When Should Time be Continuous?
Volatility modeling and estimation of high-frequency data

Yue Fang

Lundquist College of Business
University of Oregon
Eugene, OR 97403

Phone: (541) 346-3265
Email: yfang@darkwing.uoregon.edu

Abstract

The paper studies the problem of volatility modeling and estimation of high-frequency data under continuous record asymptotics. The approach decomposes the observed data into price diffusion and stationary components. The diffusion component may be identified as the "true" value of the underlying asset. The stationary component, termed as the high-frequency "noise" (HFN), accommodates pertinent market microstructure features. A simple condition, characterizing the HFN component on which conventional volatility estimators on the basis of noisy observations will be consistent for diffusion volatility, is derived, and is applied to Reuters FXFX data. It is shown that conventional volatility estimators lead to substantial spurious volatility in high-frequency returns. The failure of conventional estimators in providing consistent estimates is due to the higher irregularities of the HFN sample path, which is induced, at least in part, by trader heterogeneity. In addition, the optimal sampling frequency is acquired which justifies the appropriateness of the use of the 10- to 15-minute sampling intervals - the benchmark noise filter used in many recent empirical studies dealing with high-frequency foreign exchange data.

When Should Time be Continuous?
Volatility modeling and estimation of high-frequency data

Abstract

The paper studies the problem of volatility modeling and estimation of high-frequency data under continuous record asymptotics. The approach decomposes the observed data into price diffusion and stationary components. The diffusion component may be identified as the "true" value of the underlying asset. The stationary component, termed as the high-frequency "noise" (HFN), accommodates pertinent market microstructure features. A simple condition, characterizing the HFN component on which conventional volatility estimators on the basis of noisy observations will be consistent for diffusion volatility, is derived, and is applied to Reuters FXFX data. It is shown that conventional volatility estimators lead to substantial spurious volatility in high-frequency returns. The failure of conventional estimators in providing consistent estimates is due to the higher irregularities of the HFN sample path, which is induced, at least in part, by trader heterogeneity. In addition, the optimal sampling frequency is acquired which justifies the appropriateness of the use of the 10- to 15-minute sampling intervals - the benchmark noise filter used in many recent empirical studies dealing with high-frequency foreign exchange data.

Recent availability of high-frequency data offers empirical researchers the opportunity of addressing problems that are difficult to resolve with daily data or data collected at lower frequencies. On the other hand, modeling and analyzing near-continuous records of data pose new challenges. The extent to which either the empirical findings or the theoretical concepts developed in daily or weekly data are applicable in the high-frequency domain is not apparent and needs to be explored. For example, several recent studies have raised questions about the usefulness of standard volatility models in high-frequency data analysis. The autoregressive conditional heteroskedastic (ARCH) model and its various extensions have enjoyed considerable success in modeling dynamics of volatility for daily and weekly data across different asset classes and institutional settings (Bollerslev et al. (1992, 1994)). It has been well-known that under some regularity conditions, if one can observe a continuous record of data the volatility of the underlying asset can be modeled by the ARCH family and estimated consistently even if mis-specified volatility models are used. This robustness phenomenon of volatility modeling has important implications both in theory and in practice. The theoretical implication is that there is only one true variance of any continuously observable process and all investors should agree on it without debate. Nelson (1992) provided formal analysis of the robustness of volatility estimates to certain types of model mis-specification and he termed this phenomenon of the uniqueness of volatility as "getting the right variance with the wrong model." From a practical point of view, it implies that the volatility could be estimated at any desired accuracy as long as one collects data at finer intervals. Unfortunately, such continuous record asymptotics will ultimately break down. In practice, sampling cannot be done continuously. The more problematic feature of this asymptotic notion is that some insidious side effects, so called high-frequency noise (HFN), such as those related to market microstructures, may arise. As a consequence, dynamic properties of the observed process may be distorted and often lead to puzzling results. In fact, as was pointed out by Merton (1980) and will be shown in this paper, the benefit of a shorter sampling interval will be swamped long before the continuous time limit is reached. In this respect, the continuous record asymptotics of volatility estimation may never be realized in practice.

Volatility plays a central role in many areas such as portfolio management, risk management and option pricing. A better understanding on how the micro-behavior of volatility differs with respect to different sampling rules would certainly be an important step in the use of high-frequency data in modeling and analyzing volatility dynamics. In this paper, we study the problem of high-frequency volatility modeling and estimation when prices follow a diffusion process which is contaminated by a random noise. Rather than seeking to model specified types of HFN, we presume that we are observing a discrete-time sample of a stochastic process which is a summation of the underlying "true" price diffusion (the signal) and stationary transitory noise, which is a combination of several sources related to market microstructure effects.[1] It is of intrinsic interest to study the behavior of conventional volatility estimators (commonly used in analyzing daily data or data with lower frequencies) under the continuous record asymptotics. We show that if the HFN component is "smooth enough" conventional volatility estimates can still provide a consistent estimate of volatility of the diffusion. However, conventional volatility estimates cannot, in general, serve as consistent volatility estimators. Their failure is due to the high irregularities of the sample path of the HFN component.

The multiple components interpretation of market volatility as resulting from the proposed structural model applies equally well across most financial markets and instruments. However,

[1]There is voluminous literature on market microstructure and its impact on volatility modeling and estimation. However, empirical findings have led to a certain fragmentation of the literature, with most studies focusing predominately on one of various specified types of HFN: nonsynchronous trading (Lo and MacKinlay (1990)), the bid-ask spread (Roll (1984), Glosten (1987), Kaul and Nimalendran (1990)), price discreteness (Hausman, Lo and MacKinlay (1992)), et al. This is ultimately not satisfactory since interactions across these dimensions may produce illusory and misleading results.

for concreteness the empirical analysis in this paper is focused on the foreign exchange market, and one-year tick-by-tick Deutschemark-U.S. Dollar (DM/$) Reuters FXFX data. Our empirical findings clearly point towards the existence of the HFN component in the Reuters FXFX series. By analyzing data sampled at different frequencies, we are able to quantify the smoothness of the HFN sample path and to provide evidence on the source of spurious volatility in high-frequency returns. Our results reveal a clear picture on how "excessive volatility" is generated as one samples at finer intervals; they also demonstrate the importance of the trade-off between increasing of sample size and minimizing HFN biases.

We show that high HFN sample path irregularities are at least partially induced by trader heterogeneity. The hypothesis of trader heterogeneity goes beyond the classical view of rational agents by assuming that tick-by-tick price movement reflects traders' specific characteristics which are determined by factors such as inventories (Lyons (1995) and Froot at al. (1992)), information asymmetry (Morris (1994)), and time horizon (M*ü*ller et al. (1997)).[2] As will become clear, heterogeneous behaviors of market participants govern the short-run price dynamics. It generates high sample path irregularities of price series at the micro level and causes "excessive volatility" in high-frequency returns. The results highlight the importance of the microstructure perspective in understanding volatility dynamics in the foreign exchange market.

The remainder of the paper is organized as follows. Section 1 introduces the volatility model and the estimation problem. Section 2 discusses sample path smoothness and explores sources of high irregularities of the sample path of the HFN component. Estimating volatility is examined in Section 3. An application involving high-frequency Reuters FXFX data is undertaken in Section 4. Section 5 is devoted to analysis of the optimal sampling frequency. Section 6 concludes.

1 High-frequency Data Modeling

1.1 The Model

In this section, we present a simple model for high-frequency data that provides a framework for studying the potential impact of the HFN on volatility estimation. Let X_t, the asset price at time

[2]Traders may even rely on information that bears no relation at all to fundamentals and trade on perception; see, for example, a new survey on United States foreign exchange traders by Cheung and Chinn (1999) and the paper "Making book on the buck," *Wall Street Journal*, September 23, 1988.

t, satisfy the following diffusion process described by Itô stochastic differential equation[3]

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad 0 \le t \le T, \tag{1}$$

where $\sigma : \mathcal{R} \to \mathcal{R}^+$ and W_t is a standard Wiener process. To ensure the existence and uniqueness of a solution to (1), we require the usual Lipschitz and growth conditions:

A1. There exists some constant $K > 0$ such that the function μ and σ satisfy the following conditions for all x and y in $S \subset \mathcal{R}$:

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \le K|x - y|,$$

and

$$\mu^2(x) + \sigma^2(x) \le k^2(1 + x^2).$$

For the convenience, we further assume

A2. $\sigma(.)$ is bounded over the state space of X_t.[4]

Equation (1) is a representation of weak convergence results for sequences of stochastic difference equations such as ARCH models under the continuous-time limit (Nelson (1990)). The model in (1) will serve as a benchmark model for the true return volatility. If prices are characterized by the model in (1), the assumption A1 assures that their volatilities could in principle be estimated without error for any given finite interval $[0, T]$, given that X_t can be observed continuously throughout $[0, T]$.[5]

Now consider an alternative model that allows observed prices to contain errors due to HFN. In particular, the following specification

$$Y_t = X_t + \mathcal{E}_t, \quad 0 \le t \le T. \tag{2}$$

X_t and $\mathcal{E}_t$ are called signal and (HFN) noise, respectively. The noise process satisfies the following assumptions:

A3. $\mathcal{E}_t$ is a stationary process with mean zero and a covariance function $\Gamma(s,t) = \gamma(s-t)$. We make three additional assumptions on the smoothness of γ: as $h \downarrow 0$,

$$E(\mathcal{E}_{t+h} - \mathcal{E}_t)^2 = O(h^l),$$

[3]There is a substantial body of evidence that documents strong intra-daily seasonality in volatility. Much of this intra-daily volatility pattern arises from the time-of-day phenomena which mainly depends upon the daily business cycle. Seasonal effects are incorporated into the model by making the drift $\mu(.)$ in (1) time dependent. Note also that the drift term is less important in the study of volatility using high-frequency data. We discuss this further in Section 2.2.

[4]$\sigma(.)$ is the conditional standard derivation per unit time of the differenced X_t. A2 is equivalent to the assumption that the disturbance term of X_t is of type I (Chapter 3, Merton (1992)).

[5]See, for example, Karatzas and Shreve (1988).

$$E(\mathcal{E}_{t+h} - \mathcal{E}_t)^4 = O(h^{2l}),$$

and

$$E[(\mathcal{E}_h - \mathcal{E}_0)^2(\mathcal{E}_{t+h} - \mathcal{E}_t)^2] = O([t/h]^{-\theta} h^{2l}),$$

where $l \geq 0$ and $\theta > 0$.[6]

Finally, we assume

A4. X_t and $\mathcal{E}_t$ are independent.[7]

The model (2) is a structural model. It is different from the ARMA-GARCH process used in Baillie and Bollerslev (1990), Bollerslev and Domowitz (1993) and Bollerslev and Melvin (1994). As an alternative model, the ARMA-GARCH tracks the temporal dependencies in the conditional mean and variance of observed high-frequency returns but has no capability to model high HFN sample path irregularities. As pointed out by Harvey et al. (1992), Baillie and Bollerslev's model may be viewed as the *reduced form* of models such as (2). In contrast to the ARMA-GARCH process, the model (2) is set up in terms of components which have a direct interpretation.[8]

1.2 The Estimation Problem

To facilitate the study we focus on the quadratic variation and then extend the discussion to the nonparametric kernel estimator considered in Florens-Zmirou (1993). Our choice of the quadratic variation is determined by several considerations. First, since economic theory typically does not provide much information about the functional form of the variance, it is appropriate to consider estimators which require few assumptions on the data generating process. Asymptotic properties of the quadratic variation assume only knowledge on the degree of smoothness of the diffusion function in contrast to parametric cases in which the diffusion function is known up to the value of some finite-dimensional parameters, such as models used by Wiggins (1987), Chesney and Scott (1989), Scott (1989), and Melino and Turnbull (1990). Second, the primary interest in this study is the estimation of the accumulated volatility on a given time interval, the quantity playing a

[6] $\mathcal{E}_t$ need not be a Gaussian process. If the Gaussian assumption is added, the fourth-moment and mixing conditions can be dropped.

[7] The independence assumption is not essential to our results. It can be relaxed by having assumptions on the order of $(X_{t+h} - X_t)(\mathcal{E}_{t+h} - \mathcal{E}_t)(X_{s+h} - X_s)(\mathcal{E}_{s+h} - \mathcal{E}_s)$ as $h \downarrow 0$.

[8] Multiple components ARCH models have been proposed by Müller, et al. (1997) and Andersen and Bollerslev (1997). Müller, et al. focused on the time horizon and the temporal resolution with which different traders are viewing and influencing the market. Andersen and Bollerslev used a frequency domain approach to analyze the long- and short-term volatility components. Our approach is based on the time domain, which is similar to that in Zhou (1996) where a simple structural model of the random walk model plus a white noise process has been examined.

crucial role in option pricing models. The quadratic variation is an ideal vehicle for this purpose.[9]

Suppose that the process Y_t is sampled on an equally spaced partition of $[0,T]$, $\mathcal{T} \equiv \{0, h, 2h, ..., nh = T\}$. To simplify notation, denote Y_i as Y_{ih} for $i = 0, 1, 2, ..., n$. The quadratic variation based on $\{Y_i\}$ is obtained by

$$\widehat{\sigma}_n^2 = \sum_{i=1}^{n} (\Delta_h Y_i)^2, \tag{3}$$

where $\Delta_h Y_i = Y_i - Y_{i-1}$.

In general, the diffusion term σ in (1) is stochastic. If $\sigma^2(X_\tau) \neq T^{-1} \int_0^T \sigma^2(X_t)dt$ alomost surely for $\tau \in [0,T]$, the quadratic variation, $\widehat{\sigma}_n^2$, is not consistent for $\sigma^2(X_\tau)$. The estimation problem based on $\widehat{\sigma}_n^2$ is naturally in the form of its accumulation on $[0,T]$, that is, $\int_0^T \sigma^2(X_t)dt$.[10] The estimation problem of $\sigma^2(X_\tau)$ will be considered when we discuss Florens-Zmirou's nonparametric estimator in Section 3.4.[11]

The consistency of an estimator $\tilde{\sigma}_n^2$ for $\int_0^T \sigma^2(X_t)dt$ is understood in the sense of convergence in probability under the limit that h approaches zero as n increases without bound so as to keep $T \equiv hn$ fixed. More precisely, $\tilde{\sigma}_n^2$ is said to be a consistent estimator for signal volatility if for fixed T, as $h \downarrow 0$,

$$\tilde{\sigma}_n^2 - \int_0^T \sigma^2(X_t)dt = o(1). \tag{4}$$

When $h \downarrow 0$, we obtain in the limit a continuous record of observations over a finite time span, comparable to a conventional ($T \uparrow \infty$) data recording. This type of asymptotics, *the continuous record asmptotics*, is also different from those in other studies. For example, Bergstrom (1976), Nelson (1992), and Foster and Nelson (1994) considered cases when $T \uparrow \infty$ and, at the same time, $h \downarrow 0$ at an appropriate rate.

We view the continuous record asymptotics as a proper framework for the problem of volatility modeling and estimation using high-frequency data. However, the notion of the continuous record asymptotics may not be appropriate for studying the problem of estimating returns. This can be seen clearly in the case that if X_t follows the geometric Brownian motion, the accuracy of the maximum likelihood (ML) estimator for returns depends only upon the total length of the observation period, regardless of what the sampling frequency is (Merton (1980)). Although the conventional ($T \uparrow \infty$) large sample is essential for the estimation consistency of returns,

[9]The quadratic variation has been used in several recent studies. For example, it has been examined by Andersen and Bollerslev (1988b) in a study of volatility forecasting based on high-frequency intradaily data.

[10]The knowledge of the quantity $\int_0^T \sigma^2(X_t)dt$ is essential in pricing of various options with stochastic volatility; see, for example, Hull and White (1987) and Stein and Stein (1991) for more details.

[11]To simplify terminology, we shall refer to $\sigma^2(X_\tau)$ and $\int_0^T \sigma^2(X_t)dt$ volatility, although it should be clear on which of $\sigma^2(X_\tau)$ or $\int_0^T \sigma^2(X_t)dt$ we refer to from context.

the asymptotic properties of return estimators do depend on whether data are available on the entire sample path for general diffusion processes.[12] Nevertheless, we use the continuous record asymptotics for the future argument since the drift term μ in (1), which is in higher order, is less important in the study of volatility using high-frequency data.

Without loss of generality, we assume that $T \equiv 1$ in the rest of discussion. Unless otherwise indicated, the convergence is in probability as $h \downarrow 0$ and all summation, $\sum$, are from 1 to n.

2 Sample Path Regularities

Our approach relies on the analysis of the noise to signal ratio (which is defined below). It leads to rather simple characterizations of necessary and sufficient conditions for the consistency of the quadratic variation estimator in terms of the regularities of sample paths.

Let $\Delta_h X_t$ and $\Delta_h \mathcal{E}_t$ be increments in the processes accrued over a time interval of length h. Thus, $\Delta_h X_t = X_{t+h} - X_t$ and $\Delta_h \mathcal{E}_t = \mathcal{E}_{t+h} - \mathcal{E}_t$. The noise to signal ratio, ρ, defined by

$$\rho(h) \equiv \frac{E(\Delta_h \mathcal{E}_t)^2}{E(\Delta_h X_t)^2}, \tag{5}$$

captures the relative changes of X_t and $\mathcal{E}_t$ on $[t, t+h]$.

The diffusion process X_t has continuous sample paths but almost all sample paths have infinite variation in any arbitrary small interval of time due to the diffusion part. The process X_t will not have a smooth sample path no matter what smoothness conditions on $\sigma(.)$ we make as long as $\sigma(.)$ does not vanish. The law of the iterated logarithm provides a previous statement about how X_t given in (1) oscillates on a neighborhood of t: $[t, t+h]$. Under the assumption A1, for every fixed $t \in [0, 1]$,

$$\limsup_{h \downarrow 0} \frac{|\Delta_h X_t|}{\sqrt{2h \log\log 1/h}} = \sigma(X_t),$$

with probability 1 (see, for example, Theorem 7.2.5, Arnold (1974)). Subject to the existence of the finite second moment of $\Delta_h X_t$,

$$E[(\Delta_h X_t)^2 | X_t = x] = \sigma^2(x) h + o(h) \text{ for } t \in [0, 1].$$

Under the assumption A2,

$$E(\Delta_h X_t)^2 = O(h). \tag{6}$$

[12] See, for example, Goldenberg and Schmidt (1996).

Similarly, $E(\Delta_h \mathcal{E}_t)^2$ can be written as a function of h for sufficiently small h. Let $B_h\gamma(t)$, a difference operator, be defined as $\gamma(t+h)-\gamma(t)$. Since $E(\Delta_h \mathcal{E}_t)^2 = B_{-h}B_h\gamma(0)$, under the assumption A3,

$$E(\Delta_h \mathcal{E}_t)^2 = O(h^l) \tag{7}$$

for some $l \geq 0$.

Combining (6) and (7), we obtain the noise to signal ratio

$$\rho(h) \equiv \frac{E(\Delta_h \mathcal{E}_t)^2}{E(\Delta_h X_t)^2} = O(h^{l-1}). \tag{8}$$

As $h \downarrow 0$, both $E(\Delta_h \mathcal{E}_t)^2$ and $E(\Delta_h X_t)^2$ decline but may have different declining rates. In general, $\rho(h)$ is a function of both h and l. It will therefore be useful to partition l according to the outcomes of $\rho(h)$:

$$\lim_{h\downarrow 0} \rho(h) = \begin{cases} 0, & l > 1, \\ O(1), & l = 1, \\ \infty, & l < 1. \end{cases} \tag{9}$$

It becomes evident from (9) that asymptotic properties of $\widehat{\sigma}_n^2$ will depend critically upon the level of l. The smoother the sample paths of $\mathcal{E}_t$ has, the higher value of l and therefore, the more likelihood of $\widehat{\sigma}_n^2$ to qualify as a consistent volatility estimator. Note that $E(\Delta_h \mathcal{E}_t)^2 = 2(\gamma(0)-\gamma(h))$. Hence, the behavior of $\rho(h)$ near the origin depends on the convergence rate of $\gamma(h)$ as $h \downarrow 0$.

To develop a sense of the behavior of the autocovariance (or, equivalently autocorrelation) function with $l > 1$, we compute autocorrelations for different h and various levels of l. From (7), $2(\gamma(0)-\gamma(h)) = Ch^l + o(h^l)$ for some constant $C > 0$. Hence, $\gamma(h) \approx \gamma(0) - \frac{C}{2}h^l$ and the autocorrelation $\frac{\gamma(h)}{\gamma(0)} = 1 - \frac{C}{2\gamma(0)}h^l$. For illustrative purpose, let the basic time unit be one hour. That is, $h = 1$ for hourly data. Suppose that the hourly series has an autocorrelation 0.1. Then, $\frac{C}{2\gamma(0)}$ will be 0.9 and autocorrelations for data sampled at different sampling frequencies are ready to be calculated. For example, autocorrelations for data sampled at every 10 minute are 0.895 and 0.938 for $l = 1.2$ and 1.5, respectively. Autocorrelations for minute-by-minute data become 99.3 (for $l = 1.2$) and 99.8 ($l = 1.5$), which are essentially one.

In fact, if $\gamma(h)$ changes very little near the origin, the noise component separated by a small time interval effectively cancels out since their correlation is considerably high. Unfortunately, as shown in Section 4, this is very unlikely to be satisfied in reality. In FXFX data, each bid and ask quote pair is input by a single dealer and therefore, it reflects each dealer's specific characteristics mentioned at the beginning of the paper. Since FXFX data consist of quotes collected from heterogeneous agents it is rare to have an autocorrelation so close to 1 (as illustrated in the above example) even if they are separated by tiny time intervals.

3 Estimating Volatility

3.1 The Characterization of Estimation Consistency

The main result in this section is that the quadratic variation of observable process Y_t turns out to be a consistent estimator of the signal volatility $\int_0^1 \sigma^2(X_t)dt$ if and only if $\rho(h) = o(1)$ as $h \downarrow 0$. Together with (9), the condition can be further formulated in terms of l.

Theorem 1 *Under assumptions A1-A4, the quadratic variation $\widehat{\sigma}_n^2$ is consistent for $\int_0^1 \sigma^2(X_t)dt$ if and only if*

$$\rho(h) = o(1) \tag{10}$$

as $h \downarrow 0$. If $B_{-h}B_h\gamma(0) = O(h^l)$ as $h \downarrow 0$, the condition (10) is equivalent to $l > 1$. Furthermore,

$$\widehat{\sigma}_n^2 = \begin{cases} \int_0^1 \sigma^2(X_t)dt + o(1) & \text{if } l > 1 \\ \int_0^1 \sigma^2(X_t)dt + \lim_{h\downarrow 0} h^{-1}B_{-h}B_h\gamma(0) + o(1) & \text{if } l = 1 \\ \infty & \text{if } l < 1 \end{cases} \tag{11}$$

The proof of the theorem is outlined in the Appendix. We provide an informal discussion here. The basic message of Theorem 1 is that if $\mathcal{E}_t$ is "smoother" than X_t (i.e., the case in which $l > 1$) the variability of $\mathcal{E}_t$ on $[t, t+h]$ is in higher-order for sufficiently small h and hence, decays faster than that of X_t as $h \downarrow 0$. Therefore, the ratio $\rho(h)$ decreases and ultimately approaches zero as we collect and use more sample data. As a consequence, the variability calculated by $\widehat{\sigma}_n^2$ is attributed by and large to the variability of X_t. In fact, $\widehat{\sigma}_n^2$ can be written as

$$\widehat{\sigma}_n^2 \equiv \sum(\Delta_h Y_i)^2 = \sum(\Delta_h X_i)^2 + \sum 2(\Delta_h X_i)(\Delta_h \mathcal{E}_i) + \sum(\Delta_h \mathcal{E}_i)^2. \tag{12}$$

The first summation in (12) is the quadratic variation based on $\{X_i\}$ and it converges to $\int_0^1 \sigma^2(X_t)dt$ in probability.[13] If $l > 1$, the last two summations are $o(1)$, which are in higher order than the first summation. Therefore, we conclude that the condition $l > 1$ preserves the consistency of $\widehat{\sigma}_n^2$.

In contrast to the case in which $l > 1$, if $l < 1$, $\sum(\Delta_h \mathcal{E}_i)^2$ is in order of $O(|h|^{l-1})$, which dominates the first and the second summations in (12). As a consequence, the variability of X_t on $[t, t+h]$ is in higher-order and the estimator $\widehat{\sigma}^2$ is made up mainly by those tiny variability of $\mathcal{E}_t$. The sum $\sum(\Delta_h \mathcal{E}_i)^2$ is unbounded above and hence, $\widehat{\sigma}_n^2$ grows without bound and fails to converge to $\int_0^1 \sigma^2(X_t)dt$ as $h \downarrow 0$.

[13] See, for example, Goldstein (1969).

The condition that $l = 1$ separates the two extreme cases ($l > 1$ and $l < 1$). If $l = 1$, decay rates of X_t and $\mathcal{E}_t$ are in the same order. The ratio $\rho(h)$ stays at a constant level as the sampling frequency changes. The estimator $\widehat{\sigma}_n^2$ will have a limit but provides a biased estimate for $\int_0^1 \sigma^2(X_t)dt$.

Although $\sum(\Delta_h \mathcal{E}_i)^2$ is unbounded for $l < 1$, the scaled $h^{1-l}\sum(\Delta_h \mathcal{E}_i)^2$ has a limit as $h \downarrow 0$. Similarly, the second cross summation in (12) scaled by the same factor, h^{1-l}, is in the order $o(1)$. Therefore, the scaled sum $h^{1-l}\widehat{\sigma}_n^2$ has a limit. This result comes directly from the proof of Theorem 1. Because of its importance in understanding dynamics of the HFN as $h \downarrow 0$ (see Section 4), we state the result as a theorem.

Theorem 2 *Under the assumptions A1-A4, if* $l < 1$,

$$h^{1-l}\widehat{\sigma}_n^2 - \frac{B_{-h}B_h\gamma(0)}{h^l} = o(1). \tag{13}$$

3.2 The Diffusion with Jumps

Because the quadratic variation shares some similar properties when applied to a mixed diffusion-jump process, a few remarks are in order before we discuss several examples of noise processes in the next section. A number of studies have investigated a model in which stock prices follow a mixed diffusion-jump process (see, for example, Merton (1976)). Although almost all sample paths of diffusion are continuous, the introduction of a jump component allows for simple sample-path discontinuities. Similar to the mixed diffusion-stationary processes, the fluctuation of the diffusion is overestimated if a model without jumps is used in a situation where jumps do occur. However, unlike the mixed diffusion-stationary process specified in Section 1.2, estimators which account for jump components can be derived in some simple cases. For example, consider a diffusion with Poisson jumps:

$$dX_t = \mu X_t dt + \sigma X_t dW_t + X_t d\eta_t, \tag{14}$$

where η_t is a compound Poisson process

$$\eta_t = \sum_{i=1}^{N_t} a_t,$$

Here N_t is a Poisson process while the a_ts are independent identically distributed (i.i.d.) random variables. Equation (14) yields a unique solution which is given by

$$X_t = x_0 e^{(\mu-1/2)t+\sigma W_t + \sum_{i=1}^{N_t}\log(1-a_i)}.$$

It follows that as $h \downarrow 0$,

$$h(\sum_{i=1}^{n}(\Delta_h \text{log} X_i)^2 - \sum_{s \leq h}(\Delta_h R_s)^2) = \sigma^2 + o(1)$$

in probability, where $R_t = \log(X_t)$ (Sørensen (1991)). The Poisson process is not a stationary process and, as a result, the previous section does not apply directly. However, the analysis may be generalized for processes with jumps in X_t. Although such an exercise is certainly interesting, it is mathematically intimidating and is beyond the scope of this paper.

3.3 More on Noise Processes

An important example of the noise process is the Gaussian process with a general exponential covariance function

$$\gamma(h) = ce^{-\alpha h^l}, \tag{15}$$

with c and $\alpha > 0$. It can be proved that (15) is a covariance function for $0 < l \leq 2$. (15) is necessarily a correlation function if $0 < l \leq 1$ by Pòlya (1949) and if $1 < l \leq 2$ by Lèvy (1925). It is also known that (15) can not be a correlation function if $l > 2$ (Yaglom (1987)). In the case in which $0 < l \leq 2$, it is easy to verify that as $h \downarrow 0$,

$$B_{-h}B_h\gamma(0) = 2c\alpha h^l + o(h^l).$$

Therefore, we have

$$\frac{B_{-h}B_h\gamma(0)}{h^l} - 2c\alpha = o(1). \tag{16}$$

In the case where $l > 1$, the variability of $\mathcal{E}_t$ on any tiny time interval $[t, t+h]$, $E(\Delta_h\mathcal{E}_t)^2$, collapses to zero at a fast (faster than those of diffusion processes) rate. Because $E(\Delta_h\mathcal{E}_t)^2$ dies down quickly, $\widehat{\sigma}_n^2$ is consistent for $\int_0^1 \sigma^2(X_t)dt$. In fact, considerable short-range dependence characterized by an autocovariance function $\gamma(h)$ varying very little near the origin is crucial for consistency of high-frequency volatility estimators.

If $l = 1$, (15) leads to the first-order autoregressive (the Ornstein-Uhlenbeck) process. Applying results of Theorem 1 as well as (16), we have

$$\widehat{\sigma}_n^2 - (\int_0^1 \sigma^2(X_t)dt + 2c\alpha) = o(1). \tag{17}$$

Since $2c\alpha > 0$, the volatility of X_t is overestimated. It is also important to notice that the estimation bias, $2c\alpha$, depends not only on the variance of $\mathcal{E}_t$ but on the parameter α as well.

The case where $0 < l < 1$ is more interesting and appears more empirically relevant (see empirical results for FXFX data in Section 4). If $l < 1$, since $h^{1-l}\widehat{\sigma}_n^2$ converges to $h^{-l}B_{-h}B_h\gamma(0)$ in probability by Theorem 2, the shape of $h^{-1}B_{-h}B_h\gamma(0)$, or $2c\alpha h^{l-1}$, approximates the un-scaled quadratic variation $\widehat{\sigma}_n^2$. For all $l < 1$, $2c\alpha h^{l-1}$ or $\widehat{\sigma}_n^2$ explodes as $h \downarrow 0$. However, they differ dramatically in terms of the explosion rate. The curve with a smaller l value grows more quickly than the one with a larger l value.

An alternative specification of autocovariance is that γ has a spike at $h = 0$. That is,

$$\gamma(h) = \begin{cases} c_0, & h = 0 \\ ce^{-\alpha h^k}, & \text{otherwise} \end{cases} \tag{18}$$

with c and $\alpha > 0$, and $k > 0$. If $c_0 \neq c$,

$$B_{-h}B_h\gamma(0) = O(1)$$

as $h \downarrow 0$. Then it follws from Theorem 1 for $l = 0$ that the estimate of $\int_0^1 \sigma^2(X_t)dt$ given by $\widehat{\sigma}_n^2$ is inconsistent and the bias explodes as fast as h^{-1}.

If allowing $c = 0$, it follows that the autocovariance function has the form

$$\gamma(0) = \begin{cases} c_0, & h = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Although such a continuous "white noise" process does not exist, except in a highly degenerate sense, its discrete version has been widely used in modeling financial assets. For example, Merville and Pieptea (1989) considered the mean-reverting diffusion with white noise process:

$$dY_t = (\mu - kY_t)dt + \sigma dW_t + \delta dZ_t \tag{20}$$

with dZ_t drawing from i.i.d. $N(0, 1)$, independent from dW_t. The instantaneous variance of Y_t is given by

$$var(\Delta Y_t) = \sigma^2 h + \delta^2.$$

When h increases, so does $var(\Delta_h Y_t)$ at a rate σ^2. As h decreases, $var(\Delta_h Y_t)$ decreases at the same rate. However, as $h \downarrow 0$, $var(\Delta_h Y_t)$ does not converge to zero but rather to the noise-related variance δ^2.

Note that the stationary process with a covariance function defined either by (15) or (18) exhibits very slow (hyperbolic) decay rate near the origin. However, both autocovariance functions in (15) and (18) decline exponentially as $h \uparrow \infty$, implying that they will not possess the *long-memory* property under the jargon of the conventional asymptotics, which focuses on the behavior of autocovariance (or autocorrelation) function at infinity.[14]

Andersen and Bollerslev (1997, 1998a) and Chambers (1998) studied the persistence of long-memory property under temporal aggregation. Their studies suggested that by interpreting the volatility as a mixture of numerous heterogeneous short-run information arrivals, the observed volatility process may exhibit long-run dependence. If the long-memory dependence in low-frequency volatility is the direct consequence of the short-run component, the process with the autocovariance function (15) or (18) is clearly inappropriate. Nevertheless, the question of whether the process with the autocovariance function given in (15) or (18) is an appropriate approximation of the statistical distribution of the HFN may not be crucial for the statistical aspects of volatility study of high-frequency data – after all, our analysis based on the continuous record asymptotics only requires knowledge of the dependence property of the HFN near the origin. However, the behavior of the autocovariance function of the HFN at the infinite is central to the interpretation of the origination of the long-memory property of low-frequency data. Since the present study is focused exclusively on analysis in the high-frequency domain we shall not pursue this issue any further here.[15]

3.4 The Florens-Zmirou Estimator

Florens-Zmirou (1993) proposed the following nonparametric volatility estimator

$$\widehat{s}_n^2(\tau) = \frac{1}{2n\xi_n} \sum_{i=1}^{n-1} 1_{|\tau - i/n| < \xi_n} (\Delta_h X_i)^2, \tag{21}$$

where $\tau \in [0,1]$ and ξ_n is a nonstochastic sequence converging to zero at an appropriate rate as $h \downarrow 0$. The estimator $\widehat{s}_n^2(\tau)$ is a kernel estimator. Unlike the quadratic variation, $\widehat{s}_n^2(\tau)$ is consistent for $\sigma^2(X_\tau)$. The following result on $\widehat{s}_n^2(\tau)$ is an adaptation of the results in Theorem 2.3 of Corradi and White (1999).

Theorem 3 *Under the assumption A1, as $h \downarrow 0$, $h^{-1}\xi_n \uparrow \infty$ and $h^{-1}\xi_n^3 \downarrow 0$,*

$$\widehat{s}_n^2(\tau) - \sigma^2(X_\tau) = o(1)$$

[14] A process is usually called to have a long memory property if the absolute value of its autocovariance function declines hyperbolically when $h \uparrow \infty$ (i.e., $|\gamma(h)| = O(|h|^{-k})$ as $h \uparrow \infty$ for some $k > 0$).

[15] Note that since the class of all stationary correlation functions coincides with the class of positive definite functions (see, for example, Yaglom (1987)), it is easy to find a stationary process with desired statistical short- and long-run properties. For example, a simple form which adds long- and short-memory components may be appropriate. See Barndorff-Nielsen and Shephard (1998) for such examples. See also Baillie (1996) for more discussion on the fractional Gaussian noise.

pointwise in $\tau \in [0,1]$ *in probability.*

Although $\widehat{s}_n^2(\tau)$ and $\widehat{\sigma}_n^2$ are designed with different purposes in mind, they will have similar properties when applied to data contaminated by a random noise. Results derived for $\widehat{\sigma}_n^2$ hold, at least qualitatively, for $\widehat{s}_n^2(\tau)$. We illustrate this by considering a simple case when $\sigma(.)$ is a constant, σ_0 $(0 < \sigma_0 < \infty)$.

In the constant volatility case ($\sigma^2(.) \equiv \sigma_0^2$), the two estimators are asymptotically equivalent. In fact, Corradi and White (1999) showed that

$$\{diag(\sqrt{2\widehat{\sigma}_n^2})\}^{-1}(n\xi_n)^{1/2}((\widehat{s}_n^2(\tau_1) - \widehat{\sigma}_n^2, (\widehat{s}_n^2(\tau_2) - \widehat{\sigma}_n^2, ..., (\widehat{s}_n^2(\tau_m) - \widehat{\sigma}_n^2))' \sim N(0, I_m). \qquad (22)$$

The following results, establishing the condition for the consistency of $\widehat{s}_n^2(\tau)$ for σ_0^2, can be obtained by applying (22) and results in Theorem 1.

Theorem 4 *Under assumptions A1-A4, if* $\sigma(.) \equiv \sigma_0$ *with* $0 < \sigma_0 < \infty$*,*

$$\widehat{s}_n^2(\tau) - \sigma_0^2 = o(1)$$

in probability if and only if

$$\rho(h) = o(1)$$

as $h \downarrow 0$*,* $h^{-1}\xi_n \uparrow \infty$ *and* $h^{-1}\xi_n^3 \downarrow 0$*.*

4 Empirical Results for the DM/$ Exchange Rate

4.1 Spurious Volatility in High-frequency Returns

As an empirical application, we analyze the high-frequency DM/$ in the framework proposed in this paper. The raw data were recorded from the Reuters foreign exchange FXFX page and contain Greenwich Mean Time (GMT) at which quotes are recorded, bid and ask prices, and information on quote origination. These Reuters FXFX quotes are from more than 592 banks who are a part of the spot market. We will work on the mid-point of the logarithms of the bid and the ask prices as in most previous studies. We focus on one year time span from October 1, 1992 to September 30, 1993. For the almost continuously recorded quotes over the one-year period, there are total $1,472,241$ ticks, which yield about 4034 ticks per day and relatively higher numbers (5598) for weekdays because the market was usually quiet during weekends. We exclude weekends for which the absence of quotes may produce unreliable results.

Most existing empirical literature on FX market microstructure uses these indicative quote data derived from Reuters FXFX screens[16] or other competitors such as Knight Ridder and Telerate.[17] As the only information source available to all market participants, these quotes *indicate* the current foreign exchange rates. It is well-known that such high-frequency quotes, recorded on a second-by-second basis, are very "noisy" (see, for example, Goodhart and O'Hara (1997)). To keep the noise low, most current empirical research based on high-frequency FXFX data does not use full data but a sub-sample obtained by sampling data at some regular time intervals. The most popular sampling frequencies are 10- or 15-minutes (for example, see Guillaume at al. (1997), Andersen and Bollerslev (1997), and Melvin and Peiers (1997) among others).

By sampling data at 10- or 15-minutes, the HFN impact declines dramatically from that at the tick-by-tick level. Table 1 presents estimation results of daily volatilities using the quadratic variation for the DM/$ series with different sampling frequencies. As can be seen, as the sampling frequency increases, the estimates do not converge but rather inflate. This suggests that rapid quote arrivals are consisted with rapid noise arrivals. The evidence in Table 1 also suggests that the noise is less "smoother" than the diffusion at ultra-high sampling frequencies. For relatively low sampling frequencies, the general level of the process changes little, with most variation coming from the diffusion. As the sampling frequency increases, however, the general level of volatility changes, the variation due to the noise component becomes significant in relation to the variation of diffusion. At the ultra-high frequency, diffusion becomes less important and noise takes over. By increasing sampling frequency enough, we should see only the variation from the noise. See Sections 4.2 and 5 for further discussion.

The evidence that the HFN induces an upward bias in volatility estimation shown in Table 1 may be well as expected. However, the magnitude of the bias is striking.[18] Recall that the consistency of volatility estimation depends on the degree of smoothness of the autocovariance function $\gamma(h)$ around 0. If the sample path of the HFN is not smoother than that of the diffusion, the noise component separated by a small time interval effectively can not cancel out and tends to dominate the estimates obtained by the quadratic variation. Although each of $(\Delta_h \mathcal{E}_i)^2$ is tiny

[16] FXFX is the older version of the currently used EFX page. The EFX data, however, have similar shortcomings as those in the FXFX data. Like the old FXFX system, the EFX page provides only indicative quotes that are not binding commitments. It also does not contain information on traded currency volumes.

[17] According to Reuters, about 60% of transactions in the interbank market take place through the Reuters FXFX system (Evans (1998)).

[18] In fact, the HFN bias may be more serious than those reported here in empirical analysis using other similar data sets. As pointed out in Evan (1998), the FXFX indicative quotes cannot arrive faster than every 5 seconds (due to technical constraints in the Reuters system). At active times like the London afternoon, this constraint is often binding and only a small fraction of the quotes has been posted.

Table 1: *Volatility Estimates based on the Quadratic Variation. Daily volatility estimates for DM/$ from October 1, 1992 to September 30, 1993. Data are sampled at various frequencies indicated.*

	Tick-by-tick	Every One Minute	Every Five Mintues	Every Ten Minutes	Every Fifteen Minutes
Monday	2.935e-04	9.036e-05	5.810e-05	5.671e-05	5.264e-05
Tuesday	3.351e-04	1.022e-04	6.259e-05	5.995e-05	5.674e-05
Wednesday	3.505e-04	1.040e-04	6.303e-05	5.857e-05	5.421e-05
Thursday	3.810e-04	1.207e-04	7.660e-05	6.983e-05	6.758e-05
Friday	3.597e-04	1.106e-04	7.850e-05	7.302e-05	6.776e-05
Overall Mean	3.440e-04	1.056e-04	6.776e-05	6.362e-05	5.979e-05

when h is infinitesimally small, the HFN bias, approximated by $\sum(\Delta_h \mathcal{E}_i)^2$, may not be negligible if the number of terms in the summation is sufficiently large.

4.2 High Irregularities of the HFN Sample Path

To develop further intuition for these results, we use the result from Theorem 2 to estimate the smooth parameter of the noise component, l. The evidence reported in Table 1 suggests that the decay rate of γ, l, is less than 1 for the noise process. Recall that, for $l < 1$, the scaled quadratic variation has a finite limit. Taking the logarithm of (13) leads to the approximate equation

$$\ln \widehat{\sigma}_n^2 \approx \ln(lim_{h\to 0}\frac{B_{-h}B_h\gamma(0)}{h^l}) + (l-1)\ln(h). \quad (23)$$

The relationship between $\ln \widehat{\sigma}_n^2$ and $\ln(h)$ is linear. The coefficients, $\ln(lim_{h\to 0}\frac{B_{-h}B_h\gamma(0)}{h^l})$ and l, are estimated using the regression technique. For each day of the one year period examined, we estimate (23) based on 50 observations obtained using different sampling frequencies ranging from every 5 seconds up to every 25 minutes.[19]

Table 2 reports the estimation results. Results vary from Mondays to Fridays. However, the evidence that the estimated l is well less than one is consistent across different days of the week. The average estimated l for weekdays is about 0.323. The smallest average estimated l, 0.278, is on Thursdays.

The overall average estimated $\ln(lim_{h\to 0}\frac{B_{-h}B_h\gamma(0)}{h^l})$ is -14.11. Since $E(\Delta_h\mathcal{E}_t)^2 \approx B_{-h}B_h\gamma(0)$, we may have an estimate of $E(\Delta_h\mathcal{E}_t)^2$ based on $\ln(lim_{h\to 0}\frac{B_{-h}B_h\gamma(0)}{h^l})$. The biases due to the HFN,

[19]Note that (23) is an asymptotic equation. To reduce potential estimation bias due to assuming that it holds for all h, we use only sampling frequencies up to 25 minutes.

Table 2: *Irregularities of the HFN Sample Path. Daily volatility estimates for DM/$ from October 1, 1992 to September 30, 1993. Estimates are based on (22). Standard errors are given in parentheses.*

	l	$\ln(lim_{h\to 0}\frac{B_{-h}B_h\gamma(0)}{h^l})$
Monday	0.364 (0.250)	-13.93 (2.41)
Tuesday	0.291 (0.194)	-14.32 (1.08)
Wednesday	0.314 (0.361)	-14.22 (1.92)
Thursday	0.278 (0.184)	-14.48 (1.29)
Friday	0.367 (0.273)	-13.62 (2.25)
Overall Mean	0.323	-14.11

Table 3: *HFN Biases. The estimate is based on* $\sum(\Delta_h \mathcal{E}_i)^2 \approx n^{1-l} B_{-h} B_h \gamma(0)$ *with* l *and* $B_{-h}B_h\gamma(0)$ *given in Table 2.*

	n	$\sum(\Delta_h \mathcal{E}_i)^2$	$\frac{\sum(\Delta_h \mathcal{E}_i)^2}{\hat{\sigma}_n^2}$
Monday	5415	2.11e-04	72%
Tuesday	6008	2.88e-04	86%
Wednesday	5934	2.61e-04	75%
Thursday	5707	2.65e-04	70%
Friday	4926	2.64e-04	74%
Overall Mean	5598	2.57e-04	75%

which are dominated by $\sum(\Delta_h \mathcal{E}_i)^2$ (see (12)), can also be calculated. Results are reported in Table 3. To evaluate the proportion of variability due to the HFN, we also provide the ratio $\frac{\sum(\Delta_h \mathcal{E}_i)^2}{\widehat{\sigma}_n^2}$. It becomes evident in Table 3 that more than 70% of measured daily return variances from $\widehat{\sigma}_n^2$ can be induced by the HFN. This proportion may be as high as 86% (as on Tuesdays during the examined sample period).

4.3 First-order Autocorrelation in High-frequency Returns

It is well-known that there is a substantial negative first-order autocorrelation in the FXFX data. As quotes are sampled less frequently, the autocorrelation becomes less significant and usually diminishes at frequencies exceed 5 minutes. The model introduced in this paper provides an explanation on why and how this happens. According to (2), the first-order autocovariance of the observed return series $\Delta_h Y_t$ induced by the HFN is $2\gamma(h) - \gamma(0) - \gamma(2h)$. As a consequence, the autocorrelation induced by the HFN, denoted by ϑ, is $[2\gamma(h) - \gamma(0) - \gamma(2h)]/[2(\gamma(0) - \gamma(h)) + O(h)]$. By a first order Taylor series expansion on $\gamma(h)$ near $h = 0$, ϑ equals approximately $[(2^l - 2)h^l]/[2h^l + O(h)]$. For $l < 1$, $[(2^l - 2)h^l]/[2h^l + O(h)]$ is negative for sufficiently small h and approaches $(2^{l-1} - 1)$ as $h \downarrow 0$. As quotes are sampled less frequently, the autocorrelation induced by the HFN becomes less negative due to the increasing importance of the diffusion component and eventually will diminish, becoming statistically insignificant.[20]

5 The Optimal Sampling Frequency

The empirical analysis in the previous section provides distinct evidence for the existence of multiple volatility components at the high intradaily frequencies. In particular, while the volatility dynamics at relatively low frequencies are governed primarily by the persistent diffusion component, it is dominated by a much more short-run component (the HFN) as data are sampled at ultra-high frequencies. These findings suggest that there may exist an optimal sampling frequency at which one has a large number of observations to insure the consistency of the estimation while the biases inherent in high-frequency data are controlled at an appropriate level.

In this section, we show via Monte Carlo simulations that for Reuters FXFX data, 10- to 15-minute sampling intervals are optimal based on mean square error (MSE) type of criteria. The

[20]Note that if $\mathcal{E}_t$ is uncorrelated (the case discussed by Merville and Pieptea (1989)), $[2\gamma(h) - \gamma(0) - \gamma(2h)]/[2(\gamma(0) - \gamma(h)) + O(h)] \approx [-\gamma(0)]/[2\gamma(0) + O(h)]$, which is bounded below by $-1/2$ and approaches $-1/2$ as $h \downarrow 0$.

simulation is based on the following model:

$$Y_t = X_t + \mathcal{E}_t,\ \ t \in [0, T], \tag{24}$$

where X_t is assumed to be a lognormal stochastic volatility (SV) model

$$dX_t = \sigma_t dW_t \tag{25}$$

$$d\ln\sigma_t = (\alpha - \beta\ln\sigma_t)dt + \gamma dW_t^*. \tag{26}$$

W_t and W_t^* are two independent standard Wiener processes. We choose values of the parameters (α, β, γ) that yield an average 0.7% daily standard deviation of returns of the X_t process. The HFN, $\mathcal{E}_t$, is a stationary Gaussian process with the following model specification:

$$d\mathcal{E}_t = a_\varepsilon \mathcal{E}_t dt + \sigma_\varepsilon dW_t^l, \tag{27}$$

where W_t^l is a fractional Brownian motion with order $l = 0.323$, independent of both W_t and W_t^*. The parameters a_ε and σ_ε are chosen to generate a proper noise level, namely, 75% of total variability of returns of Y_t are from $\{\mathcal{E}_t\}$ at the tick-by-tick level.

The SV model is chosen because of the growing consensus in recent studies that volatilities do change over time. The use of fractional Brownian motion is primarily due to its special autocorrelation structures, which make it an ideal vehicle to generate a stationary component with properties uncovered in Reuters FXFX data in Section 4. More specifically, the process $\mathcal{E}_t$ with $l < 1$ has a sample path which is less "smoother" than that of X_t defined by (25) and (26).[21]

To visualize the variability change of X_t relative to that of $\mathcal{E}_t$ as the sampling frequency changes, Figure 1 reports several simulated series based on (24)-(27) with parameters chosen to

[21] The solution of (27) can be written as (Comte and Renault (1996)):

$$\mathcal{E}_t = \int_0^t e^{a_\varepsilon(t-s)}\sigma_\varepsilon dW_s^l$$

and the covariance function $\gamma(h)$ of $\mathcal{E}_t$ is given by

$$\gamma(h) = \frac{1}{\Gamma((l+1)/2)}\int_0^\infty (x+h)^{(l-1)/2}x^{(l-1)/2}a(x+h)a(x)dx$$

where $a(x)$ is defined as

$$a(x) = \sigma_\varepsilon x^{(l-1)/2}\frac{d}{dx}\int_0^x e^{a_\varepsilon u}(x-u)^{(l-1)/2}du.$$

Under some regularity conditions, $\gamma(h) = \gamma(0) + c_\gamma h^l + o(h^l)$ for some constant c_γ (Comte (1994)) and hence, $B_{-h}B_h\gamma(0) = O(h^l)$, where $\gamma(h)$ is the covariance function of $\mathcal{E}_t$.

generate required volatility levels, as mentioned above. One hundred observations for returns of X_t and $\mathcal{E}_t$ as well as Y_t are generated for each of the three sampling frequencies: daily, hourly, and tick-by-tick (i.e., 5600 observations per day according to the results in Table 3). The results show that daily returns of Y_t are governed by the signal process X_t and the noise component $\mathcal{E}_t$ is negligible (Figure 1 (a)). From Figure 1 (b), it can be seen that hourly returns are essentially dominated by the signal process but the impact of the noise component becomes clearly visible. At the tick-by-tick frequency, however, the observed returns are a good mixture of signal and noise and are driven mainly by the noise component (Figure 1 (c)).

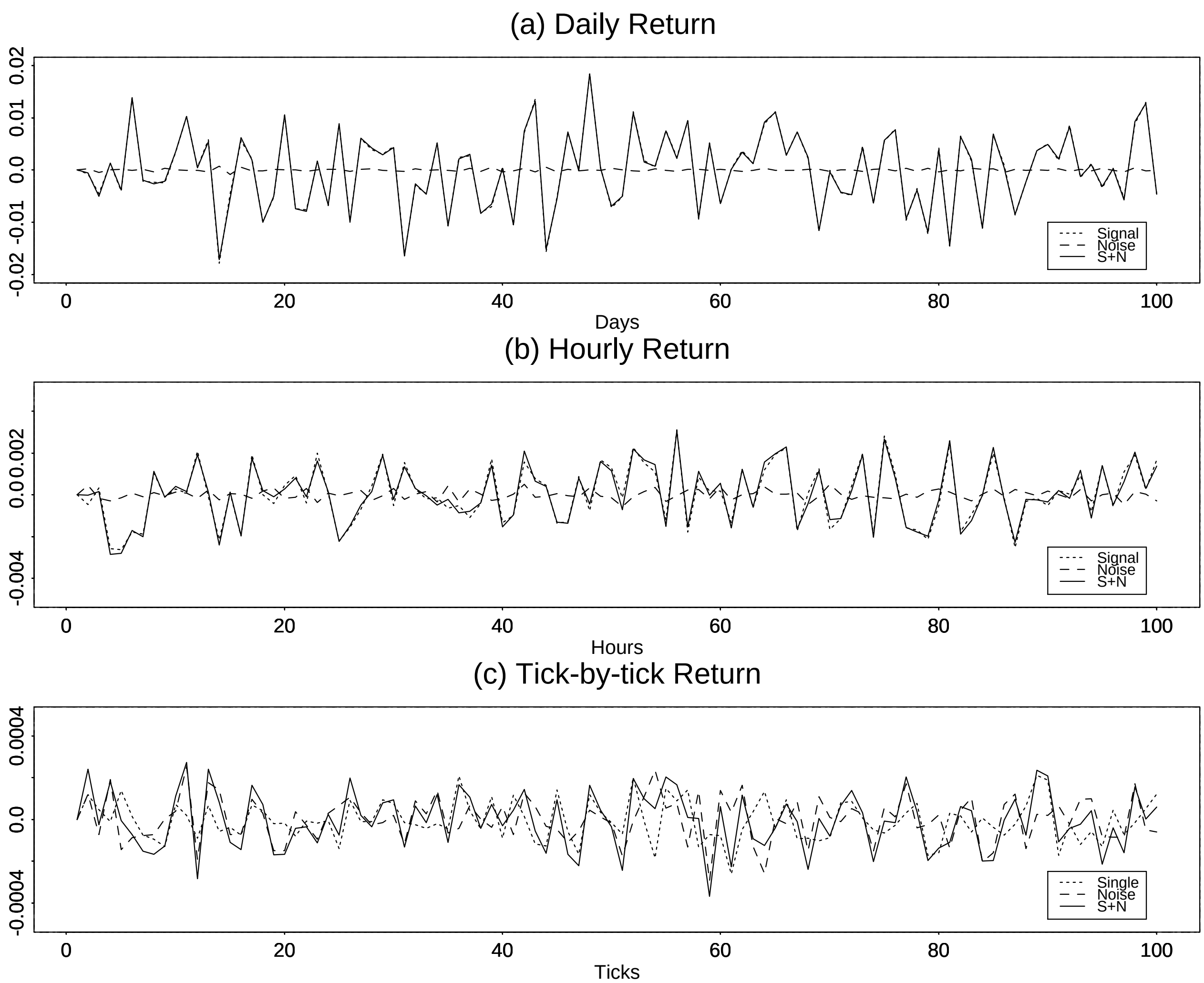

Figure 1. Simulated sample paths.

To find the optimal sampling frequency, the quadratic variation (3) is applied to simulated

Table 4: *SRMSE*[a]

	Sampling Interval							
	1 Hour	30 Min.	20 Min.	15 Min.	10 Min.	5 Min.	1 Min.	Tick-by-tick[b]
SRMSE ($\times 10^4$)	2.156	1.158	1.060	0.962	0.964	1.366	3.768	9.570

a. The simulation is based on 3,000 replications.
b. Based on 5600 observations per day.

series to estimate daily volatility. Since the time span $[0, T]$ is fixed as one day, we take $T = 1$. Hence, the relationship between the sample size n and the length of sampling interval h, is $nh = T \equiv 1$. For example, $n = 1440$ and $h = 1/1440$ for data sampled every minute; while $n = 144$ and $h = 1/144$ for data collected every ten minutes.

The square root of MSE (SRMSE) for various levels of sampling frequencies are reported in Table 4. The results show that as the sampling frequency increases, the SRMSE first decreases and then increases. The optimal sampling frequency is about 10 to 15 minutes. The SRMSE of tick-by-tick data (9.57×10^{-4}) is about 10 times higher than the optima (0.96×10^{-4}). The U-shape curve of SRMSE clearly demonstrates the importance of the trade-off between increasing of sample size and minimizing HFN biases.

6 Conclusion

We focus two research questions which are important in high-frequency data analysis. First, what are conditions characterizing the HFN component on which conventional volatility estimators will provide consistent volatility estimates? We have established such conditions under the framework of continuous record asymptotics. A tentative conclusion from this study is that more frequent samplings need not provide better volatility estimates. More specifically, conventional volatility estimators, commonly used in daily and weekly data analysis, may not create consistent continuously filtered volatility estimates under certain circumstances. We illustrate our findings with the widely used HFDF93 high-frequency Reuters FXFX DM/$ data released from Olsen and Associates in 1993. We found that the HFN leads to substantial spurious volatility in high-frequency quotation series; about 75 % of measured daily return variances can be induced by the HFN component with high irregularities in the sample path. We have argued that trader heterogeneity helps explain these empirical findings. If all traders are homogeneous, the correlation between consecutive price quotes should be nearly one and as a consequence, one observes a "very smooth" HFN sample path. If traders are heterogeneous, however, the correlation between consecutive price quotes may not be nearly one. Therefore, one observes high irregularities of the HFN sample path, which gives rise to "excessive volatility" in high-frequency returns.[22]

A second, and related, question is how much impact does the HFN component have on the price series dynamics, especially, on the behavior of volatility as the sampling frequency changes? When analyzing volatility for data sampled at ultra-high frequencies, the HFN component tends to dominate the estimates obtained with conventional methods, whereas properties of data sampled less frequently are mainly driven by the diffusion component. For example, we shown that the HFN is a potential source of some short-lived dynamics in return series such as the apparent negative first-order autocorrelation in returns based on samples collected at 5 minute time intervals or less. Our results also provided a justification for the appropriateness of the use of the 10- to 15-minute sampling intervals. Data sampled at these frequencies are the ideal compromise, yielding a large number of observations while minimizing the biases inherent in high-frequency data.

Appendix

Proof of Theorem 1: Let $n = h^{-1}$ be the number of intervals in the partition of $[0,1]$. Under the assumption A3, the sequence $\frac{1}{n}\sum(\Delta_h \mathcal{E}_i)^2$ converges to $E(\Delta_h \mathcal{E}_t)^2 = O(h^l)$ by the law of large number, and therefore $\sum(\Delta_h \mathcal{E}_t)^2 = O(h^{l-1})$ as $h \downarrow 0$. Since X_t is a diffusion process, its quadratic variation $\sum(\Delta_h X_i)^2$ converges to $\int_0^1 \sigma^2(X_t)dt$. Finally, by Hőlder's inequality together with A2-A5, the quadratic covariance $\sum \Delta_h X_i \Delta_h \mathcal{E}_i$ is $O(h^{l/2})$. So the quadratic variation $\widehat{\sigma}_n^2$ is consistent for $\int_0^1 \sigma^2(X_t)dt$ if and only if $l > 1$, and the result in (12) follows.

References

[1] Andersen, T. G. and T. Bollerslev, 1997, "Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns," *Journal of Finance,* **52**, 975-1005.

[2] Andersen, T. G. and T. Bollerslev, 1998a, "Towards a unified framework for high and low frequency return volatility modeling," forthcoming *Statistica Neerlandica.*

[22] Although how the HFN evolves over time is likely to be market specific, similar findings have been documented in high-frequency equity data. For example, Kaul and Nimalendran (1990) studies the impact of the bid-ask spread on market overreaction. They found that bid-ask errors in NASDAQ transaction prices are a predominant source of stock price reversals. After extracting measurement errors caused by bid-ask spread, they reported stock returns are positively, and not negatively, autocorrelated. It was also shown that bid-ask spread errors lead to substantial spurious volatility and the bid-ask error component can explain over 50% of daily return variance.

[3] Andersen, T. G. and T. Bollerslev, 1998b, "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts," *International Economic Review*, **39**, 885-905.

[4] Baillie, R., 1996, "Long memory processes and fractional integration in econometrics," *Journal of Econometrics*, **73**, 5-59.

[5] Baillie, R. T. and T. Bollerslev, 1990, "Intra-day and inter-market volatility in foreign exchange rates," *Review of Economic Studies,* **58**, 565-585.

[6] Barndorff-Nielsen, O. and N. Shephard, 1998, "Aggregation and model construction for volatility models, Working paper, University of Aarhus and Nuffield College.

[7] Bergstrom, A., 1976, *Statistical Inference in Continuous Time Economic Models,* North Holland, Amsterdam.

[8] Bollerslev, T., R. Chou, and K. Kroner, 1992, "ARCH modeling in finance: A review of the theory and empirical evidence," *Journal of Econometrics,* **52**, 5-59.

[9] Bollerslev, T., R. Engle, and D. Nelson, 1994, "ARCH models," in R. Engle and D. McFadden (eds.), *Handbook of Econometrics,* Vol. IV, Elsevier, Amsterdam.

[10] Bollerslev, T. and I. Domowitz, 1993, "Trading patterns and prices in the interbank foreign exchange market," *Journal of Finance*, **48**, 1421-1443.

[11] Bollerslev, T. and M. Melvin, 1994, "Bid-ask spreads and volatility in the foreign exchange market," *Journal of International Economics*, **36**, 355-372.

[12] Chambers, M., 1998, "Long memory and aggregation in macroeconomic time series," *International Economic Review,* **39**, 1053-1072.

[13] Chesney, M. and L. Scott, 1989, "Pricing European currency options: A comparison of the modified Black-Scholes model and a random variance model," *Journal of Financial and Quantitative Analysis,* **24**, pp. 267-285.

[14] Cheung, Y. and M. Chinn, 1999, "Macroeconomic implications of the beliefs and behavior of foreign exchange traders," *Working paper*, University of California, Santa Cruz.

[15] Comte, F. and E. Renault, 1996, "Long memory continuous time models," *Journal Econometrics*, **73**, 101-149.

[16] Comte, F., 1994, "Simulation and estimation of long memory continuous time models," *Working paper*, Centre de Recherche en Economie et Statistique.

[17] Corradi, V. and H. White, 1999, "Specification tests for the variance of a diffusion," *Journal of Time Series Analysis,* **20**, 254-270.

[18] Evans, M. D., 1998, "The microstructure of foreign exchange dynamics," Working paper, Georgetown University

[19] Florens-Zmirou, D., 1993, "On estimation of the diffusion coefficient from discrete observations," *Journal of Applied Probability,* **30**, 790-804.

[20] Froot, K., D. Scharfstein and J. Stein, 1992, "Herd on the street: Informational inefficiencies in a market with short-term speculation," *Journal of Finance,*, **47**, 1461-1484.

[21] Goldenbery, D. and R. Schmidt, 1996, "On estimating the expected rate of return in diffusion price models with application to estimating the expected return on the market." *Journal of Financial and Quantitative Analysis,* **31**, 605-631.

[22] Glosten, L., 1987, "Components of the bid-ask spread and statistical properties of transaction prices," *Journal of Finance*, **42**, 1293-1307.

[23] Goldstein, J., 1969, "Second Order Itô processes," *Nagoya Math. J. ,* **36**, pp. 27-63.

[24] Goodhart, C. and M. O'Hara, 1997, "High frequency data in financial markets: Issues and applications," *Journal of Empirical Finance,* **4**, 73-114.

[25] Guillaume, D. M., M. M. Dacorogna, R. R. Davé, U. A. Müller, R. B. Olsen, and O. V. Pictet, 1997, "From the bird's eye to the microscope," *Finance and Stochastics*, **1**, 95-129.

[26] Harvey, A., E. Ruiz, and E. Sentana, 1992, "Unobserved component time series models with ARCH distributions," *Journal of Econometrics,* **52**, 129-157.

[27] Hausman, J., A. Lo and A. C. MacKinlay, 1992, "An ordered probit analysis of transaction stock prices," *Journal of Financial Economics*, **31**, 319-379.

[28] Hull, J. and A. White, 1987, "The pricing of options on assets with stochastic volatilities," *Journal of Finance*, **42**, 281-300.

[29] Karatzas, I. and S. Shreve, 1988, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

[30] Kaul G. and M. Nimalendran, 1990, "Price reversals: Bid-ask errors or market overreaction?" *Journal of Financial Economics,* **28**, 67-93.

[31] Levỳ, P., 1925, *Calcul des probabilitès,* Gauthier-Villars, Paris.

[32] Lyons, R., 1995, "Tests of microstructural hypotheses in the foreign exchange market," *Journal of Financial Economics,* **39**, 321-351.

[33] Lo, A. and A. C. MacKinlay, 1990, "An econometric analysis of nonsynchronous-trading," *Journal of Econometrics*, **45**, 181-212.

[34] Melino, A. and S. Turnbull, 1990, "Pricing foreign currency options with stochastic volatility," *Journal of Econometrics,* **45**, pp. 239-265.

[35] Melvin, M. and B. Peiers, 1998, "The global transmission of volatility in the foreign exchange market," *Working paper*, Arizona State University.

[36] Merton, R., 1976, "Option pricing when underlying stock returns are discontinuous," *Journal of Financial Economics,* **3**, 125-144.

[37] Merton, R., 1980, "On estimating the expected return on the market," *Journal of Financial Economics,* **8**, pp. 323-361.

[38] Merton, R., 1992, *Continuous-time Finance,* Blackwell Publishers, Cambridge, MA.

[39] Merville, L. and D. Pieptea, 1989, "Stock-price volatility, mean-reverting diffusion, and noise," *Journal of Financial Economics,* **24**, 193-214.

[40] Morris, S., 1994, "Trade with heterogeneous prior beliefs and asymmetric information," *Econometrica,* **62**, 1327-1347.

[41] Müller, U. A., M. M. Dacorogna, R. D. Davé, R. B. Olsen, O. V. Pictet, and J. E. Weizsắcker, 1997, "Volatilities of different time resolutions: Analyzing the dynamics of market components," *Journal of Empirical Finance,* **4**, 213-239.

[42] Nelson, D., 1990, "ARCH models as diffusion approximations," *Journal of Econometrics*, **45**, 7-38.

[43] Nelson, D., 1992, "Filtering and forecasting with mis-specified ARCH models I: Getting the right variance with the wrong model," *Journal of Econometrics*, **52**, 61-90.

[44] Nelson, D. and D. Foster (1994). "Asymptotic filtering theory for univariate ARCH models," *Econometrica,* **62**, 1-42.

[45] Pòlya, G., 1949, "Remarks on characteristic functions," *Proc. Berkeley Symp. Math. Stat. Prob.*, University of California Press, Berkeley, CA.

[46] Roll, R., 1984 "A simple implicit measure of the effective bid-ask spread in an efficient market," *Journal of Finance*, **39**, 1127-1140.

[47] Scott, L., 1989, "Option pricing when the variance changes randomly: Theory, estimation, and an application," *Journal of Financial and Quantitative Analysis,* **22**, pp. 419-438.

[48] Sørensen, M., 1991, "Likelihood methods for diffusions with jumps," *Statistical Inference in Stochastic Processes,* Prabhu, N. and I. Basawa (ed.), Marcel Dekker, Inc., New York.

[49] Wiggins, J., 1987, "Option values under stochastic volatility: Theory and empirical estimates," *Journal of Financial Economics,* **19**, pp. 351-372.

[50] Stein, E. M. and J. C. Stein, 1991, "Stock price distributions with stochastic volatility: An analytic approach," *The Review of Financial Studies,* **4**, 727-752.

[51] Yaglom, A., 1987, *Correlation Theory of Stationary and Related Random Functions,* Springer-Verlag, New York.

[52] Zhou, B., 1996, "High frequency data and volatility in foreign exchange rates," *Journal of Business & Economic Statistics*, **19**, 45-52.