \beta < 1/2$,

$$\Delta_n^{-2} \sum_{i=1}^n E_{i-1} [v_{2,i}^2] \xrightarrow{a.s.} 4 \int_0^1 \mu_s^2 \sigma_s^2 (1-s)^{-2\alpha-2\beta} ds;$$

(ii) When $\alpha + \beta > 1/2$, we have

$$\Delta_n^{2\alpha+2\beta-3} \sum_{i=1}^n E_{i-1} [v_{2,i}^2] \xrightarrow{p} 4\zeta_{(\alpha,2,0),(2\beta,1,0)}^{\mu^2 \sigma^2};$$

(iii) When $\alpha + \beta = 1/2$,

$$\Delta_n^{-2} \log(n)^{-1} \sum_{i=1}^n E_{i-1} [v_{2,i}^2] \xrightarrow{p} 4\zeta_{(2,0)}^{\mu^2\sigma^2}.$$

This implies that

$$\begin{cases} nR_{2,n} \xrightarrow{d} N\left(0, 4 \int_0^1 \mu_s^2 \sigma_s^2 (1-s)^{-2\alpha-2\beta} ds\right) & \text{if } \alpha + \beta < 1/2 \\ n^{\frac{3}{2}-(\alpha+\beta)} R_{2,n} \xrightarrow{d} N\left(0, 4\zeta_{(\alpha,2,0),(2\beta,1,0)}^{\mu^2\sigma^2}\right) & \text{if } \alpha + \beta > 1/2 \\ \frac{n}{(\log n)^{1/2}} R_{2,n} \xrightarrow{d} N\left(0, 4\zeta_{(2,0)}^{\mu^2\sigma^2}\right) & \text{if } \alpha + \beta = 1/2. \end{cases}$$

For $R_{3,n}$, we have

$$R_{3,n} \stackrel{p}{\approx} 2 \sum_{i=1}^n \sigma_{t_{i-1}}^2 \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^s (1-u)^{-\beta} dW_u \right) (1-s)^{-\beta} dW_s := \sum_{i=1}^n v_{3,i}.$$

We again have $\sum_{i=1}^n E_{i-1} [v_{3,i}] = 0$ and the crucial order is

$$\begin{aligned} \sum_{i=1}^n E_{i-1} [v_{3,i}^2] &= 4 \sum_{i=1}^n \sigma_{t_{i-1}}^4 \int_{t_{i-1}}^{t_i} E_{i-1} \left(\int_{t_{i-1}}^s (1-u)^{-\beta} dW_u \right)^2 (1-s)^{-2\beta} ds \\ &= 4 \sum_{i=1}^n \sigma_{t_{i-1}}^4 \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^s (1-u)^{-2\beta} ds \right) (1-s)^{-2\beta} ds \\ &= 2 \sum_{i=1}^n \sigma_{t_{i-1}}^4 \left[\int_{t_{i-1}}^{t_i} (1-s)^{-2\beta} ds \right]^2 \\ &= \frac{2\Delta_n^{2-4\beta}}{(1-2\beta)^2} \sum_{j=1}^n \sigma_{t_{n-j}}^4 \left[j^{1-2\beta} - (j-1)^{1-2\beta} \right]^2 \\ &\stackrel{p}{\approx} 2\Delta_n^{2-4\beta} \sum_{j=1}^{n-1} \sigma_{t_{n-j}}^4 \xi_j^{-4\beta}, \end{aligned}$$

and reasoning exactly as before we have three cases and we can prove that:

$$\begin{cases} n^{1/2} R_{3,n} \xrightarrow{d} N\left(0, 2 \int_0^1 \sigma_s^4 (1-s)^{-4\beta} ds\right) & \text{if } \beta < 1/4 \\ n^{1-2\beta} R_{3,n} \xrightarrow{d} N\left(0, 2\zeta_{(2\beta,2,0)}^{\sigma^4}\right) & \text{if } 1/4 < \beta < 1/2 \\ \left(\frac{n}{\log(n)}\right)^{1/2} R_{3,n} \xrightarrow{d} N\left(0, 2\zeta_{(2,0)}^{\sigma^4}\right) & \text{if } \beta = 1/4. \end{cases}$$

One can see that when $0 \leq \alpha < 1$ and $0 \leq \beta < 1/2$, the three quantities converge to zero at various rates as $n \rightarrow \infty$. Therefore,

$$RV - \int_0^1 \sigma_s^2 (1-s)^{-2\beta} ds = R_{1,n} + R_{2,n} + R_{3,n} \xrightarrow{p} 0.$$

The limiting distribution of $\text{RV} - \int_0^1 \sigma_s^2 (1-s)^{-2\beta} ds$ in various cases comes from the specific analysis of different orders. For example, in case (4) $R_{1,n}$ is the dominating term, followed by $R_{2,n}$. \square

Proof of Theorem 3.2 and Corollary B.2. It is convenient to write:

$$\text{RiceV} = \text{RV} - \sum_{i=k+1}^n \Delta_i^n X \Delta_{i-k}^n X + o_p(1),$$

where the $o_p(1)$ is an end effect. Decomposing RiceV as in Theorem 3.1, we can now write:

$$\text{RiceV} = \tilde{R}_{1,n} + \tilde{R}_{2,n} + \tilde{R}_{3,n},$$

where

$$\tilde{R}_{1,n} := R_{1,n} - \sum_{i=k+1}^n \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \mu_s (1-s)^{-\alpha} ds \right) \left(\int_{(i-k-1)\Delta_n}^{(i-k)\Delta_n} \mu_s (1-s)^{-\alpha} ds \right)$$

$$\begin{aligned} \tilde{R}_{2,n} := & R_{2,n} - \sum_{i=k+1}^n \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \mu_s (1-s)^{-\alpha} ds \right) \left(\int_{(i-k-1)\Delta_n}^{(i-k)\Delta_n} \sigma_s (1-s)^{-\beta} dW_s \right) \\ & - \sum_{i=k+1}^n \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s (1-s)^{-\beta} dW_s \right) \left(\int_{(i-k-1)\Delta_n}^{(i-k)\Delta_n} \mu_s (1-s)^{-\alpha} ds \right) \end{aligned}$$

$$\tilde{R}_{3,n} := R_{3,n} - \sum_{i=k+1}^n \left(\int_{(i-1)\Delta_n}^{i\Delta_n} \sigma_s (1-s)^{-\beta} dW_s \right) \left(\int_{(i-k-1)\Delta_n}^{(i-k)\Delta_n} \sigma_s (1-s)^{-\beta} dW_s \right).$$

For the first term, we have, using again the stochastic continuity and $j = n - i + 1$,

$$\begin{aligned} \tilde{R}_{1,n} & \stackrel{p}{\approx} \frac{\Delta_n^{2(1-\alpha)}}{(1-\alpha)^2} \sum_{j=1}^n \mu_{t_{n-j}}^2 (j^{1-\alpha} - (j-1)^{1-\alpha})^2 \\ & - \frac{\Delta_n^{2(1-\alpha)}}{(1-\alpha)^2} \sum_{j=1}^{n-k} \mu_{t_{n-j}}^2 (j^{1-\alpha} - (j-1)^{1-\alpha}) ((j+k)^{1-\alpha} - (j+k-1)^{1-\alpha}). \end{aligned}$$

Now, notice that the the series

$$S'_n = \sum_{j=1}^{n-k} (j^{1-\alpha} - (j-1)^{1-\alpha}) [j^{1-\alpha} - (j-1)^{1-\alpha} - (j+k)^{1-\alpha} + (j+k-1)^{1-\alpha}]$$

is almost surely convergent for all $\alpha > 0$. Indeed, by the mean-value theorem applied twice

$$S'_n = (1-\alpha)^2 \sum_{j=1}^{n-k} \xi_j^{-\alpha} (\xi_j^{-\alpha} - \xi_{j+k}^{-\alpha}) = k\alpha(1-\alpha)^2 \sum_{j=1}^{n-k} \xi_j^{-\alpha} (\xi'_j)^{-\alpha-1}$$

$$\leq k\alpha(1-\alpha)^2 \sum_{j=1}^{n-k} j^{-2\alpha-1}$$

for suitable numbers $\xi'_j \in]\xi_j, \xi_{j+k}[$. Thus, S'_n is always convergent when $\alpha > 0$. This proves that, when $\alpha > 0$,

$$\Delta_n^{2\alpha-2} \tilde{R}_{1,n} \xrightarrow{p} \zeta_{(\alpha,2,0)}^{\mu^2} - \zeta_{(\alpha,1,0),(\alpha,1,k)}^{\mu^2}$$

while when $\alpha = 0$ we simply have:

$$\tilde{R}_{1,n} \stackrel{p}{\sim} \frac{\Delta_n^2}{(1-\alpha)^2} \left[\sum_{j=1}^n \mu_{t_{n-j}}^2 - \sum_{j=1}^{n-k} \mu_{t_{n-j}}^2 \right] = \frac{\Delta_n^2}{(1-\alpha)^2} \sum_{j=n-k+1}^n \mu_{t_{n-j}}^2 = O_p(\Delta_n^2).$$

For the third term, we write: $\tilde{R}_{3,n} = \sum_{i=k+1}^n \tilde{v}_{3,i} + o_p(1)$, where the $o_p(1)$ term is taking care of the end effect and the continuity of the coefficients (since $k\Delta_n \rightarrow 0$ for every fixed k), and

$$\begin{aligned} \tilde{v}_{3,i} = & 2\sigma_{t_{i-1}}^2 \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^s (1-u)^{-\beta} dW_u \right) (1-s)^{-\beta} dW_s \\ & - \sigma_{t_{i-1}}^2 \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\beta} dW_s \right) \left(\int_{t_{i-k-1}}^{t_{i-k}} (1-s)^{-\beta} dW_s \right). \end{aligned}$$

We now we apply Theorem 2.2.14 in [Jacod and Protter \(2011\)](#). First,

$$\sum_{i=k+1}^n E_{i-1} [\tilde{v}_{3,i}] = 0,$$

and,

$$\begin{aligned} \sum_{i=k+1}^n E_{i-1} [\tilde{v}_{3,i}^2] = & 4\sigma_{t_{i-1}}^4 \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^s (1-u)^{-2\beta} du \right) (1-s)^{-2\beta} ds \\ & + \sigma_{t_{i-1}}^4 \left(\int_{t_{i-1}}^{t_i} (1-s)^{-2\beta} ds \right) \left(\int_{t_{i-k-1}}^{t_{i-k}} (1-s)^{-\beta} dW_s \right)^2. \end{aligned}$$

Now, invoking again a suitable law of large numbers and changing index to $j = n - i + 1$, we have:

$$\begin{aligned} \sum_{i=k+1}^n E_{i-1} [\tilde{v}_{3,i}^2] & \stackrel{p}{\sim} 2 \frac{\Delta_n^{2-4\beta}}{(1-2\beta)^2} \sum_{j=1}^{n-k} \sigma_{t_{n-j}}^4 \left[j^{1-2\beta} - (j-1)^{1-2\beta} \right]^2 \\ & + \frac{\Delta_n^{2-4\beta}}{(1-2\beta)^2} \sum_{j=1}^{n-k} \sigma_{t_{n-j}}^4 \left[(j+k)^{1-2\beta} - (j+k-1)^{1-2\beta} \right] \left[j^{1-2\beta} - (j-1)^{1-2\beta} \right] \\ & \stackrel{p}{\sim} 3\Delta_n^{2-4\beta} \sum_{j=1}^{n-k} \sigma_{t_{n-j}}^4 j^{-4\beta} \end{aligned}$$

which allows us to conclude, when $\beta > 1/4$,

$$\Delta_n^{2\beta-1} \tilde{R}_{3,n} \xrightarrow{d} N\left(0, 2\zeta_{(2\beta,2,0)}^{\sigma^4} + \zeta_{(2\beta,1,k),(2\beta,1,0)}^{\sigma^4}\right).$$

When instead $\beta < 1/4$, using $(1-t_i-k\Delta_n)^{-4\beta} \sim (1-t_i)^{-4\beta} + 4\beta k\Delta_n(1-t_i)^{-4\beta-1}$ we have:

$$\Delta_n^{-1/2} \tilde{R}_{3,n} \xrightarrow{d} N\left(0, 3 \int_0^1 \sigma_s^4 (1-s)^{-4\beta} ds\right).$$

Finally, when $\beta = 1/4$, reasoning as above,

$$\left(\frac{n}{\log(n)}\right)^{1/2} \tilde{R}_{3,n} \xrightarrow{d} N\left(0, 2\zeta_{(2,0)}^{\sigma^4} + \zeta_{(1,k),(1,0)}^{\sigma^4}\right).$$

For the second term, we write $\tilde{R}_{2,n} = \sum_{i=k+1}^n \tilde{v}'_{2,i} + o_p(1)$, where the $o_p(1)$ term is again taking care of the end effect and the continuity of the coefficients, and

$$\begin{aligned} \tilde{v}'_{2,i} &= 2\mu_{t_{i-1}}\sigma_{t_{i-1}} \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\alpha} ds \right) \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\beta} dW_s \right) \\ &\quad - \mu_{t_{i-1}}\sigma_{t_{i-1}} \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\beta} dW_s \right) \left(\int_{t_{i-k-1}}^{t_{i-k}} (1-s)^{-\alpha} ds \right) \\ &\quad - \mu_{t_{i-1}}\sigma_{t_{i-1}} \left(\int_{t_{i-k-1}}^{t_{i-k}} (1-s)^{-\beta} dW_s \right) \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\alpha} ds \right). \end{aligned}$$

We now rearrange the sum and write: $\tilde{R}_{2,n} = \sum_{i=k+1}^{n-k} \tilde{v}_{2,i} + o_p(1)$, where

$$\tilde{v}_{2,i} = \mu_{t_{i-1}}\sigma_{t_{i-1}} \left(\int_{t_{i-1}}^{t_i} (1-s)^{-\beta} dW_s \right) \left(2 \int_{t_{i-1}}^{t_i} (1-s)^{-\alpha} ds - \int_{t_{i-k-1}}^{t_{i-k}} (1-s)^{-\alpha} ds - \int_{t_{i+k-1}}^{t_{i+k}} (1-s)^{-\alpha} ds \right),$$

so that we have $\sum_{i=k+1}^{n-k} E_{i-1} [\tilde{v}_{2,i}] = 0$. The crucial order is now, using again $j = n - i + 1$,

$$\begin{aligned} \sum_{i=k+1}^{n-k} E_{i-1} [\tilde{v}_{2,i}^2] &= \sum_{i=k+1}^{n-k} \mu_{t_{i-1}}^2 \sigma_{t_{i-1}}^2 \left(\int_{t_{i-1}}^{t_i} (1-s)^{-2\beta} ds \right) \\ &\quad \times \left(2 \int_{t_{i-1}}^{t_i} (1-s)^{-\alpha} ds - \int_{t_{i-k-1}}^{t_{i-k}} (1-s)^{-\alpha} ds - \int_{t_{i+k-1}}^{t_{i+k}} (1-s)^{-\alpha} ds \right)^2 \\ &= \frac{\Delta_n^{3-2\alpha-2\beta}}{(1-\alpha)^2 (1-2\beta)} \sum_{j=k+1}^{n-k} \mu_{t_{n-j}}^2 \sigma_{t_{n-j}}^2 (j^{1-2\beta} - (j-1)^{1-2\beta}) \\ &\quad \times \left(2(j^{1-\alpha} - (j-1)^{1-\alpha}) - ((j+k)^{1-\alpha} - (j+k-1)^{1-\alpha}) - ((j-k)^{1-\alpha} - (j-k-1)^{1-\alpha}) \right)^2 \end{aligned}$$

which allows us to conclude, when $\alpha + \beta > 1/2$ (which is the only case in which

$\tilde{R}_{2,n}$ dominates $\tilde{R}_{3,n}$),

$$\Delta_n^{\alpha+\beta-\frac{3}{2}} R_{2,n} \xrightarrow{d} N(0, V_4^{\text{RiceV}}(\alpha, \beta)),$$

where $V_4^{\text{RiceV}}(\alpha, \beta)$ is given by Eq. (11). □

Proof of Theorem 3.3 and Corollary B.3. The proof follows among the same line of the proof of Theorems 3.1 and 3.2. □

Assumption A.1. *The weight function $g(\cdot)$ has the following properties: (1) $g : [0, 1] \rightarrow \mathbb{R}$ continuous and piecewise continuously differentiable with a piecewise Lipschitz derivative g' ; (2) $g(0) = g(1) = 0$; (3) $\int_0^1 g(s)^2 ds < \infty$.*