CHAPTER 2

The Econometrics of High Frequency Data

Per. A. Mykland and Lan Zhang

Department of Statistics, University of Chicago
5734 University Avenue, Chicago, IL 60637, USA
and
Department of Finance, University of Illinois at Chicago
601 S Morgan Street, Chicago, IL 60607-7124, USA

2.1 Introduction

2.1.1 Overview

This is a course on estimation in high frequency data. It is intended for an audience that includes people interested in finance, econometrics, statistics, probability and financial engineering.

There has in recent years been a vast increase in the amount of high frequency data available. There has also been an explosion in the literature on the subject. In this course, we start from scratch, introducing the probabilistic model for such data, and then turn to the estimation question in this model. We shall be focused on the (for this area) emblematic problem of estimating volatility. Similar techniques to those we present can be applied to estimating leverage effects, realized regressions, semivariances, doing analyses of variance, detecting jumps, measuring liquidity by measuring the size of the microstructure noise, and many other objects of interest.

The applications are mainly in finance, ranging from risk management to options hedging (see Section 2.2.6 below), execution of transactions, portfolio optimization (Fleming, Kirby, and Ostdiek (2001; 2003)), and forecasting. The latter literature has been particularly active, with contributions including

Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold, and Labys (2001, 2003), Andersen, Bollerslev, and Meddahi (2005), Dacorogna, Gençay, Müller, Olsen, and Pictet (2001), and Meddahi (2001). Methodologies based on high frequency data can also be found in neural science (see, for example, Valdés-Sosa, Bornot-Sánchez, Melie-García, Lage-Castellanos, and Canales-Rodriguez (2007)) and climatology (see Ditlevsen, Ditlevsen and Andersen (2002) and Ditlevsen and Sørensen (2004) on Greenlandic ice cores).

The purpose of this article, however, is not so much to focus on the applications as on the probabilistic setting and the estimation methods. The theory was started, on the probabilistic side, by Jacod (1994) and Jacod and Protter (1998), and on the econometric side by Foster and Nelson (1996) and Comte and Renault (1998). The econometrics of integrated volatility was pioneered in Andersen et al. (2001, 2003), Barndorff-Nielsen and Shephard (2002, 2004b) and Dacorogna et al. (2001). The authors of this article started to work in the area through Zhang (2001), Zhang, Mykland, and Aït-Sahalia (2005), and Mykland and Zhang (2006). For further references, see Section 2.5.5.

Parametric estimation for discrete observations in a fixed time interval is also an active field. This problem has been studied by Genon-Catalot and Jacod (1994), Genon-Catalot, Jeantheau, and Larédo (1999; 2000), Gloter (2000), Gloter and Jacod (2001b, 2001a), Barndorff-Nielsen and Shephard (2001), Bibby, Jacobsen, and Sørensen (2009), Elerian, Siddhartha, and Shephard (2001), Jacobsen (2001), Sørensen (2001), and Hoffmann (2002). This is, of course, only a small sample of the literature available. Also, these references only concern the type of asymptotics considered in this paper, where the sampling interval is $[0, T]$. There is also a substantial literature on the case where $T \to \infty$ (see Section 1.2 of Mykland (2010b) for some of the main references in this area).

This article is meant to be a moderately self-contained course on the basics of this material. The introduction assumes some degree of statistics/econometric literacy, but at a lower level than the standard probability text. Some of the material is research front and not published elsewhere. This is not meant as a full review of the area. Readers with a good probabilistic background can skip most of Section 2.2, and occasional other sections.

The text also mostly overlooks (except Sections 2.3.5 and 2.6.3) the questions that arise in connection with multidimensional processes. For further literature in this area, one should consult Barndorff-Nielsen and Shephard (2004a), Hayashi and Yoshida (2005) and Zhang (2011).

2.1.2 High Frequency Data

Recent years have seen an explosion in the amount of financial high frequency data. These are the records of transactions and quotes for stocks, bonds, currencies, options, and other financial instruments.

A main source of such data is the *Trades and Quotes (TAQ)* database, which covers the stocks traded on the New York Stock Exchange (NYSE). For example, here is an excerpt of the transactions for Monday, April 4, 2005, for the pharmaceutical company Merck (MRK):

```
symbol date      time    price   size
MRK 20050405    9:41:37 32.69   100
MRK 20050405    9:41:42 32.68   100
MRK 20050405    9:41:43 32.69   300
MRK 20050405    9:41:44 32.68   1000
MRK 20050405    9:41:48 32.69   2900
MRK 20050405    9:41:48 32.68   200
MRK 20050405    9:41:48 32.68   200
MRK 20050405    9:41:51 32.68   4200
MRK 20050405    9:41:52 32.69   1000
MRK 20050405    9:41:53 32.68   300
MRK 20050405    9:41:57 32.69   200
MRK 20050405    9:42:03 32.67   2500
MRK 20050405    9:42:04 32.69   100
MRK 20050405    9:42:05 32.69   300
MRK 20050405    9:42:15 32.68   3500
MRK 20050405    9:42:17 32.69   800
MRK 20050405    9:42:17 32.68   500
MRK 20050405    9:42:17 32.68   300
MRK 20050405    9:42:17 32.68   100
MRK 20050405    9:42:20 32.69   6400
MRK 20050405    9:42:21 32.69   200
MRK 20050405    9:42:23 32.69   3000
MRK 20050405    9:42:27 32.70   8300
MRK 20050405    9:42:29 32.70   5000
MRK 20050405    9:42:29 32.70   1000
MRK 20050405    9:42:30 32.70   1100
```

“Size” here refers to the number of stocks that changed hands in the given transaction. This is often also called “volume”.

There are 6302 transactions recorded for Merck for this day. On the same day, Microsoft (MSFT) had 80982 transactions. These are massive amounts of data, and they keep growing. Four years later, on April 3, 2009, there were 74637

Merck transactions, and 211577 Microsoft transactions. What can we do with such data? This course is about how to approach this question.

2.1.3 A First Model for Financial Data: The GBM

Finance theory suggests the following description of prices, that they must be so-called *semimartingales*. We defer a discussion of the general concept until later (see also Delbaen and Schachermayer (1995)), and go instead to the most commonly used such semimartingale: the *Geometric Brownian Motion (GBM)*. This is a model where the stock price movement is additive on the log scale, as follows.

Set

$$X_t = \log S_t = \text{ the logarithm of the stock price } S_t \text{ at time } t. \tag{2.1}$$

The GBM model is now that

$$X_t = X_0 + \mu t + \sigma W_t, \tag{2.2}$$

where μ and σ are constants, and W_t is a *Brownian Motion (BM)*, a concept we now define. The "time zero" is an arbitrary reference time.

Definition 2.1 *The process $(W_t)_{0 \le t \le T}$ is a Brownian motion provided*
(1) $W_0 = 0$;
(2) $t \to W_t$ is a continuous function of t;
(3) W has independent increments: if $t > s > u > v$, then $W_t - W_s$ is independent of $W_u - W_v$;
(4) for $t > s$, $W_t - W_s$ is normal with mean zero and variance $t - s$ (N(0,t-s)).

2.1.4 Estimation in the GBM model

It is instructive to consider estimation in this model. We take time $t = 0$ to be the beginning of the trading day, and time $t = T$ to be the end of the day.

Let's assume that there are n observations of the process (transactions). We suppose for right now that the transactions are spaced equally in time, so that an observation is had every $\Delta t_n = T/n$ units of time. This assumption is quite unrealistic, but it helps a straightforward development which can then be modified later.

The observations (log transaction prices) are therefore $X_{t_{n,i}}$, where $t_{n,i} = i\Delta t_n$. If we take differences, we get observations

$$\Delta X_{t_{n,i+1}} = X_{t_{n,i+1}} - X_{t_{n,i}},\ i = 0, ..., n-1.$$

The $\Delta X_{t_{n,i+1}}$ are independent and identically distributed (iid) with law $\mathrm{N}(\mu\Delta \mathrm{t_n}, \sigma^2 \Delta \mathrm{t_n})$. The natural estimators are:

$$\hat{\mu}_n = \frac{1}{n\Delta t_n} \sum_{i=0}^{n-1} \Delta X_{t_{n,i+1}} = (X_T - X_0)/T \text{ both MLE and UMVU;}$$

and

$$\hat{\sigma}^2_{n,MLE} = \frac{1}{n\Delta t_n} \sum_{i=0}^{n-1} (\Delta X_{t_{n,i+1}} - \overline{\Delta X}_{t_n})^2 \text{ MLE; or} \tag{2.3}$$

$$\hat{\sigma}^2_{n,UMVU} = \frac{1}{(n-1)\Delta t_n} \sum_{i=0}^{n-1} (\Delta X_{t_{n,i+1}} - \overline{\Delta X}_{t_n})^2 \text{ UMVU.}$$

Here, MLE is the maximum likelihood estimator, and UMVU is the uniformly minimum variance unbiased estimator (see Lehmann (1983) or Rice (2006)). Also, $\overline{\Delta X}_{t_n} = \frac{1}{n}\sum_{i=0}^{n-1} \Delta X_{t_{n,i+1}} = \hat{\mu}_n \Delta t_n$.

The estimators (2.3) clarify some basics. First of all, μ *cannot be consistently estimated* for fixed length T of time interval. In fact, the $\hat{\mu}_n$ does not depend on n, but only on T and the value of the process at the beginning and end of the time period. This is reassuring from a common sense perspective. If we could estimate μ for actual stock prices, we would know much more about the stock market than we really do, and in the event that μ changes over time, benefit from a better allocation between stocks and bonds. – Of course, if $T \to \infty$, then μ *can* be estimated consistently. Specifically, $(X_T - X_0)/T \xrightarrow{p} \mu$ as $T \to \infty$. This is because $\mathrm{Var}((X_T - X_0)/T) = \sigma^2/T \to 0$.

It is perhaps more surprising that σ^2 *can* be estimated consistently for fixed T, as $n \to \infty$. In other words, $\hat{\sigma}^2_n \xrightarrow{p} \sigma^2$ as $n \to \infty$. Set $U_{n,i} = \Delta X_{t_{n,i}}/(\sigma \Delta t_n^{1/2})$. Then the $U_{n,i}$ are iid with distribution $\mathrm{N}((\mu/\sigma)\Delta \mathrm{t_n^{1/2}}, 1)$. Set $\bar{U}_{n,\cdot} = n^{-1}\sum_{i=0}^{n-1} U_{n,i}$. It follows from considerations for normal random variables (Cochran (1934)) that

$$\sum_{i=0}^{n-1} (U_{n,i} - \bar{U}_{n,\cdot})^2$$

is χ^2 distributed with $n-1$ degrees of freedom (and independent of $\bar{U}_{n,\cdot}$). Hence, for the UMVU estimator,

$$\begin{aligned} \hat{\sigma}^2_n &= \sigma^2 \Delta t_n \frac{1}{(n-1)\Delta t_n} \sum_{i=0}^{n-1} (U_{n,i} - \bar{U}_{n,\cdot})^2 \\ &\stackrel{\mathcal{L}}{=} \sigma^2 \frac{\chi^2_{n-1}}{n-1}. \end{aligned}$$

It follows that

$$E(\hat{\sigma}_n^2) = \sigma^2 \text{ and } \mathrm{Var}(\hat{\sigma}_\mathrm{n}^2) = \frac{2\sigma^4}{\mathrm{n}-1},$$

since $E\chi_m^2 = m$ and $\mathrm{Var}(\chi_\mathrm{m}^2) = 2\mathrm{m}$. Hence $\hat{\sigma}_n^2$ is consistent for σ^2: $\hat{\sigma}_n^2 \to \sigma^2$ in probability as $n \to \infty$.

Similarly, since χ_{n-1}^2 is the sum of $n-1$ iid χ_1^2 random variables, by the central limit theorem we have the following convergence in law:

$$\frac{\chi_{n-1}^2 - E\chi_{n-1}^2}{\sqrt{\mathrm{Var}(\chi_\mathrm{n-1}^2)}} = \frac{\chi_{n-1}^2 - (n-1)}{\sqrt{2(n-1)}} \xrightarrow{\mathcal{L}} \mathrm{N}(0,1),$$

and so

$$\begin{aligned} n^{1/2}(\hat{\sigma}_n^2 - \sigma^2) &\sim (n-1)^{1/2}(\hat{\sigma}_n^2 - \sigma^2) \\ &\stackrel{\mathcal{L}}{=} \sqrt{2}\sigma^2 \frac{\chi_{n-1}^2 - (n-1)}{\sqrt{2(n-1)}} \\ &\xrightarrow{\mathcal{L}} \sigma^2 N(0,2) = N(0, 2\sigma^4). \end{aligned}$$

This provides an asymptotic distribution which permits the setting of intervals. For example, $\sigma^2 = \hat{\sigma}_n^2 \pm 1.96 \times \sqrt{2}\hat{\sigma}_n^2$ would be an asymptotic 95 % confidence interval for σ^2.

Since $\hat{\sigma}_{n,MLE}^2 = \frac{n-1}{n}\hat{\sigma}_{n,UMVU}^2$, the same asymptotics apply to the MLE.

2.1.5 Behavior of Non-Centered Estimators

The above discussion of $\hat{\sigma}_{n,UMVU}^2$ and $\hat{\sigma}_{n,MLE}^2$ is exactly the same as in the classical case of estimating variance on the basis of iid observations. More unusually, for high frequency data, the mean is often not removed in estimation. The reason is as follows. Set

$$\hat{\sigma}_{n,nocenter}^2 = \frac{1}{n\Delta t_n}\sum_{i=0}^{n-1}(\Delta X_{t_{n,i+1}})^2.$$

Now note that for the MLE version of $\hat{\sigma}_n$,

$$\begin{aligned}
\hat{\sigma}^2_{n,MLE} &= \frac{1}{n\Delta t_n}\sum_{i=0}^{n-1}(\Delta X_{t_{n,i+1}} - \overline{\Delta X}_{t_n})^2 \\
&= \frac{1}{n\Delta t_n}\left(\sum_{i=0}^{n-1}(\Delta X_{t_{n,i+1}})^2 - n(\overline{\Delta X}_{t_n})^2\right) \\
&= \hat{\sigma}^2_{n,nocenter} - \Delta t_n \hat{\mu}^2_n \\
&= \hat{\sigma}^2_{n,nocenter} - \frac{T}{n}\hat{\mu}^2_n.
\end{aligned}$$

Since $\hat{\mu}^2_n$ does not depend on n, it follows that

$$n^{1/2}\left(\hat{\sigma}^2_{n,MLE} - \hat{\sigma}^2_{n,nocenter}\right) \xrightarrow{p} 0.$$

Hence, $\hat{\sigma}^2_{n,nocenter}$ is consistent and has the same asymptotic distribution as $\hat{\sigma}^2_{n,UMVU}$ and $\hat{\sigma}^2_{n,MLE}$. It can therefore also be used to estimate variance. This is quite common for high frequency data.

2.1.6 GBM and the Black-Scholes-Merton formula

The GBM model is closely tied in to other parts of finance. In particular, following the work of Black and Scholes (1973), Merton (1973), Harrison and Kreps (1979), and Harrison and Pliska (1981), precise option prices can be calculated in this model. See also Duffie (1996), Neftci (2000), Øksendal (2003), or Shreve (2004) for book sized introductions to the theory.

In the case of the *call option*, the price is as follows. A *European call option* on stock S_t with *maturity (expiration) time* T and *strike price* K is the option to buy one unit of stock at price K at time T. It is easy to see that the value of this option at time T is $(S_T - K)^+$, where $x^+ = x$ if $x \geq 0$, and $x^+ = 0$ otherwise.

If we make the assumption that S_t is a GBM, which is to say that it follows (2.1)-(2.2), and also the assumption that the short term interest rate r is constant (in time), then the price at time t, $0 \leq t \leq T$ of this option must be

$$\text{price } = C(S_t, \sigma^2(T-t), r(T-t)),$$

where

$$\begin{aligned}
C(S,\Xi,R) &= S\Phi(d_1(S,\Xi,R)) - K\exp(-R)\Phi(d_2(S,\Xi,R)), \text{ where} \\
d_{1,2}(S,\Xi,R) &= (\log(S/K) + R \pm \Xi/2)/\sqrt{\Xi}\ (+ \text{ in } d_1 \text{ and } - \text{ in } d_2) \text{ and} \qquad (2.4) \\
\Phi(x) &= P(N(0,1) \leq x), \text{ the standard normal cdf.}
\end{aligned}$$

This is the Black-Scholes-Merton formula.

We shall see later on how high frequency estimates can be used in this formula. For the moment, note that the price only depends on quantities that are either observed (the interest rate r) or (perhaps) nearly so (the volatility σ^2). It does not depend on μ. Unfortunately, the assumption of constant r and σ^2 is unrealistic, as we shall discuss in the following.

The GBM model is also heavily used in portfolio optimization

2.1.7 Our Problem to be Solved: Inadequacies in the GBM Model

We here give a laundry list of questions that arise and have to be dealt with.

The Volatility Depends on t

It is empirically the case that σ^2 depends on t. We shall talk about the *instantaneous volatility* σ_t^2. This concept will be defined carefully in Section 2.2.

Non-Normal Returns

Returns are usually assumed to be non-normal. This behavior can be explained through random volatility and/or jumps.

- *The Volatility is Random; Leverage Effect.* Non-normality can be achieved in a continuous model by letting σ_t^2 have random evolution. It is also usually assumed that σ_t^2 can be correlated with the (log) stock price. This is often referred to as *Leverage Effect*. More about this in Section 2.2.
- *Jumps.* The GBM model assumes that the log stock price X_t is continuous as a function of t. The evolution of the stock price, however, is often thought to have a jump component. The treatment of jumps is largely not covered in this article, though there is some discussion in Section 2.6.4, which also gives some references.

Jumps and random volatility are often confounded, since any martingale can be embedded in a Brownian motion (Dambis (1965), Dubins and Schwartz (1965), see also Mykland (1995) for a review and further discussion). The difficulty in distinguishing these two sources of non-normality is also studied by Bibby, Skovgaard, and Sørensen (2005).

Microstructure Noise

An important feature of actual transaction prices is the existence of *microstructure noise*. Transaction prices, as actually observed, are typically best modeled on the form $Y_t = \log S_t =$ the logarithm of the stock price S_t at time t, where for transaction at time t_i,

$$Y_{t_i} = X_{t_i} + \text{ noise},$$

and X_t is a semimartingale. This is often called the *hidden semimartingale model*. This issue is an important part of our narrative, and is further discussed in Section 2.5, see also Section 2.6.4.

Unequally Spaced Observations

In the above, we assumed that the transaction times t_i are equally spaced. A quick glance at the data snippet in Section 2.1.2 reveal that this is typically not the case. This leads to questions that will be addressed as we go along.

2.1.8 A Note on Probability Theory, and other Supporting Material

We will extensively use probability theory in these notes. To avoid making a long introduction on stochastic processes, we will define concepts as we need them, but not always in the greatest depth. We will also omit other concepts and many basic proofs. As a compromise between the rigorous and the intuitive, we follow the following convention: the notes will (except when the opposite is clearly stated) use mathematical terms as they are defined in Jacod and Shiryaev (2003). Thus, in case of doubt, this work can be consulted.

Other recommended reference books on stochastic process theory are Karatzas and Shreve (1991), Øksendal (2003), Protter (2004), and Shreve (2004). For introduction to measure theoretic probability, one can consult Billingsley (1995). Mardia, Kent, and Bibby (1979) provides a handy reference on normal distribution theory.

2.2 A More General Model: Time Varying Drift and Volatility

2.2.1 Stochastic Integrals, Itô-Processes

We here make some basic definitions. We consider a process X_t, where the time variable $t \in [0, T]$. We mainly develop the univariate case here.

Information Sets, σ-fields, Filtrations

Information is usually described with so-called σ-fields. The setup is as follows. Our basic space is $(\Omega, \mathcal{F})$, where Ω is the set of all possible outcomes ω, and $\mathcal{F}$ is the collection of subsets $A \subseteq \Omega$ that will eventually be decidable (it will be observed whether they occured or not). All random variables are thought to be a function of the basic outcome $\omega \in \Omega$.

We assume that $\mathcal{F}$ is a so-called *σ-field*. In general,

Definition 2.2 *A collection $\mathcal{A}$ of subsets of Ω is a σ-field if*

(i) $\emptyset$, $\Omega \in \mathcal{A}$;
(ii) if $A \in \mathcal{A}$, then $A^c = \Omega - A \in \mathcal{A}$; and
(iii) if $A_n, n = 1, 2, ...$ are all in $\mathcal{A}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{A}$.

If one thinks of $\mathcal{A}$ as a collection of decidable sets, then the interpretation of this definition is as follows:

(i) $\emptyset$, Ω are decidable ($\emptyset$ didn't occur, Ω did);
(ii) if A is decidable, so is the complement A^c (if A occurs, then A^c does not occur, and *vice versa*);
(iii) if all the A_n are decidable, then so is the event $\cup_{n=1}^{\infty} A_n$ (the union occurs if and only if at least one of the A_i occurs).

A random variable X is called $\mathcal{A}$-*measurable* if the value of X can be decided on the basis of the information in $\mathcal{A}$. Formally, the requirement is that for all x, the set $\{X \leq x\} = \{\omega \in \Omega \ : \ X(\omega) \leq x\}$ be decidable ($\in \mathcal{A}$).

The evolution of knowledge in our system is described by the *filtration* (or sequence of σ-fields) $\mathcal{F}_t$, $0 \leq t \leq T$. Here $\mathcal{F}_t$ is the knowledge available at time t. Since increasing time makes more sets decidable, the family $(\mathcal{F}_t)$ is taken to satisfy that if $s \leq t$, then $\mathcal{F}_s \subseteq \mathcal{F}_t$.

Most processes will be taken to be *adapted* to $(\mathcal{F}_t)$: (X_t) is adapted to $(\mathcal{F}_t)$ if for all $t \in [0, T]$, X_t is $\mathcal{F}_t$-measurable. A vector process is adapted if each component is adapted.

We define the filtration $(\mathcal{F}_t^X)$ *generated* by the process (X_t) as the smallest filtration to which X_t is adapted. By this we mean that for any filtration $\mathcal{F}'_t$ to which (X_t) is adapted, $\mathcal{F}_t^X \subseteq \mathcal{F}'_t$ for all t. (Proving the existence of such a filtration is left as an exercise for the reader).

Wiener Processes

A Wiener process is Brownian motion relative to a filtration. Specifically,

Definition 2.3 *The process* $(W_t)_{0 \leq t \leq T}$ *is an* $(\mathcal{F}_t)$*-Wiener process if it is adpted to* $(\mathcal{F}_t)$ *and*
(1) $W_0 = 0$*;*
(2) $t \to W_t$ *is a continuous function of* t*;*
(3) W *has independent increments relative to the filtration* $(\mathcal{F}_t)$*: if* $t > s$*, then* $W_t - W_s$ *is independent of* $\mathcal{F}_s$*;*
(4) for $t > s$*,* $W_t - W_s$ *is normal with mean zero and variance* $t - s$ *(N(0,t-s)).*

Note that a Brownian motion (W_t) is an $(\mathcal{F}_t^W)$-Wiener process.

Predictable Processes

For defining stochastic integrals, we need the concept of *predictable process*. "Predictable" here means that one can forecast the value over infinitesimal time intervals. The most basic example would be a "simple process". This is given by considering break points $0 = s_0 = t_0 \leq s_1 < t_1 \leq s_2 < t_2 < ... \leq s_n < t_n \leq T$, and random variables $H^{(i)}$, observable (measurable) with respect to $\mathcal{F}_{s_i}$.

$$H_t = \begin{cases} H^{(0)} \text{ if } t = 0 \\ H^{(i)} \text{ if } s_i < t \leq t_i \end{cases} \tag{2.5}$$

In this case, at any time t (the beginning time $t = 0$ is treated separately), the value of H_t is known *before* time t.

Definition 2.4 *More generally, a process* H_t *is predictable if it can be written as a limit of simple functions* $H_t^{(n)}$*. This means that* $H_t^{(n)}(\omega) \to H_t(\omega)$ *as* $n \to \infty$*, for all* $(t, \omega) \in [0, T] \times \Omega$*.*

All adapted continuous processes are predictable. More generally, this is also true for adapted processes that are left continuous (*càg*, for *continue à gauche*). (Proposition I.2.6 (p. 17) in Jacod and Shiryaev (2003)).

Stochastic Integrals

We here consider the meaning of the expression

$$\int_0^T H_t dX_t. \tag{2.6}$$

The ingredients are the integrand H_t, which is assumed to be predictable, and the integrator X_t, which will generally be a semi-martingale (to be defined below in Section 2.2.3).

The expression (2.6) is defined for simple process integrands as

$$\sum_i H^{(i)}(X_{t_i} - X_{s_i}). \tag{2.7}$$

For predictable integrands H_t that are bounded and limits of simple processes $H_t^{(n)}$, the integral (2.6) is the limit in probability of $\int_0^T H_t^{(n)} dX_t$. This limit is well defined, *i.e.,* independent of the sequence $H_t^{(n)}$.

If X_t is a Wiener process, the integral can be defined for any predictable process H_t satisfying

$$\int_0^T H_t^2 dt \ < \ \infty.$$

It will always be the case that the integrator X_t is right continuous with left limits (*càdlàg*, for *continue à droite, limites à gauche*).

The integral process

$$\int_0^t H_s dX_s = \int_0^T H_s I\{s \le t\} dX_s$$

can also be taken to be *càdlàg*. If (X_t) is continuous, the integral is then automatically continuous.

Itô Processes

We now come to our main model, the Itô process. X_t is an Itô process relative to filtration $(\mathcal{F}_t)$ provided (X_t) is $(\mathcal{F}_t)$ adapted; and if there is an $(\mathcal{F}_t)$-Wiener process (W_t), and $(\mathcal{F}_t)$-adapted processes (μ_t) and (σ_t), with

$$\int_0^T |\mu_t| dt < \infty, \text{ and}$$
$$\int_0^T \sigma_t^2 dt < \infty$$

so that

$$X_t \ = \ X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s. \tag{2.8}$$

The process is often written on differential form:

$$dX_t \ = \ \mu_t dt + \sigma_t dW_t. \tag{2.9}$$

We note that the Itô process property is preserved under stochastic integration. If H_t is bounded and predictable, then

$$\int_0^t H_s dX_s = \int_0^t H_s \mu_s dt + \int_0^t H_s \sigma_s dW_s.$$

It is clear from this formula that predictable processes H_t can be used for integration w.r.t. X_t provided

$$\int_0^T |H_t\mu_t|dt < \infty \text{ and}$$
$$\int_0^T (H_t\sigma_t)^2 dt < \infty.$$

2.2.2 Two Interpretations of the Stochastic Integral

One can use the stochastic integral in two different ways: as model, or as a description of trading profit and loss (P/L).

Stochastic Integral as Trading Profit or Loss (P/L)

Suppose that X_t is the value of a security. Let H_t be the number of this stock that is held at time t. In the case of a simple process (2.5), this means that we hold $H^{(i)}$ units of X from time s_i to time t_i. The trading profit and loss (P/L) is then given by the stochastic integral (2.7). In this description, it is quite clear that $H^{(i)}$ must be known at time s_i, otherwise we would base the portfolio on future information. More generally, for predictable H_t, we similarly avoid using future information.

Stochastic Integral as Model

This is a different genesis of the stochastic integral model. One simply uses (2.8) as a model, in the hope that this is a sufficiently general framework to capture most relevant processes. The advantage of using predictable integrands come from the simplicity of connecting the model with trading gains.

For simple μ_t and σ_t^2, the integral

$$\sum_i \mu^{(i)}(t_i - s_i) + \sum_i \sigma^{(i)}(W_{t_i} - W_{s_i})$$

is simply a sum of contitionally normal random variables, with mean $\mu^{(i)}(t_i - s_i)$ and variance $(\sigma^{(i)})^2(t_i - s_i)$. The sum need not be normal, since μ and σ^2 can be random.

It is worth noting that in this model, $\int_0^T \mu_t dt$ is the sum of instantaneous means (drift), and $\int_0^T \sigma_t^2 dt$ is the sum of intstantaneous variances. To make the latter statement precise, note that in the model (2.8), one can show the following: Let $\mathrm{Var}(\cdot|\mathcal{F}_t)$ be the conditional variance given the information at time t. If X_t is an Itô process, and if $0 = t_{n,0} < t_{n,i} < ... < t_{n,n} = T$, then

$$\sum_i \mathrm{Var}(X_{t_{n,i+1}} - X_{t_{n,i}} | \mathcal{F}_{t_{n,i}}) \xrightarrow{p} \int_0^T \sigma_t^2 dt \tag{2.10}$$

when

$$\max_i |t_{n,i+1} - t_{n,i}| \ \rightarrow \ 0.$$

If the μ_t and σ_t^2 processes are nonrandom, then X_t is a Gaussian process, and X_T is normal with mean $X_0 + \int_0^T \mu_t dt$ and variance $\int_0^T \sigma_t^2 dt$.

The Heston model

A popular model for volatility is due to Heston (1993). In this model, the process X_t is given by

$$\begin{aligned} dX_t &= \left(\mu - \frac{\sigma_t^2}{2}\right) dt + \sigma_t dW_t \\ d\sigma_t^2 &= \kappa(\alpha - \sigma_t^2)dt + \gamma\sigma_t dZ_t \ , \ \text{with} \\ Z_t &= \rho W_t + (1-\rho^2)^{1/2} B_t \end{aligned}$$

where (W_t) and (B_t) are two independent Wiener processes, $\kappa > 0$, and $|\rho| \leq 1$. To assure that σ_t^2 does not hit zero, one must also require (Feller (1951)) that $2\kappa\alpha \geq \gamma^2$.

2.2.3 Semimartingales

Conditional Expectations

Denote by $E(\cdot|\mathcal{F}_t)$ the conditional expectation given the information available at time t. Formally, this concept is defined as follows:

Theorem 2.5 *Let $\mathcal{A}$ be a σ-field, and let X be a random variable so that $E|X| < \infty$. There is a $\mathcal{A}$-measurable random variable Z so that for all $A \in \mathcal{A}$,*

$$EZI_A = EXI_A,$$

where I_A is the indicator function of A. Z is unique "almost surely", which means that if Z_1 and Z_2 satisfy the two criteria above, then $P(Z_1 = Z_2) = 1$.

We thus define

$$E(X|\mathcal{A}) \ = \ Z$$

where Z is given in the theorem. The conditional expectation is well defined "almost surely".

For further details and proof of theorem, see Section 34 (p. 445-455) of Billingsley (1995).

This way of defining conditional expectation is a little counterintuitive if unfamiliar. In particular, the conditional expectation is a random variable. The heuristic is as follows. Suppose that Y is a random variable, and that $\mathcal{A}$ carries the information in Y. Introductory textbooks often introduce conditional expectation as a non-random quantity $E(X|Y=y)$. To make the connection, set

$$f(y) = E(X|Y=y).$$

The conditional expectation we have just defined then satisfies

$$E(X|\mathcal{A}) = f(Y). \tag{2.11}$$

The expression in (2.11) is often written $E(X|Y)$.

Properties of Conditional Expectations

- Linearity: for constant c_1, c_2:

$$E(c_1X_1 + c_2X_2 \mid \mathcal{A}) = c_1E(X_1 \mid \mathcal{A}) + c_2E(X_2 \mid \mathcal{A})$$

- Conditional constants: if Z is $\mathcal{A}$-measurable, then

$$E(ZX|\mathcal{A}) = ZE(X|\mathcal{A})$$

- Law of iterated expectations (iterated conditioning, Tower property): if $\mathcal{A}' \subseteq \mathcal{A}$, then

$$E[E(X|\mathcal{A})|\mathcal{A}'] \;=\; E(X|\mathcal{A}')$$

- Independence: if X is independent of $\mathcal{A}$:

$$E(X|\mathcal{A}) = E(X)$$

- Jensen's inequality: if $g: \; x \to g(x)$ is convex:

$$E(g(X)|\mathcal{A}) \geq g(E(X|\mathcal{A}))$$

Note: g is convex if $g(ax + (1-a)y) \leq ag(x) + (1-a)g(y)$ for $0 \leq a \leq 1$.

For example: $g(x) = e^x$, $g(x) = (x-K)^+$. Or g'' exists and is continuous, and $g''(x) \geq 0$.

Martingales

An $(\mathcal{F}_t)$ adapted process M_t is called a *martingale* if $E|M_t| < \infty$, and if, for all $s < t$,

$$E(M_t|\mathcal{F}_s) = M_s.$$

This is a central concept in our narrative. A martingale is also known as a *fair game*, for the following reason. In a gambling situation, if M_s is the amount of money the gambler has at time s, then the gambler's expected wealth at time $t > s$ is also M_s. (The concept of martingale applies equally to discrete and continuous time axis).

Example 2.6 *A Wiener process is a martingale. To see this, for $t > s$, since $W_t - W_s$ is N(0,t-s) given $\mathcal{F}_s$, we get that*

$$\begin{aligned} E(W_t|\mathcal{F}_s) &= E(W_t - W_s|\mathcal{F}_s) + W_s \\ &= E(W_t - W_s) + W_s \textit{ by independence} \\ &= W_s. \end{aligned}$$

A useful fact about martingales is the *representation by final value*: M_t is a martingale for $0 \le t \le T$ if and only if one can write (with $E|X| < \infty$)

$$M_t = E(X|\mathcal{F}_t) \text{ for all } t \in [0, T]$$

(only if by definition ($X = M_T$), if by Tower property). Note that for $T = \infty$ (which we do not consider here), this property may not hold. (For a full discussion, see Chapter 1.3.B (p. 17-19) of Karatzas and Shreve (1991)).

Example 2.7 *If H_t is a bounded predictable process, then for any martingale X_t,*

$$M_t = \int_0^t H_s dX_s$$

is a martingale. To see this, consider first a simple process (2.5), for which $H_s = H^{(i)}$ when $s_i < s \le t_i$. For given t, if $s_i > t$, by the properties of conditional expectations,

$$\begin{aligned} E\left(H^{(i)}(X_{t_i} - X_{s_i})|\mathcal{F}_t\right) &= E\left(E(H^{(i)}(X_{t_i} - X_{s_i})|\mathcal{F}_{s_i})|\mathcal{F}_t\right) \\ &= E\left(H^{(i)}E(X_{t_i} - X_{s_i}|\mathcal{F}_{s_i})|\mathcal{F}_t\right) \\ &= 0, \end{aligned}$$

and similarly, if $s_i \le t \le t_i$, then

$$E\left(H^{(i)}(X_{t_i} - X_{s_i})|\mathcal{F}_t\right) = H^{(i)}(X_t - X_{s_i})$$

so that

$$\begin{aligned} E(M_T|\mathcal{F}_t) &= E\left(\sum_i H^{(i)}(X_{t_i} - X_{s_i})|\mathcal{F}_t\right) \\ &= \sum_{i:t_i<t} H^{(i)}(X_{t_i} - X_{s_i}) + I\{t_i \le t \le s_i\}H^{(i)}(X_t - X_{s_i}) \\ &= M_t. \end{aligned}$$

The result follows for general bounded predicable integrands by taking limits and using uniform integrability. (For definition and results on uniform integrability, see Billingsley (1995).)

Thus, any bounded trading strategy H in an asset M which is a martingale results in a martingale profit and loss (P/L).

Stopping Times and Local Martingales

The concept of local martingale is perhaps best understood by considering the following integral with respect to a Wiener process (see also Duffie (1996)):

$$X_t = \int_0^t \frac{1}{\sqrt{T-s}} dW_s$$

Note that for $0 \leq t < T$, X_t is a zero mean Gaussian process with independent increments. We shall show below (in Section 2.2.4) that the integral has variance

$$\begin{aligned} \mathrm{Var(X_t)} &= \int_0^t \frac{1}{T-s} ds \\ &= \int_{T-t}^T \frac{1}{u} du \\ &= \log \frac{T}{T-t}. \end{aligned} \tag{2.12}$$

Since the dispersion of X_t goes to infinity as we approach T, X_t is not defined at T. However, one can *stop* the process at a convenient time, as follows: Set, for $A > 0$,

$$\tau = \inf\{t \geq 0 : \ X_t = A\}. \tag{2.13}$$

One can show that $P(\tau < T) = 1$. Define the modified integral by

$$\begin{aligned} Y_t &= \int_0^t \frac{1}{\sqrt{T-s}} I\{s \leq \tau\} dW_s \\ &= X_{\tau \wedge t}, \end{aligned} \tag{2.14}$$

where

$$s \wedge t = \min(s, t).$$

The process (2.14) has the following trading interpretation. Suppose that W_t is the value of a security at time t (the value can be negative, but that is possible for many securities, such as futures contracts). We also take the short term interest rate to be zero. The process X_t comes about as the value of a portfolio which holds $1/\sqrt{T-t}$ units of this security at time t. The process Y_t is obtained by holding this portfolio until such time that $X_t = A$, and then liquidating the portfolio.

In other words, we have displayed a trading strategy which starts with wealth $Y_0 = 0$ at time $t = 0$, and ends with wealth $Y_T = A > 0$ at time $t = T$. In trading terms, this is an arbitrage. In mathematical terms, this is a stochastic integral w.r.t. a martingale which is no longer a martingale.

We note that from (2.12), the conditions for the existence of the integral (2.14) are satisfied.

For trading, the lesson we can learn from this is that some condition has to be imposed to make sure that a trading strategy in a martingale cannot result in arbitrage profit. The most popular approach to this is to require that the traders wealth at any time cannot go below some fixed amount $-K$. This is the so-called credit constraint. (So strategies are required to satisfy that the integral never goes below $-K$). This does not quite guarantee that the integral w.r.t. a martingale is a martingale, but it does prevent arbitrage profit. The technical result is that the integral is a *super-martingale* (see the next section).

For the purpose of characterizing the stochastic integral, we need the concept of a *local martingale.* For this, we first need to define:

Definition 2.8 *A stopping time is a random variable τ satisfying $\{\tau \leq t\} \in \mathcal{F}_t$, for all t.*

The requirement in this definition is that we must be able to know at time t wether τ occurred or not. The time (2.13) given above is a stopping time. On the other hand, the variable $\tau = \inf\{t : W_t = \max_{0 \leq s \leq T} W_s\}$ is not a stopping time. Otherwise, we would have a nice investment strategy.

Definition 2.9 *A process M_t is a local martingale for $0 \leq t \leq T$ provided there is a sequence of stopping times τ_n so that*
(i) $M_{\tau_n \wedge t}$ is a martingale for each n; and
(ii) $P(\tau_n \to T) = 1$ as $n \to \infty$.

The basic result for stochastic integrals is now that the integral with respect to a local martingale is a local martingale, cf. result I.4.34(b) (p. 47) in Jacod and Shiryaev (2003).

Semimartingales

X_t is a semimartingale if it can be written

$$X_t = X_0 + M_t + A_t, 0 \leq t \leq T,$$

where X_0 is $\mathcal{F}_0$-measurable, M_t is a local martingale, and A_t is a process of finite variation, *i.e.*,

$$\sup \sum_i |A_{t_{i+1}} - A_{t_i}| < \infty,$$

where the supremum is over all grids $0 = t_0 < t_1 < ... < t_n = T$, and all n.

In particular, an Itô process is a semimartingale, with

$$M_t = \int_0^t \sigma_s dW_s \text{ and}$$
$$A_t = \int_0^t \mu_s ds.$$

A *supermartingale* is semimartingale for which A_t is nonincreasing. A *submartingale* is a semimartingale for which A_t is nondecreasing.

2.2.4 Quadratic Variation of a Semimartingale

Definitions

We start with some notation. A grid of observation times is given by

$$\mathcal{G} = \{t_0, t_1, ..., t_n\},$$

where we suppose that

$$0 = t_0 < t_1 < ... < t_n = T.$$

Set

$$\Delta(\mathcal{G}) = \max_{1 \le i \le n} (t_i - t_{i-1}).$$

For any process X, we define its *quadratic variation* relative to grid $\mathcal{G}$ by

$$[X, X]_t^{\mathcal{G}} = \sum_{t_{i+1} \le t} (X_{t_{i+1}} - X_{t_i})^2. \tag{2.15}$$

We note that the quadratic variation is path-dependent. One can more generally define the quadratic covariation

$$[X, Y]_t^{\mathcal{G}} = \sum_{t_{i+1} \le t} (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i}).$$

An important theorem of stochastic calculus now says that

Theorem 2.10 *For any semimartingale, there is a process $[X, Y]_t$ so that*

$$[X, Y]_t^{\mathcal{G}} \xrightarrow{p} [X, Y]_t \textit{ for all } t \in [0, T], \textit{ as } \Delta(\mathcal{G}) \to 0.$$

The limit is independent of the sequence of grids $\mathcal{G}$.

The result follows from Theorem I.4.47 (p. 52) in Jacod and Shiryaev (2003). The t_i can even be stopping times.

For an Itô process,

$$[X,X]_t = \int_0^t \sigma_s^2 ds. \tag{2.16}$$

(Cf Thm I.4.52 (p. 55) and I.4.40(d) (p. 48) of Jacod and Shiryaev (2003)). In particular, for a Wiener process W, $[W,W]_t = \int_0^t 1 ds = t$.

The process $[X,X]_t$ is usually referred to as the quadratic variation of the semimartingale (X_t). This is an important concept, as seen in Section 2.2.2. The theorem asserts that this quantity can be estimated consistently from data.

Properties

Important properties are as follows:

(1) Bilinearity: $[X,Y]_t$ is linear in each of X and Y: so for example, $[aX + bZ, Y]_t = a[X,Y]_t + b[Z,Y]_t$.

(2) If (W_t) and (B_t) are two independent Wiener processes, then

$$[W,B]_t = 0.$$

Example 2.11 *For the Heston model in Section 2.2.2, one gets from first principles that*

$$\begin{aligned}[W,Z]_t &= \rho[W,W]_t + (1-\rho^2)^{1/2}[W,B]_t \\ &= \rho t,\end{aligned}$$

since $[W,W]_t = t$ *and* $[W,B]_t = 0$.

(3) For stochastic integrals over Itô processes X_t and Y_t,

$$U_t = \int_0^t H_s dX_s \text{ and } V_t = \int_0^t K_s dY_s,$$

then

$$[U,V]_t = \int_0^t H_s K_s d[X,Y]_s.$$

This is often written on "differential form" as

$$d[U,V]_t = H_t K_t d[X,Y]_t.$$

by invoking the same results that led to (2.16). For a rigorous statement, see Property I.4.54 (p.55) of Jacod and Shiryaev (2003).

(4) For any Itô process X, $[X,t] = 0$.

Example 2.12 *(Leverage Effect in the Heston model).*

$$\begin{aligned}d[X,\sigma^2] &= \gamma\sigma_t^2 d[W,Z]_t \\ &= \gamma\sigma^2\rho dt.\end{aligned}$$

(5) Invariance under discounting by the short term interest rate. Discounting is important in finance theory. The typical discount rate is the risk free short term interest rate r_t. Recall that $S_t = \exp\{X_t\}$. The discounted stock price is then given by

$$S_t^* = \exp\{-\int_0^t r_s ds\} S_t.$$

The corresponding process on the log scale is $X_t^* = X_t - \int_0^t r_s ds$, so that if X_t is given by (2.9), then

$$dX_t^* \ = \ (\mu_t - r_t)dt + \sigma_t dW_t.$$

The quadratic variation of X_t^* is therefore the same as for X_t.

It should be emphasized that while this result remains true for certain other types of discounting (such as those incorporating cost-of-carry), it is not true for many other relevant types of discounting. For example, if one discounts by the zero coupon bond Λ_t maturing at time T, the discounted log price becomes $X_t^* = X_t - \log \Lambda_t$. Since the zero coupon bond will itself have volatility, we get

$$[X^*, X^*]_t = [X, X]_t + [\log \Lambda, \log \Lambda]_t - 2[X, \log \Lambda]_t.$$

Variance and Quadratic Variation

Quadratic variation has a representation in terms of variance. The main result concerns martingales. For $E(X^2) < \infty$, define the conditional variance by

$$\text{Var}(X|\mathcal{A}) = E((X - E(X|\mathcal{A}))^2|\mathcal{A}) = E(X^2|\mathcal{A}) - E(X|\mathcal{A})^2$$

and similarly $\text{Cov}(X, Y|\mathcal{A}) = E((X - E(X|\mathcal{A}))(Y - E(Y|\mathcal{A})|A)$.

Theorem 2.13 *Let M_t be a martingale, and assume that $E[M, M]_T < \infty$. Then, for all $s < t$,*

$$\text{Var}(M_t|\mathcal{F}_s) \ = \ E((M_t - M_s)^2|\mathcal{F}_s) \ = \ E([M, M]_t - [M, M]_s|\mathcal{F}_s). \quad (2.17)$$

This theorem is the beginning of something important: the left hand side of (2.17) relates to the central limit theorem, while the right hand side only concerns the law of large numbers. We shall see this effect in more detail in the sequel.

A quick argument for (2.17) is as follows. Let $\mathcal{G} = \{t_0, t_1, ..., t_n\}$, and let $t_* = \max\{u \in \mathcal{G} : \ u \leq t\}$, and similarly for s_*. Suppose for simplicity that $s, t \in \mathcal{G}$. Then, for $s_* \leq t_i < t_j$,

$$\begin{aligned} & E((M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j})|\mathcal{F}_{t_j}) \\ = \ & (M_{t_{i+1}} - M_{t_i})E((M_{t_{j+1}} - M_{t_j})|\mathcal{F}_{t_j}) \\ = \ & 0, \end{aligned}$$

so that by the Tower rule (since $\mathcal{F}_{s_*} \subseteq \mathcal{F}_{t_j}$)

$$\begin{aligned}
&\text{Cov}(M_{t_{i+1}} - M_{t_i}, M_{t_{j+1}} - M_{t_j}|\mathcal{F}_{s_*}) \\
&= E((M_{t_{i+1}} - M_{t_i})(M_{t_{j+1}} - M_{t_j})|\mathcal{F}_{s_*}) \\
&= 0.
\end{aligned}$$

It follows that, for $s < t$,

$$\begin{aligned}
\text{Var}(M_{t_*} - M_{s_*}|\mathcal{F}_{s_*}) &= \sum_{s_* \le t_i < t_*} \text{Var}(M_{t_{i+1}} - M_{t_i}|\mathcal{F}_{s_*}) \\
&= \sum_{s_* \le t_i < t_*} E((M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_{s_*}) \\
&= E(\sum_{s_* \le t_i < t_*} (M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_{s_*}) \\
&= E([M, M]_{t_*}^{\mathcal{G}} - [M, M]_{s_*}^{\mathcal{G}}|\mathcal{F}_{s_*}).
\end{aligned}$$

The result as $\Delta(\mathcal{G}) \to 0$ then follows by uniform integrability (Theorem 25.12 (p. 338) in Billingsley (1995)).

On the basis of this, one can now show for an Itô process that

$$\lim_{h \downarrow 0} \frac{1}{h} \text{Cov}(\mathrm{X_{t+h} - X_t, Y_{t+h} - Y_t}|\mathcal{F}_\mathrm{t}) = \frac{\mathrm{d}}{\mathrm{dt}}[\mathrm{X, Y}]_\mathrm{t}.$$

A similar result holds in the integrated sense, cf. formula (2.10). The reason this works is that the dt terms are of smaller order than the martingale terms.

Sometimes instantaneous correlation is important. We define

$$\mathrm{cor(X, Y)_t} = \lim_{\mathrm{h} \downarrow 0} \mathrm{cor}(\mathrm{X_{t+h} - X_t, Y_{t+h} - Y_t}|\mathcal{F}_\mathrm{t}),$$

and note that

$$\mathrm{cor(X, Y)_t} = \frac{\mathrm{d[X, Y]_t/dt}}{\sqrt{\mathrm{(d[X, X]_t/dt)(d[Y, Y]_t/dt)}}}.$$

We emphasize that these results only hold for Itô processes. For general semimartingales, one needs to involve the concept of predictable quadratic variation, cf. Section 2.2.4.

To see the importance of the instantaneous correlation, note that in the Heston model,

$$\mathrm{cor}(\mathrm{X}, \sigma^2)_\mathrm{t} = \rho.$$

In general, if $dX_t = \sigma_t dW_t + dt$-term, and $dY_t = \gamma_t dB_t + dt$-term, where W_t and B_t are two Wiener processes, then

$$\mathrm{cor(X, Y)_t} = \mathrm{sgn}(\sigma_\mathrm{t}\gamma_\mathrm{t})\mathrm{cor(W, B)_t}. \qquad (2.18)$$

Lévy's Theorem

An important result is now the following:

Theorem 2.14 *Suppose that M_t is a continuous $(\mathcal{F}_t)$-local martingale, $M_0 = 0$, so that $[M, M]_t = t$. Then M_t is an $(\mathcal{F}_t)$-Wiener process.*

(Cf. Thm II.4.4 (p. 102) in Jacod and Shiryaev (2003)). More generally, from properties of normal random variables, the same result follows in the vector case: If $M_t = (M_t^{(1)}, ..., M_t^{(p)})$ is a continuous $(\mathcal{F}_t)$-martingale, $M_0 = 0$, so that $[M^{(i)}, M^{(j)}]_t = \delta_{ij} t$, then M_t is a vector Wiener process. (δ_{ij} is the Kronecker delta: $\delta_{ij} = 1$ for $i = j$, and $= 0$ otherwise.)

Predictable Quadratic Variation

One can often see the symbol $\langle X, Y \rangle_t$. This can be called the predictable quadratic vartiation. Under regularity conditions, it is defined as the limit of $\sum_{t_i \le t} \mathrm{Cov}(\mathrm{X_{t_{i+1}} - X_{t_i}, Y_{t_{i+1}} - Y_{t_i}} | \mathcal{F}_{\mathrm{t_i}})$ as $\Delta(\mathcal{G}) \to 0$.

For Itô processes, $\langle X, Y \rangle_t = [X, Y]_t$. For general semimartingales this equality does not hold. Also, except for Itô processes, $\langle X, Y \rangle_t$ cannot generally be estimated consistently from data without further assumptions. For example, If N_t is a Poisson process with intensity λ, then $M_t = N_t - \lambda t$ is a martingale. In this case, $[M, M]_t = N_t$ (observable), while $\langle M, M \rangle_t = \lambda t$ (cannot be estimated in finite time). For further discussion of such discontinuous processes, see the references mentioned in Section 2.1.8, and also, in the context of survival analysis, Andersen, Borgan, Gill, and Keiding (1992).

For continuous semimartingales, The symbol $\langle X, Y \rangle_t$ is commonly used in the literature in lieu of $[X, Y]_t$ (including in our papers).

2.2.5 Itô's Formula for Itô processes

Main Theorem

Theorem 2.15 *Suppose that f is a twice continuously differentiable function, and that X_t is an Itô process. Then*

$$df(X_t) = f'(X_t) dX_t + \frac{1}{2} f''(X_t) d[X, X]_t. \tag{2.19}$$

Similarly, in the multivariate case, for $X_t = (X_t^{(1)}, ..., X_t^{(p)})$,

$$df(X_t) = \sum_{i=1}^{p} \frac{\partial f}{\partial x^{(i)}}(X_t) dX_t^{(i)} + \frac{1}{2} \sum_{i,j=1}^{p} \frac{\partial^2 f}{\partial x^{(i)} \partial x^{(j)}}(X_t) d[X^{(i)}, X^{(j)}]_t.$$

(Reference: Theorem I.4.57 in Jacod and Shiryaev (2003).)

We emphasize that (2.19) is the same as saying that

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)dX_s + \frac{1}{2}\int_0^t f''(X_s)d[X,X]_s.$$

If we write out $dX_t = \mu_t dt + \sigma_t dW_t$ and $d[X,X]_t = \sigma_t^2 dt$, then equation (2.19) becomes

$$\begin{aligned} df(X_t) &= f'(X_t)(\mu_t dt + \sigma_t dW_t) + \frac{1}{2}f''(X_t)\sigma_t^2 dt \\ &= (f'(X_t)\mu_t + \frac{1}{2}f''(X_t)\sigma_t^2)dt + f'(X_t)\sigma_t dW_t. \end{aligned}$$

We note, in particular, that if X_t is an Itô process, then so is $f(X_t)$.

Example of Itô's Formula: Stochastic Equation for a Stock Price

We have so far discussed the model for a stock on the log scale, as $dX_t = \mu_t dt + \sigma_t dW_t$. The price is given as $S_t = \exp(X_t)$. Using Itô's formula, with $f(x) = \exp(x)$, we get

$$dS_t = S_t(\mu_t + \frac{1}{2}\sigma_t^2)dt + S_t\sigma_t dW_t. \qquad (2.20)$$

Example of Itô's Formula: Proof of Lévy's Theorem (Section 2.2.4)

Take $f(x) = e^{ihx}$, and go on from there. Left to the reader.

Example of Itô's Formula: Genesis of the Leverage Effect

We here see a case where quadratic covariation between a process and it's volatility can arise from basic economic principles. The following is the origin of the use of the word "leverage effect" to describe such covariation. We emphasize that this kind of covariation can arise from many considerations, and will later use the term leverage effect to describe the phenomenon broadly.

Suppose that the $\log$ value of a firm is Z_t, given as a GBM,

$$dZ_t = \nu dt + \gamma dW_t.$$

For simplicity, suppose that the interest rate is zero, and that the firm has borrowed C dollars (or euros, yuan, ...). If there are M shares in the company, the value of one share is therefore

$$S_t = (\exp(Z_t) - C)/M.$$

On the log scale, therefore, by Itô's Formula,

$$\begin{aligned} dX_t &= d\log(S_t) \\ &= \frac{1}{S_t}dS_t - \frac{1}{2}\frac{1}{S_t^2}d[S,S]_t \\ &= \frac{M}{\exp(Z_t)-C}dS_t - \frac{1}{2}\left(\frac{M}{\exp(Z_t)-C}\right)^2 d[S,S]_t. \end{aligned}$$

Since, in the same way as for (2.20)

$$\begin{aligned} dS_t &= \frac{1}{M}d\exp(Z_t) \\ &= \frac{1}{M}\exp(Z_t)[(\nu+\frac{1}{2}\gamma^2)dt+\gamma dW_t]. \end{aligned}$$

Hence, if we set

$$U_t = \frac{\exp(Z_t)}{\exp(Z_t)-C},$$

$$\begin{aligned} dX_t &= U_t[(\nu+\frac{1}{2}\gamma^2)dt+\gamma dW_t] - \frac{1}{2}U_t^2\gamma^2 dt \\ &= (\nu U_t + \frac{1}{2}\gamma^2(U_t-U_t^2))dt + U_t\gamma dW_t. \end{aligned}$$

In other words,

$$dX_t = \mu_t dt + \sigma_t dW_t$$

where

$$\begin{aligned} \mu_t &= \nu U_t + \frac{1}{2}\gamma^2(U_t-U_t^2) \text{ and} \\ \sigma_t &= U_t\gamma. \end{aligned}$$

In this case, the log stock price and the volatility are, indeed, correlated. When the stock price goes down, the volatility goes up (and the volatility will go to infinity if the value of the firm approaches the borrowed amount C, since in this case $U_t \to \infty$. In terms of quadratic variation, the leverage effect is given as

$$\begin{aligned} d[X,\sigma^2]_t &= U_t\gamma^3 d[W,U^2]_t \\ &= 2U_t^2\gamma^3 d[W,U]_t \text{ since } dU_t^2 = 2U_t dU_t + d[U,U]_t \\ &= -2U_t^4\gamma^4 C\exp(-Z_t)dt. \end{aligned}$$

The last transition follows since, by taking $f(x) = (1-C\exp(-x))^{-1}$

$$\begin{aligned} dU_t &= df(Z_t) \\ &= f'(Z_t)dZ_t + dt\text{-terms} \end{aligned}$$

so that

$$\begin{aligned} d[W,U]_t &= f'(Z_t)d[W,Z]_t \\ &= f'(Z_t)\gamma dt \\ &= -U_t^2 C\exp(-Z_t)\gamma dt, \end{aligned}$$

since $f'(x) = -f(x)^2 C\exp(-x)$.

A perhaps more intuitive result is obtained from (2.18), by observing that $\mathrm{sgn}(\mathrm{d}[\mathrm{X},\sigma^2]_\mathrm{t}/\mathrm{dt}) = -1$: on the correlation scale, the leverage effect is

$$\mathrm{cor}(\mathrm{X},\sigma^2)_\mathrm{t} = -1.$$

2.2.6 Nonparametric Hedging of Options

Suppose we can set the following prediction intervals at time $t = 0$:

$$R^+ \geq \int_0^T r_u du \geq R^- \text{ and } \Xi^+ \geq \int_0^T \sigma_u^2 du \geq \Xi^- \tag{2.21}$$

Is there any sense that we can hedge an option based on this interval?

We shall see that for a European call there is a strategy, beginning with wealth $C(S_0, \Xi^+, R^+)$, which will be solvent for the option payoff so long as the intervals in (2.21) are realized.

First note that by direct differentiation in (2.4), one obtains the two (!!!) Black-Scholes-Merton differential equations

$$\frac{1}{2}C_{SS}S^2 = C_\Xi \text{ and } -C_R = C - C_S S \tag{2.22}$$

(recall that $C(S,\Xi,R) = S\Phi(d_1) - K\exp(-R)\Phi(d_2)$ and $d_{1,2} = (\log(S/K) + R \pm \Xi/2)/\sqrt{\Xi}$).

In analogy with Section 2.1.6, consider the financial instrument with price at time t:

$$V_t = C(S_t, \Xi_t, R_t),$$

where

$$R_t = R^+ - \int_0^t r_u du \text{ and } \Xi_t = \Xi^+ - \int_0^t \sigma_u^2 du.$$

We shall see that the instrument V_t can be *self financed* by holding, at each time t,

$C_S(S_t, \Xi_t, R_t)$ units of stock, in other words $S_t C_S(S_t, \Xi_t, R_t)$ \$ of stock, and
$V_t - S_t C_S(S_t, \Xi_t, R_t) = -C_R(S_t, \Xi_t, R_t)$ \$ in bonds. (2.23)

where the equality follows from the first equation in (2.22). Note first that, from Itô's formula,

$$\begin{aligned} dV_t &= dC(S_t, \Xi_t, R_t) \\ &= C_S dS_t + C_R dR_t + C_\Xi d\Xi_t + \frac{1}{2} C_{SS} d[S,S]_t \\ &= C_S dS_t - C_R r_t dt - C_\Xi \sigma_t^2 dt + \frac{1}{2} C_{SS} S_t^2 \sigma_t^2 dt \\ &= C_S dS_t - C_R r_t dt \end{aligned} \tag{2.24}$$

because of the second equation in (2.22).

From equation (2.24), we see that holding C_S units of stock, and $-C_R$ \$ of bonds at all times t does indeed produce a P/L $V_t - V_0$, so that starting with V_0 \$ yields V_t \$ at time t.

From the second equation in (2.23), we also see that V_t \$ is exactly the amount needed to maintain these positions in stock and bond. Thus, V_t has a self financing strategy.

Estimated volatility can come into this problem in two ways:

(1) In real time, to set the hedging coefficients: under discrete observation, use

$$\hat{\Xi}_t = \Xi^+ - \text{estimate of integrated volatility from 0 to } t.$$

(2) As an element of a forecasting procedure, to set intervals of the form (2.21).

For further literature on this approach, consult Mykland (2000, 2003a, 2003b, 2005, 2010b). The latter paper discusses, among other things, the use of this method for setting reserve requirements based on an exit strategy in the event of model failure.

For other ways of using realized volatility and similar estimators in options trading, we refer to Zhang (2001), Hayashi and Mykland (2005), Mykland and Zhang (2008), and Zhang (2009).

2.3 Behavior of Estimators: Variance

2.3.1 The Emblematic Problem: Estimation of Volatility

In this section, we develop the tools to show convergence in high frequency data. As example throughout, we consider the problem of estimation of volatility. (In the absence of microstructure.) This classical problem is that of estimating $\int_0^t \sigma_s^2 ds$. The standard estimator, *Realized Volatility (RV)*, is simply $[X,X]_t^{\mathcal{G}}$ in (2.15). The estimator is consistent as $\Delta(\mathcal{G}) \to 0$, from the very definition of quadratic variation.

This raises the question of what other properties one can associate with this estimator. For example, does the asymptotic normality continue to hold. This is a rather complex matter, as we shall see.

There is also the question of what to do in the presence of microstructure, to which we return in Section 2.5.

2.3.2 A Temporary Martingale Assumption

For now consider the case where

$$X_t = X_0 + \int_0^t \sigma_s dW_s, \tag{2.25}$$

i.e., X_t is a local martingale. We shall see in Section 2.4.4 that drift terms can easily be incorporated into the analysis.

We shall also, for now, assume that σ_t is bounded, *i.e.*, there is a nonrandom σ_+ so that

$$\sigma_t^2 \leq \sigma_+^2 \text{ for all } t. \tag{2.26}$$

This makes X_t a martingale. We shall see in Section 2.4.5 how to remove this assumption.

2.3.3 The Error Process

On a grid $\mathcal{G} = \{t_0, t_1, ..., t_n\}$, we get from Itô's formula that

$$(X_{t_{i+1}} - X_{t_i})^2 = 2\int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) dX_s + \int_{t_i}^{t_{i+1}} \sigma_s^2 ds.$$

If we set

$$t_* = \max\{t_i \in \mathcal{G} : t_i \leq t\}, \tag{2.27}$$

the same equation will hold with (t_*, t) replacing (t_i, t_{i+1}). Hence

$$M_t = \sum_{t_{i+1} \leq t} (X_{t_{i+1}} - X_{t_i})^2 + (X_t - X_{t_*})^2 - \int_0^t \sigma_s^2 ds$$

is a local martingale on the form

$$M_t = 2\sum_{t_{i+1} \leq t} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) dX_s + 2\int_{t_*}^t (X_s - X_{t_*}) dX_s.$$

On differential form $dM_t = 2(X_t - X_{t_*})dX_t$. We shall study the behavior of martingales such as M_t.

Of course, we only observe $[X, X]_t^{\mathcal{G}} = \sum_{t_{i+1} \leq t} (X_{t_{i+1}} - X_{t_i})^2$, but we shall see next that the same results apply to this quantity. ($(X_t - X_{t_*})^2$ is negligible.)

2.3.4 Stochastic Order Symbols

We also make use of the following notation:

Definition 2.16 *(stochastic order symbols) Let Z_n be a sequence of random variables. We say that $Z_n = o_p(1)$ if $Z_n \to 0$ in probability, and that $Z_n = o_p(u_n)$ if $Z_n/u_n = o_p(1)$. Similarly, we say that $Z_n = O_p(1)$ – "bounded in probability" – if for all $\epsilon > 0$, there is an M so that $\sup_n P(|Z_n| > M) \leq \epsilon$. There is a theorem to the effect that this is the same as saying that for every subsequence n_k, there is a further subsequence n_{k_l} so that $Z_{n_{k_l}}$ converges in law. (See Theorem 29.3 (p. 380) in Billingsley (1995)). Finally, $Z_n = O_p(u_n)$ if $Z_n/u_n = O_p(1)$.*

For further discusion of this notation, see the Appendix A in Pollard (1984). (This book is out of print, but can at the time of writing be downloaded from http://www.stat.yale.edu/∼pollard/).

To see an illustration of the usage: under (2.26), we have that

$$\begin{aligned} E(X_t - X_{t_*})^2 &= E([X,X]_t - [X,X]_{t_*}) \\ &= E\int_{t_*}^{t} \sigma_s^2 ds \\ &\leq E(t - t_*)\sigma_+^2 \\ &\leq E\Delta(\mathcal{G})\sigma_+^2 \end{aligned}$$

so that $(X_t - X_{t_*})^2 = O_p(E\Delta(\mathcal{G}))$, by Chebychev's inequality.

2.3.5 Quadratic Variation of the Error Process: Approximation by Quarticity

An Important Result

To find the variance of our estimate, we start by computing the quadratic variation

$$\begin{aligned} [M,M]_t &= 4\sum_{t_{i+1}\leq t}\int_{t_i}^{t_{i+1}} (X_s - X_{t_i})^2 d[X,X]_s \\ &\quad + 4\int_{t_*}^{t} (X_s - X_{t_*})^2 d[X,X]_s. \end{aligned} \tag{2.28}$$

It is important here that we mean $[M,M]_t$, and not $[M,M]_t^{\mathcal{G}}$.

A nice result, originally due to Barndorff-Nielsen and Shephard (2002), concerns the estimation of this variation. Define the quarticity by

$$[X,X,X,X]_t^{\mathcal{G}} = \sum_{t_{i+1}\leq t} (X_{t_{i+1}} - X_{t_i})^4 + (X_t - X_{t_*})^4.$$

Use Itô's formula to see that (where M_t is the error process from Section 2.3.3)

$$d(X_t - X_{t_i})^4 = 4(X_t - X_{t_i})^3 dX_t + 6(X_t - X_{t_i})^2 d[X,X]_t$$
$$= 4(X_t - X_{t_i})^3 dX_t + \frac{6}{4} d[M,M]_t,$$

since $d[M,M]_t = 4(X_t - X_{t_i})^2 d[X,X]_t$. It follows that if we set

$$M_t^{(2)} = \sum_{t_{i+1} \le t} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i})^3 dX_s + \int_{t_*}^{t} (X_s - X_{t_*})^3 dX_s$$

we obtain

$$[X,X,X,X]_t^{\mathcal{G}} = \frac{3}{2}[M,M]_t + 4M_t^{(2)}.$$

It turns out that the $M_t^{(2)}$ term is of order $o_p(n^{-1})$, so that $(2/3)n[X,X,X,X]_t^{\mathcal{G}}$ is a consistent estimate of the quadratic variation (2.28):

Proposition 2.17 *Assume (2.26). Consider a sequence of grids* $\mathcal{G}_n = \{0 = t_{n,0} < ... < t_{n,n} = T\}$. *Suppose that, as* $n \to \infty$, $\Delta(\mathcal{G}_n) = o_p(1)$, *and*

$$\sum_{i=0}^{n-1} (t_{n,i+1} - t_{n,i})^3 = O_p(n^{-2}). \tag{2.29}$$

Then

$$\sup_{0 \le t \le T} \mid [M,M]_t - \frac{2}{3}[X,X,X,X]_t^{\mathcal{G}_n} \mid \; = \; o_p(n^{-1}) \text{ as } n \to \infty.$$

Note that in the following, we typically suppress the double subscript on the times:

$$t_i \text{ means } t_{n,i}.$$

The Conditions on the Times – Why They are Reasonable

Example 2.18 *We first provide a simple example to emphasize that Proposition 2.17 does the right thing. Assume for simplicity that the observation times are equidistant:* $t_i = t_{n,i} = iT/n$, *and that the volatility is constant:* $\sigma_t \equiv \sigma$. *It is then easy to see that the conditions, including (2.29), are satisfied. On the other hand,* $[X,X,X,X]_t^{\mathcal{G}}$ *has the distribution of* $(T/n)^2\sigma^4 \sum_{i=1}^n U_i^4$, *where the* U_i *are iid standard normal. Hence,* $n\frac{2}{3}[X,X,X,X]_t^{\mathcal{G}} \xrightarrow{p} \frac{2}{3}T^2\sigma^4 \times E(N(0,1)^4) = 2T^2\sigma^4$. *It then follows from Proposition 2.17 that* $n[M,M]_t^{\mathcal{G}} \xrightarrow{p} 2T^2\sigma^4$.

Example 2.19 *To see more generally why (2.29) is a natural condition, consider a couple of cases for the spacings.*

(i) The spacings are sufficiently regular to satisfy

$$\Delta(\mathcal{G}) = \max_i (t_{i+1} - t_i) = O_p(n^{-1}).$$

Then

$$\sum_{i=0}^{n} (t_{i+1} - t_i)^3 \le \sum_{i=0}^{n} (t_{i+1} - t_i) \left(\max_i (t_{i+1} - t_i) \right)^2 = T \times O_p(n^{-2})$$

(ii) On the other hand, suppose that the sampling times follow a Poisson process with parameter λ (still with $t_0 = 0$). Denote by N the number of sampling points in the interval $[0, T]$, i.e., $N = \inf\{i : t_i > T\}$. If one conditions on N, say, $N = n$, the conditional distribution of the points $t_i, i = 1, ..., n-1$, behave like the order statistics of $n-1$ uniformly distributed random variables (see, for example, Chapter 2.3 in Ross (1996)). In other words, $t_i = TU_{(i)}$ (for $0 < i < n$), where $U_{(i)}$ is the i'th order statistic of $U_1, ..., U_{n-1}$, which are iid U[0,1]. Without any asymptotic impact, now also impose $t_n = T$ (to formally match the rest of our theory).

Now define $U_{(0)} = 0$ and $U_{(n)} = 1$. With these definitions, note that for $i = 1, ..., n$, $U_{(i)} - U_{(i-1)}$ are identically distributed with the same distribution as $U_{(1)}$, which has density $(n-1)(1-x)^{n-2}$. (See, for example, Exercise 3.67 (p. 110) in Rice (2006).) The expression in (2.29) becomes

$$\sum_{i=0}^{n-1} (t_{i+1} - t_i)^3 = T^3 \sum_{i=1}^{n} (U_{(i)} - U_{(i-1)})^3 = T^3 n E U_{(1)}^3 (1 + o_p(1))$$

by the law of large numbers. Since $EU_{(1)}^3 = \frac{6}{(n+1)n(n-1)} = O(n^{-3})$, (2.29) follows.

Application to Refresh Times

We here briefly consider the case of multidimensional processes of the form $(X_t^{(1)}, ..., X_t^{(p)})$. It will often be the case that the observation occurs at asynchronous times. In other words, process $(X_t^{(r)})$ is observed at times $\mathcal{G}_n^{(r)} = \{0 \le t_{n,0}^{(r)} < t_{n,1}^{(r)} < ... < t_{n,n_r}^{(r)} \le T\}$, and the grids $\mathcal{G}_n^{(r)}$ are not the same. Note that in this case, there is latitude in what meaning to assign to the symbol n. It is an index that goes to infinity with each n_r, for example $n = n_1 + ... + n_p$. One would normally require that n_r/n is bounded away from zero.

A popular way of dealing with this problem is to use *refresh times*, as follows.

Set $u_{n,0} = 0$, and then define recursively for $i > 0$

$$u_{n,i} = \max_{r=1,\ldots,p} \min\{t \in \mathcal{G}_n^{(r)} : t > u_{n,i-1}\}.$$

The $u_{n,i}$ is called the i'th refresh time, and is the time when all the p processes have undergone an update of observation. Successful uses of refresh times can be found in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2011) and Zhang (2011).

Now note that if the conditions (on times) in Proposition 2.17 are satisfied for each grid $\mathcal{G}_n^{(r)}$, the conditions are also satisfied for the grid of refresh times. This is because each $\Delta u_{n,i+1}$ must be matched or dominated by a spacing in one of each grid $\mathcal{G}_n^{(r)}$. Specifically, for each i, define $j_{r,i} = \max\{j : t_{n,j}^{(r)} \leq u_{n,i}\}$ and note that $j_{r,i} + 1 = \min\{j : t_{n,j}^{(r)} > u_{n,i}\}$. Hence, there is an r_i so that

$$u_{n,i+1} = \max_r t_{n,j_{r,i}+1}^{(r)} = t_{n,j_{r_i,i}+1}^{(r_i)}$$

and so

$$u_{n,i+1} - u_{n,i} \leq t_{n,j_{r_i,i}+1}^{(r_i)} - t_{n,j_{r_i,i}}^{(r_i)} \leq \max_r \left(t_{n,j_{r,i}+1}^{(r)} - t_{n,j_{r,i}}^{(r)} \right).$$

In particular, for (2.29),

$$\begin{aligned} \sum_i (u_{n,i+1} - u_{n,i})^3 &\leq \sum_i \max_r \left(t_{n,j_{r,i}+1}^{(r)} - t_{n,j_{r,i}}^{(r)} \right)^3 \\ &\leq \sum_i \sum_r \left(t_{n,j_{r,i}+1}^{(r)} - t_{n,j_{r,i}}^{(r)} \right)^3 \\ &\leq \sum_{r=1}^{p} \sum_i \left(t_{n,i+1}^{(r)} - t_{n,i}^{(r)} \right)^3, \end{aligned}$$

and similarly for the condition $\Delta(\mathcal{G}_n) = o_p(1)$.

The theory in this article is therefore amenable to developments involving refresh times. This issue is not further pursued here, though we return to asynchronous times in Section 2.6.3.

2.3.6 Moment Inequalities, and Proof of Proposition 2.17

L^p Norms, Moment Inequalities, and the Burkholder-Davis-Gundy Inequality

For $1 \leq p < \infty$, define the *L^p-norm*:

$$||X||_p = (E|X|^p)^{\frac{1}{p}},$$

The Minkowski and Hölder inequalities say that

$$||X+Y||_p \leq ||X||_p + ||Y||_p$$
$$||XY||_1 \leq ||X||_p ||Y||_q \text{ for } \frac{1}{p}+\frac{1}{q}=1.$$

Example 2.20 *A special case of the Hölder inequalitiy is* $||X||_1 \leq ||X||_p$ *(take* $Y=1$*). In particular, under (2.29), for for* $1 \leq v \leq 3$*:*

$$\left(\frac{1}{n}\sum_{i=0}^{n-1}(t_{i+1}-t_i)^v\right)^{\frac{1}{v}} \leq \left(\frac{1}{n}\sum_{i=0}^{n-1}(t_{i+1}-t_i)^3\right)^{\frac{1}{3}}$$
$$= \left(\frac{1}{n}\times O_p(n^{-2})\right)^{\frac{1}{3}} = \left(O_p(n^{-3})\right)^{\frac{1}{3}} = O_p(n^{-1}),$$

so that

$$\sum_{i=0}^{n}(t_{i+1}-t_i)^v = O_p(n^{1-v}). \tag{2.30}$$

To show Proposition 2.17, we need the Burkholder-Davis-Gundy inequality (see Section 3 of Ch. VII of Dellacherie and Meyer (1982), or p. 193 and 222 in Protter (2004)), as follows. For $1 \leq p < \infty$, there are universal constants c_p and C_p so that for all continuous martingales N_t,

$$c_p||[N,N]_T||_{p/2}^{1/2} \leq ||\sup_{0\leq t\leq T}|N_t|\;||_p \leq C_p||[N,N]_T||_{p/2}^{1/2}.$$

Note, in particular, that for $1 < p < \infty$,

$$C_p^2 = q^p\left(\frac{p(p-1)}{2}\right)$$

where q is given by $p^{-1}+q^{-1}=1$.

Proof of Proposition 2.17

From applying Itô's Formula to $(X_t - X_{t_i})^8$:

$$\begin{aligned}[M^{(2)},M^{(2)}]_t &= \sum_{t_{i+1}\leq t}\int_{t_i}^{t_{i+1}}(X_s-X_{t_i})^6 d[X,X]_s \\ &\quad + \int_{t_*}^{t}(X_s-X_{t_*})^6 d[X,X]_s \\ &= \frac{1}{28}[X;8]_t^{\mathcal{G}} + \text{martingale term}\end{aligned}$$

where $[X;8]_t^{\mathcal{G}} = \sum_{t_{i+1}\leq t}(X_{t_{i+1}}-X_{t_i})^8 + (X_t - X_{t_*})^8$ is the *ochticity*.

Note that for stopping time $\tau \le T$, $[X;8]^{\mathcal{G}}_\tau = \sum_i (X_{t_{i+1}\wedge\tau} - X_{t_i\wedge\tau})^8$. Hence, by the Burkholder-Davis-Gundy inequality (with $p = 8$)

$$\begin{aligned} E[M^{(2)}, M^{(2)}]_\tau &= \frac{1}{28} E[X;8]^{\mathcal{G}}_\tau \\ &\le \frac{1}{28} C_8^8 E \sum_i ([X,X]_{t_{i+1}\wedge\tau} - [X,X]_{t_i\wedge\tau})^4 \\ &\le \frac{1}{28} C_8^8 \sigma_+^8 E \sum_i (t_{i+1}\wedge\tau - t_i\wedge\tau)^4. \end{aligned}$$

Let $\epsilon > 0$, and set

$$\tau_n = \inf\{\, t \in [0,T] : n^2 \sum_i (t_{i+1}\wedge t - t_i \wedge t)^4 > \epsilon \,\}.$$

Then

$$E[M^{(2)}, M^{(2)}]_{\tau_n} \le n^{-2} \frac{1}{28} C_8^8 \sigma_+^8 \epsilon. \tag{2.31}$$

By assumption, $n^2 \sum_i (t_{i+1}\wedge t - t_i\wedge t)^4 \ \le\ \Delta(\mathcal{G}) n^2 \sum_i (t_{i+1} - t_i)^3 \xrightarrow{p} 0$, and hence

$$P(\tau_n \neq T) \to 0 \text{ as } n \to \infty. \tag{2.32}$$

Hence, for any $\delta > 0$,

$$\begin{aligned} &P(n \sup_{0\le t\le T} |M_t^{(2)}| \ > \delta) \qquad (2.33) \\ \le\ & P(n \sup_{0\le t\le \tau_n} |M_t^{(2)}| \ > \delta) + P(\tau_n \neq T) \\ \le\ & \frac{1}{\delta^2} E\left(n \sup_{0\le t\le\tau_n} |M_t^{(2)}| \right)^2 + P(\tau_n \neq T) \text{ (Chebychev)} \\ \le\ & \frac{1}{\delta^2} C_2^2 n^2 E[M^{(2)}, M^{(2)}]_{\tau_n} + P(\tau_n \neq T) \text{ (Burkholder-Davis-Gundy)} \\ \le\ & \frac{1}{\delta^2} C_2^2 \frac{1}{28} C_8^8 \sigma_+^8 \epsilon + P(\tau_n \neq T) \text{ (from (2.31))} \\ \to\ & \frac{1}{\delta^2} C_2^2 \frac{1}{28} C_8^8 \sigma_+^8 \epsilon \text{ as } n \to \infty \text{ (from (2.32))}. \end{aligned}$$

Hence Proposition 2.17 has been shown.

2.3.7 Quadratic Variation of the Error Process: When Observation Times are Independent of the Process

Main Approximation

We here assume that the observation times are independent of the process X. The basic insight for the following computation is that over small intervals,

$(X_t - X_{t_*})^2 \approx [X,X]_t - [X,X]_{t_*}$. To the extent that this approximation is valid, it follows from (2.28) that

$$\begin{aligned}[M,M]_t &\approx 4\sum_{t_{i+1}\le t}\int_{t_i}^{t_{i+1}}([X,X]_s - [X,X]_{t_i})d[X,X]_s \\ &\quad + 4\int_{t_*}^{t}([X,X]_s - [X,X]_{t_*})d[X,X]_s \\ &= 2\sum_{t_{i+1}\le t}([X,X]_{t_{i+1}} - [X,X]_{t_i})^2 + 2([X,X]_t - [X,X]_{t_*})^2.\end{aligned}$$

We shall use this device several times in the following, and will this first time do it rigorously.

Proposition 2.21 *Assume (2.26), and that σ_t^2 is continuous in mean square:*

$$\sup_{0\le t-s\le\delta} E(\sigma_t^2 - \sigma_s^2)^2 \to 0 \text{ as } \delta \to \infty.$$

Also suppose that the grids $\mathcal{G}_n$ are nonrandom, or independent of the process X_t. Also suppose that, as $n \to \infty$, $\Delta(\mathcal{G}) = o_p(n^{-1/2})$, and assume (2.29). Then

$$\begin{aligned}[M,M]_t &= 2\sum_{t_{i+1}\le t}([X,X]_{t_{i+1}} - [X,X]_{t_i})^2 + 2([X,X]_t - [X,X]_{t_*})^2 \\ &\quad + o_p(n^{-1}). \end{aligned} \tag{2.34}$$

If σ_t is continuous, it is continuous in mean square (because of (2.26)). More generally, σ_t can, for example, also have Poisson jumps.

In the rest of this section, we shall write all expectations implicitly as conditional on the times.

To show Proposition 2.21, we need some notation and a lemma, as follows:

Lemma 2.22 *Let $t_* = \max\{t_i \in \mathcal{G} : t_i \le t\}$ (as in (2.27)). Let N_t be an Itô process martingale, for which (for $a, b > 0$), for all t,*

$$\frac{d}{dt}E[N,N]_t \le a(t-t_*)^b.$$

Let H_t be a predictable process, satisfying $|H_t| \le H_+$ for some constant H_+. Set

$$R_v(\mathcal{G}) = \left(\sum_{i=0}^{n-1}(t_{i+1}-t_i)^v\right).$$

Then

$$\begin{aligned}
&|| \sum_{t_{i+1} \leq t} \int_{t_i}^{t_{i+1}} (N_s - N_{t_i}) H_s ds + \int_{t_*}^{t} (N_s - N_{t_*}) H_s ds ||_1 \\
&\qquad \leq \left(H_+^2 \frac{a}{b+3} R_{b+3}(\mathcal{G}) \right)^{1/2} \\
&\qquad\quad + R_{(b+3)/2}(\mathcal{G}) \frac{2}{b+3} \left(\frac{a}{b+1} \right)^{1/2} \sup_{0 \leq t-s \leq \Delta(\mathcal{G})} ||H_s - H_t||_2.
\end{aligned} \tag{2.35}$$

Proof of Proposition 2.21. Set $N_t = M_t$ and $H_t = \sigma_t^2$. Then

$$\begin{aligned}
d[M, M]_t &= 4(X_t - X_{t_i})^2 d[X, X]_t \\
&= 4([X, X]_t - [X, X]_{t_i}) d[X, X]_t + 4((X_t - X_{t_i})^2 \\
&\quad - ([X, X]_t - [X, X]_{t_i})) d[X, X]_t \\
&= 4([X, X]_t - [X, X]_{t_i}) d[X, X]_t + 2(N_t - N_{t_i}) \sigma_t^2 dt.
\end{aligned}$$

Thus, the approximation error in (2.34) is exactly of the form of the left hand side in (2.35). We note that

$$\begin{aligned}
Ed[N, N]_t &= 4E(X_t - X_{t_i})^2 d[X, X]_t \\
&= 4E(X_t - X_{t_i})^2 \sigma_+^2 dt \\
&= 4(t - t_i) \sigma_+^4 dt
\end{aligned}$$

hence the conditions of Lemma 2.22 are satisfied with $a = 4\sigma_+^4$ and $b = 1$. The result follows from (2.30). □

Proof of Lemma 2.22 (Technical Material, can be omitted).

Decompose the original problem as follows:

$$\begin{aligned}
&\int_{t_i}^{t_{i+1}} (N_s - N_{t_i}) H_s ds \\
&\quad = \int_{t_i}^{t_{i+1}} (N_s - N_{t_i}) H_{t_i} ds + \int_{t_i}^{t_{i+1}} (N_s - N_{t_i})(H_s - H_{t_i}) ds.
\end{aligned}$$

For the first term, from Itô's formula, $d(t_{i+1} - s)(N_s - N_{t_i}) = -(N_s - N_{t_i}) ds + (t_{i+1} - s) dN_s$, so that

$$\int_{t_i}^{t_{i+1}} (N_s - N_{t_i}) H_{t_i} ds = H_{t_i} \int_{t_i}^{t_{i+1}} (t_{i+1} - s) dN_s$$

hence

$$\begin{aligned}
&\sum_{t_{i+1}\le t}\int_{t_i}^{t_{i+1}}(N_s - N_{t_i})H_s ds \\
&= \sum_{t_{i+1}\le t} H_{t_i}\int_{t_i}^{t_{i+1}}(t_{i+1}-t)dN_s \\
&\quad + \sum_{t_{i+1}\le t}\int_{t_i}^{t_{i+1}}(N_s - N_{t_i})(H_s - H_{t_i})ds. \qquad (2.36)
\end{aligned}$$

The first term is the end point of a martingale. For each increment,

$$\begin{aligned}
E\left(\int_{t_i}^{t_{i+1}}(N_s - N_{t_i})H_{t_i}ds\right)^2 &= E\left(H_{t_i}\int_{t_i}^{t_{i+1}}(t_{i+1}-s)dN_s\right)^2 \\
&\le H_+^2 E\left(\int_{t_i}^{t_{i+1}}(t_{i+1}-s)dN_s\right)^2 \\
&= H_+^2 E\left(\int_{t_i}^{t_{i+1}}(t_{i+1}-s)^2 d[N,N]_s\right) \\
&= H_+^2 \int_{t_i}^{t_{i+1}}(t_{i+1}-s)^2 dE[N,N]_s \\
&= H_+^2 \int_{t_i}^{t_{i+1}}(t_{i+1}-s)^2 \frac{d}{ds}E[N,N]_s ds \\
&= H_+^2 \int_{t_i}^{t_{i+1}}(t_{i+1}-s)^2 a(s-t_i)^b ds \\
&= H_+^2 \frac{a}{b+3}(t_{i+1}-t_i)^{b+3}
\end{aligned}$$

and so, by the uncorrelatedness of martingale increments,

$$\begin{aligned}
E\left(\sum_{t_{i+1}\le t} H_{t_i}\int_{t_i}^{t_{i+1}}(t_{i+1}-t)dN_s\right)^2 &\le H_+^2\frac{a}{b+3}\left(\sum_{t_{i+1}\le t}(t_{i+1}-t_i)^3\right) \\
&\le H_+^2\frac{a}{b+3}R_{b+3}(\mathcal{G}). \qquad (2.37)
\end{aligned}$$

On the other hand, for the second term in (2.36),

$$\begin{aligned}
&||(N_s - N_{t_i})(H_s - H_{t_i})||_1 \\
&\quad \leq \quad ||N_s - N_{t_i}||_2 ||H_s - H_{t_i}||_2 \\
&\quad \leq \quad \left(E(N_s - N_{t_i})^2\right)^{1/2} ||H_s - H_{t_i}||_2 \\
&\quad = \quad (E([N,N]_s - [N,N]_{t_i}))^{1/2} ||H_s - H_{t_i}||_q \\
&\quad = \quad \left(\int_{t_i}^{s} \frac{d}{du} E[N,N]_u du\right)^{1/2} ||H_s - H_{t_i}||_2 \\
&\quad \leq \quad \left(\int_{t_i}^{s} a(u - t_i)^b du\right)^{1/2} ||H_s - H_{t_i}||_2 \\
&\quad = \quad \left(\frac{a}{b+1}(s - t_i)^{b+1}\right)^{1/2} ||H_s - H_{t_i}||_2 \\
&\quad = \quad (s - t_i)^{(b+1)/2} \left(\frac{a}{b+1}(s - t_i)^{b+1}\right)^{1/2} ||H_s - H_{t_i}||_2,
\end{aligned}$$

and from this

$$\begin{aligned}
&|| \int_{t_i}^{t_{i+1}} (N_s - N_{t_i})(H_s - H_{t_i}) ds||_1 \\
&\quad \leq \quad \int_{t_i}^{t_{i+1}} ||(N_s - N_{t_i})(H_s - H_{t_i})||_1 ds \\
&\quad \leq \quad \int_{t_i}^{t_{i+1}} (s - t_i)^{(b+1)/2} ds \left(\frac{a}{b+1}\right)^{1/2} \sup_{t_i \leq s \leq t_{i+1}} ||H_s - H_{t_i}||_2 \\
&\quad = \quad (t_{i+1} - t_i)^{(b+3)/2} \frac{2}{b+3} \left(\frac{a}{b+1}\right)^{1/2} \sup_{t_i \leq s \leq t_{i+1}} ||H_s - H_{t_i}||_2.
\end{aligned}$$

Hence, finally, for the second term in (2.36),

$$\begin{aligned}
&|| \sum_{t_{i+1} \leq t} \int_{t_i}^{t_{i+1}} (N_s - N_{t_i})(H_s - H_{t_i}) dt||_1 \\
&\quad \leq \quad \left(\sum_{t \leq t_{i+1}} (t_{i+1} - t_i)^{(b+3)/2}\right) \frac{2}{b+3} \left(\frac{a}{b+1}\right)^{1/2} \sup_{0 \leq t-s \leq \Delta(\mathcal{G})} ||H_s - H_t||_2 \\
&\quad = \quad R_{(b+3)/2}(\mathcal{G}) \frac{2}{b+3} \left(\frac{a}{b+1}\right)^{1/2} \sup_{0 \leq t-s \leq \Delta(\mathcal{G})} ||H_s - H_t||_2. \qquad (2.38)
\end{aligned}$$

Hence, for the overall sum (2.36), from (2.37) and (2.38) and

$$
\begin{aligned}
&|| \sum_{t_{i+1} \le t} \int_{t_i}^{t_{i+1}} (N_s - N_{t_i}) H_s ds ||_1 \\
\le \quad & || \sum_{t_{i+1} \le t} H_{t_i} \int_{t_i}^{t_{i+1}} (t_{i+1} - t) dN_s ||_1 \\
& + || \sum_{t_{i+1} \le t} \int_{t_i}^{t_{i+1}} (N_s - N_{t_i})(H_s - H_{t_i}) ds ||_1 \\
\le \quad & || \sum_{t_{i+1} \le t} H_{t_i} \int_{t_i}^{t_{i+1}} (t_{i+1} - t) dN_s ||_2 \\
& + || \sum_{t_{i+1} \le t} \int_{t_i}^{t_{i+1}} (N_s - N_{t_i})(H_s - H_{t_i}) ds ||_1 \\
\le \quad & \left(H_+^2 \frac{a}{b+3} R_{b+3}(\mathcal{G}) \right)^{1/2} \\
& + R_{(b+3)/2}(\mathcal{G}) \frac{2}{b+3} \left(\frac{a}{b+1} \right)^{1/2} \sup_{0 \le t-s \le \Delta(\mathcal{G})} ||H_s - H_t||_2 .
\end{aligned}
$$

The part from t_* to t can be included similarly, showing the result. □

Quadratic Variation of the Error Process, and Quadratic Variation of Time

To give the final form to this quadratic variation, define the "Asymptotic Quadratic Variation of Time" (AQVT), given by

$$
H_t = \lim_{n \to \infty} \frac{n}{T} \sum_{t_{n,j+1} \le t} (t_{n,j+1} - t_{n,j})^2, \tag{2.39}
$$

provided that the limit exists. From Example 2.19, we know that dividing by n is the right order. We now get

Proposition 2.23 *Assume the conditions of Proposition 2.21, and that the AQVT exists. Then*

$$
n[M, M]_t \xrightarrow{p} 2T \int_0^t \sigma_s^4 dH_s .
$$

The proof is a straight exercise in analysis. The heuristic for the result is as follows. From (2.34),

$$\begin{aligned}
[M,M]_t &= 2\sum_{t_{i+1}\le t}([X,X]_{t_{i+1}}-[X,X]_{t_i})^2 + 2([X,X]_t-[X,X]_{t_*})^2 \\
&\quad +\ o_p(n^{-1}) \\
&= 2\sum_{t_{i+1}\le t}\Big(\int_{t_i}^{t_{i+1}}\sigma_s^2 ds\Big)^2 + 2\Big(\int_{t_*}^{t}\sigma_s^2 ds\Big)^2 + o_p(n^{-1}) \\
&= 2\sum_{t_{i+1}\le t}((t_{i+1}-t_i)\sigma_{t_i}^2)^2 + 2((t-t_*)\sigma_{t_*}^2)^2 + o_p(n^{-1}) \\
&= 2\frac{T}{n}\int_0^t \sigma_s^4 dH_s + o_p(n^{-1}).
\end{aligned}$$

Example 2.24 *We here give a couple of examples of the AQVT:*
(i) When the times are equidistant: $t_{i+1}-t_i = T/n$, *then*

$$\begin{aligned}
H_t &\approx \frac{n}{T}\sum_{t_{n,j+1}\le t}\left(\frac{T}{n}\right)^2 \\
&= \frac{T}{n}\#\{t_{i+1}\le t\} \\
&= T\times \textit{ fraction of } t_{i+1} \textit{ in } [0,t] \\
&\approx T\times\frac{t}{T} = t.
\end{aligned}$$

(ii) When the times follow a Poisson process with parameter λ*, we proceed as in case (ii) in Example 2.19. We condition on the number of sampling points* n*, and get* $t_i = TU_{(i)}$ *(for* $0<i<n$*), where* $U_{(i)}$ *is the i'th order statistic of* $U_1,\ldots,U_n$*, which are iid U[0,1]. Hence (again taking* $U_{(0)}=0$ *and* $U_{(n)}=1$*)*

$$\begin{aligned}
H_t &\approx \frac{n}{T}\sum_{t_{n,j+1}\le t}(t_{i+1}-t_i)^2 \\
&= T^2\frac{n}{T}\sum_{t_{n,j+1}\le t}(U_{(i)}-U_{(i-1)})^2 \\
&= T^2\frac{n}{T}\sum_{t_{n,j+1}\le t}EU_{(1)}^2(1+o_p(1)) \\
&= T^2\frac{n}{T}\#\{t_{i+1}\le t\}EU_{(1)}^2(1+o_p(1)) \\
&= Tn^2\frac{t}{T}EU_{(1)}^2(1+o_p(1)) \\
&= 2t(1+o_p(1))
\end{aligned}$$

by the law of large numbers, since the spacings have identical distribution, and since $EU_{(1)}^2 = 2/n(n+1)$. *Hence* $H_t = 2t$.

The Quadratic Variation of Time in the General Case

We now go back to considering the times as possibly dependent with the process X. Note that by using the Burkholder-Davis-Gundy Inequality conditionally, we obtain that

$$\begin{aligned} &c_4^4 E(([X,X]_{t_{i+1}} - [X,X]_{t_i})^2 \mid \mathcal{F}_{t_i}) \\ &\quad \leq \quad E((X_{t_{i+1}} - X_{t_i})^4 \mid \mathcal{F}_{t_i}) \quad \leq \quad C_4^4 E(([X,X]_{t_{i+1}} - [X,X]_{t_i})^2 \mid \mathcal{F}_{t_i}), \end{aligned}$$

where c_4 and C_4 are as in Section 2.3.6. In the typical law of large numbers setting, $[X,X,X,X]_t - \sum_i E((X_{t_{i+1}} - X_{t_i})^4 \mid \mathcal{F}_{t_i})$ is a martingale which is of lower order than $[X,X,X,X]_t$ itself, and the same goes for

$$\sum_i \left[([X,X]_{t_{i+1}} - [X,X]_{t_i})^2 - E(([X,X]_{t_{i+1}} - [X,X]_{t_i})^2 \mid \mathcal{F}_{t_i}) \right].$$

By the argument in Proposition 2.23, therefore, it follows that under suitable regularity conditions, if $n[X,X,X,X]_t \xrightarrow{p} U_t$ as $n \to \infty$, and if the AQVT H_t is absolutely continuous in t, then U_t is also absolutely continuous, and

$$c_4^4 2T\sigma_t^4 H_t' \leq U_t' \leq C_4^4 2T\sigma_t^4 H_t'.$$

This is of some theoretic interest in that it establishes the magnitude of the limit of $n[X,X,X,X]_t$. However, it should be noted that $C_4^4 = 2^{18}/3^6 \approx 359.6$, so the bounds are of little practical interest.

A slightly closer analysis of this particular case uses the Bartlett type identities for martingales to write

$$\begin{aligned} &E((X_{t_{i+1}} - X_{t_i})^4 \mid \mathcal{F}_{t_i}) \\ &\quad = \quad -3E(([X,X]_{t_{i+1}} - [X,X]_{t_i})^2 \mid \mathcal{F}_{t_i}) \\ &\qquad\quad + 6E((X_{t_{i+1}} - X_{t_i})^2([X,X]_{t_{i+1}} - [X,X]_{t_i}) \mid \mathcal{F}_{t_i}) \\ &\quad \leq \quad -3E(([X,X]_{t_{i+1}} - [X,X]_{t_i})^2 \mid \mathcal{F}_{t_i}) \\ &\qquad\quad + 6E((X_{t_{i+1}} - X_{t_i})^4 \mid \mathcal{F}_{t_i})^{1/2} E(([X,X]_{t_{i+1}} - [X,X]_{t_i})^2 \mid \mathcal{F}_{t_i})^{1/2}. \end{aligned}$$

Solving this quadratic inequality yields that we can take $c_4^4 = (3-\sqrt{6})^2 \approx 0.3$ and $C_4^4 = (3+\sqrt{6})^2 \approx 29.7$.

2.3.8 Quadratic Variation, Variance, and Asymptotic Normality

We shall later see that $n^{1/2}([X,X]_t^{\mathcal{G}} - [X,X]_t)$ is approximately normal. In the simplest case, where the times are independent of the process, the normal distribution has mean zero and variance $n[M,M]_t \approx 2T \int_0^t \sigma_s^4 dH_s$. From standard central limit considerations, this is unsurprising when the σ_t process is nonrandom, or more generally independent of the W_t process. (In the latter case, one simply conditions on the σ_t process).

What is surprising, and requires more concepts, is that the normality result also holds when σ_t process has dependence with the W_t process. For this we shall need new concepts, to be introduced in Section 2.4.

2.4 Asymptotic Normality

2.4.1 Stable Convergence

In order to define convergence in law, we need to deal with the following issue. Suppose $\hat{\theta}_n$ is an estimator of θ, say, $\hat{\theta}_n = [X, X]_T^{\mathcal{G}_n}$ and $\theta = [X, X]_T = \int_0^T \sigma_t^2 dt$. As suggested in Section 2.3.7, the variance of $Z_n = n^{1/2}(\hat{\theta}_n - \theta)$ converges to $2T \int_0^t \sigma_s^4 dH_s$. We shall now go on to show the following convergence in law:

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} U \times \left(2T \int_0^T \sigma_s^4 dH_s\right)^{1/2},$$

where U is a standard normal random variable, independent of the σ_t^2 process. In order to show this, we need to be able to bring along prelimiting information into the limit: U only exists in the limit, while as argued in Section 2.3.5, the asymptotic variance $2T \int_0^T \sigma_s^4 dH_s$ can be estimated consistently, and so is a limit in probability of a prelimiting quantity.

To operationalize the concept in our setting, we need the filtration $(\mathcal{F}_t)$ to which all relevant processes (X_t, σ_t, etc) are adapted. We shall assume that Z_n (the quantity that is converging in law) to be measurable with respect to a σ-field χ, $\mathcal{F}_T \subseteq \chi$. The reason for this is that it is often convenient to exclude microstructure noise from the filtration $\mathcal{F}_t$. Hence, for example, the TSRV (in Section 2.5 below) is not $\mathcal{F}_T$-measurable.

Definition 2.25 *Let Z_n be a sequence of χ-measurable random variables, $\mathcal{F}_T \subseteq \chi$. We say that Z_n converges $\mathcal{F}_T$-stably in law to Z as $n \to \infty$ if Z is measurable with respect to an extension of χ so that for all $A \in \mathcal{F}_T$ and for all bounded continuous g, $EI_A g(Z_n) \to EI_A g(Z)$ as $n \to \infty$.*

The definition means, up to regularity conditions, that Z_n converges jointly in law with all $\mathcal{F}_T$ measurable random variables. This intuition will be imprortant in the following. For further discussion of stable convergence, see Rényi (1963), Aldous and Eagleson (1978), Chapter 3 (p. 56) of Hall and Heyde (1980), Rootzén (1980) and Section 2 (p. 169-170) of Jacod and Protter (1998). We now move to the main result.

2.4.2 Asymptotic Normality

We shall be concerned with a sequence of martingales M_t^n, $0 \leq t \leq T$, $n = 1, 2, ...$, and how it converges to a limit M_t. We consider here only continuous martingales, which are thought of as random variables taking values in the set $\mathbb{C}$ of continuous functions $[0, T] \to \mathbb{R}$.

To define weak, and stable, convergence, we need a concept of continuity. We say that g is a continuous function $\mathbb{C} \to \mathbb{R}$ if:

$$\sup_{0 \leq t \leq T} |x_n(t) - x(t)| \to 0 \text{ implies } g(x_n) \to g(x).$$

We note that if $(M_t^n) \xrightarrow{\mathcal{L}} (M_t)$ in this process sense, then, for example, $M_T^n \xrightarrow{\mathcal{L}} M_T$ as a random variable. This is because the function $x \to g(x) = x(T)$ is continuous. The reason for going via process convergence is (1) sometimes this is really the result one needs, and (2) since our theory is about continuous processes converging to a continuous process, one does not need asymptotic negligibility conditions à la Lindeberg (these kinds of conditions are in place in the usual CLT precisely to avoid jumps is the asymptotic process). For a related development based on discrete time predictable quadratic variations, and Lindeberg conditions, see Theorem IX.7.28 (p. 590-591) of Jacod and Shiryaev (2003).

In order to show results about continuous martingales, we shall use the following assumption

Condition 2.26 *There are Brownian motions $W_t^{(1)}, ..., W_t^{(p)}$ (for some p) that generate $(\mathcal{F}_t)$.*

It is also possible to proceed with assumptions under which there are jumps in some processes, but for simplicity, we omit any discussion of this here.

Under Condition 2.26, it follows from Lemma 2.1 (p. 270) in Jacod and Protter (1998) that stable convergence in law of a local martingale M^n to a process M is equivalent to (straight) convergence in law of the process $(W^{(1)}, ..., W^{(p)}, M^n)$ to the process $(W^{(1)}, ..., W^{(p)}, M)$. This result does not extend to all processes and spaces, cf. the discussion in the cited paper.

Another main fact about stable convergence is that limits and quadratic variation can be interchanged:

Proposition 2.27 *(Interchangeability of limits and quadratic variation). Assume that M^n is a sequence of continuous local martingales which converges stably to a process M. Then $(M^n, [M^n, M^n])$ converges stably to $(M, [M, M])$.*

For proof, we refer to Corollary VI.6.30 (p. 385) in Jacod and Shiryaev (2003), which also covers the case of bounded jumps. More generally, consult *ibid.*, Chapter VI.6.

We now state the main central limit theorem (CLT).

Theorem 2.28 *Assume Condition 2.26. Let (M_t^n) be a sequence of continuous local martingales on $[0,T]$, each adapted to $(\mathcal{F}_t)$, with $M_0^n = 0$. Suppose that there is an $(\mathcal{F}_t)$ adapted process f_t so that*

$$[M^n, M^n]_t \xrightarrow{p} \int_0^t f_s^2 ds \text{ for each } t \in [0,T]. \tag{2.40}$$

Also suppose that, for each $i = 1, .., p$,

$$[M^n, W^{(i)}]_t \xrightarrow{p} 0 \text{ for each } t \in [0,T]. \tag{2.41}$$

There is then an extension $(\mathcal{F}'_t)$ of $(\mathcal{F}_t)$, and an $(\mathcal{F}'_t)$-martingale M_t so that (M_t^n) converges stably to (M_t). Furthermore, there is a Brownian motion (W'_t) so that $(W_t^{(1)}, ..., W_t^{(p)}, W'_t)$ is an $(\mathcal{F}'_t)$-Wiener process, and so that

$$M_t = \int_0^t f_s dW'_s. \tag{2.42}$$

It is worth while to understand the proof of this result, and hence we give it here. The proof follows more or less *verbatim* that of Theorem B.4 in Zhang (2001) (p. 65-67). The latter is slightly more general.

Proof of Theorem 2.28. Since $[M^n, M^n]_t$ is a non-decreasing process and has non-decreasing continuous limit, the convergence (2.40) is also in law in $\mathbb{D}(\mathbb{R})$ by Theorem VI.3.37 (p. 354) in Jacod and Shiryaev (2003). Thus, in their terminology (*ibid.*, Definition VI.3.25, p. 351), $[M^n, M^n]_t$ is C-tight. From this fact, *ibid.*, Theorem VI.4.13 (p. 358) yields that the sequence M^n is tight.

From this tightness, it follows that for any subsequence M^{n_k}, we can find a further subsequence $M^{n_{k_l}}$ which converges in law (as a process) to a limit M, jointly with $W^{(1)}, ..., W^{(p)}$; in other words, $(W^{(1)}, ..., W^{(p)}, M^{n_{k_l}})$ converges in law to $(W^{(1)}, ..., W^{(p)}, M)$. This M is a local martingale by *ibid.*, Proposition IX.1.17 (p. 526), using the continuity of M_t^n. Using Proposition 2.27 above, $(M^{n_{k_l}}, [M^{n_{k_l}}, M^{n_{k_l}}])$ converge jointly in law (and jointly with the $W^{(i)}$'s) to $(M, [M, M])$. From (2.40) this means that $[M, M]_t = \int_0^t f_s^2 ds$. The continuity of $[M, M]_t$ assures that M_t is continuous. By the same reasoning, from (2.41), $[M, W^{(i)}] \equiv 0$ for each $i = 1, .., p$. Now let $W'_t = \int_0^t f_s^{-1/2} dM_s$ (if f_t is zero on a set of Lebesgue measure greater than zero, follow the alternative construction in Volume III of Gikhman and Skorohod (1969)). By Property (3) in Section 2.2.4 (or refer directly to Property I.4.54 (p.55) of Jacod and Shiryaev (2003)), $[W', W']_t = t$, while $[W', W^{(i)}] \equiv 0$. By the multivariate version of Lévy's Theorem (Section 2.2.4, or refer directly to Theorem II.4.4 (p. 102) of Jacod and Shiryaev (2003)), it therefore follows that $(W_t^{(1)}, ..., W_t^{(p)}, W_t')$ is a Wiener process. The equality (2.42) follows by construction of W_t'. Hence the Theorem is shown for subsequence $M^{n_{k_l}}$. Since the subsequence M^{n_k} was arbitrary, Theorem 2.28 follows (cf. the Corollary on p. 337 of Billingsley (1995)). □

2.4.3 Application to Realized Volatility

Independent Times

We now turn our attention to the simplest application: the estimator from Section 2.3. Consider the normalized (by $\sqrt{n}$) error process

$$M_t^n = 2n^{1/2} \sum_{t_{i+1} \le t} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) dX_s + 2n^{1/2} \int_{t_*}^{t} (X_s - X_{t_*}) dX_s. \tag{2.43}$$

From Section 2.3.7, we have that Condition (2.40) of Theorem 2.28 is satisfied, with

$$f_t^2 = 2T\sigma_t^4 H_t'.$$

It now remains to check Condition (2.41). Note that

$$d[M^n, W^{(i)}]_t = 2n^{1/2}(X_t - X_{t_*}) d[X, W^{(i)}]_t.$$

We can now apply Lemma 2.22 with $N_t = X_t$ and $H_t = (d/dt)[X, W^{(i)}]_t$. From the Cauchy-Schwarz inequality (in this case known as the Kunita-Watanabe inequality)

$$\begin{aligned} &|[X, W^{(i)}]_{t+h} - [X, W^{(i)}]_t| \\ &\quad \le \sqrt{[X,X]_{t+h} - [X,X]_t}\sqrt{[W^{(i)}, W^{(i)}]_{t+h} - [W^{(i)}, W^{(i)}]_t} \\ &\quad \le \sqrt{\sigma_+^2 h}\sqrt{h} = \sigma_+ h \end{aligned}$$

(recall that the quadratic variation is a limit of sums of squares), so we can take $H_+ = \sigma_+$. On the other hand, $(d/dt)E[N,N]_t \le \sigma_+^2 = a(t - t_*)^b$ with $a = \sigma_+^2$ and $b = 0$.

Thus, from Lemma 2.22,

$$
\begin{aligned}
&||[M^n, W^{(i)}]_t||_1 \\
&= 2n^{1/2}|| \sum_{t_{i+1}\le t} \int_{t_i}^{t_{i+1}} (N_s - N_{t_i}) H_s ds + \int_{t_*}^{t} (N_s - N_{t_*}) H_s ds||_1 \\
&\le 2n^{1/2} \left(H_+^2 \frac{a}{b+3} R_{b+3}(\mathcal{G}) \right)^{1/2} \\
&\quad + R_{(b+3)/2}(\mathcal{G}) \frac{2}{b+3} \left(\frac{a}{b+1} \right)^{1/2} \sup_{0\le t-s\le \Delta(\mathcal{G})} ||H_s - H_t||_2 \\
&= O_p(n^{1/2} R_3(\mathcal{G})^{1/2}) + O_p(n^{1/2} R_{3/2}(\mathcal{G}) \sup_{0\le t-s\le \Delta(\mathcal{G})} ||H_s - H_t||) \\
&= o_p(1)
\end{aligned}
$$

under the conditions of Proposition 2.21, since $R_v(\mathcal{G}) = O_p(n^{1-v})$ from (2.30), and since $\sup_{0\le t-s\le\Delta(\mathcal{G})} ||H_s - H_t|| = o_p(1)$ (The latter fact is somewhat complex. One shows that one can take $W^{(1)} = W$ by a use of Lévy's theorem, and the result follows).

We have therefore shown:

Theorem 2.29 *Assume Condition 2.26, as well as the conditions of Proposition 2.21, and also that the AQVT $H(t)$ exists and is absolutely continuous. Let M_t^n be given by (2.43). Then (M_t^n) converges stably in law to M_t, given by*

$$M_t = \sqrt{2T} \int_0^t \sigma_s^2 \sqrt{H_s'} dW_s'.$$

As a special case:

Corollary 2.30 *Under the conditions of the above theorem, for fixed t,*

$$\sqrt{n}\left([X,X]_t^{\mathcal{G}_n} - [X,X]_t \right) \xrightarrow{\mathcal{L}} U \times \left(2T \int_0^t \sigma_s^4 dH_s \right)^{1/2}, \tag{2.44}$$

where U is a standard normal random variable independent of $\mathcal{F}_T$.

Similar techniques can now be used on other common estimators, such as the TSRV. We refer to Section 2.5.

In the context of equidistant times, this result goes back to Jacod (1994), Jacod and Protter (1998), and Barndorff-Nielsen and Shephard (2002). We emphasize that the method of proof in Jacod and Protter (1998) is quite different from the one used here, and gives rise to weaker conditions. The reason for our different treatment is that we have found the current framework more conducive to generalization to other observation time structures and other estimators. In the long run, it is an open question which general framework is the most useful.

Endogenous Times

The assumption of independent sampling times is not necessary for a limit result, though a weakening of conditions will change the result. To see what happens, we follow the development in Li, Mykland, Renault, Zhang, and Zheng (2009), and define the *tricicity* by $[X, X, X]_t^{\mathcal{G}} = \sum_{t_{i+1} \le t}(X_{t_{i+1}} - X_{t_i})^3 + (X_t - X_{t_*})^3$, and assume that

$$n[X, X, X, X]_t^{\mathcal{G}_n} \xrightarrow{p} U_t \text{ and } n^{1/2}[X, X, X]_t^{\mathcal{G}_n} \xrightarrow{p} V_t. \tag{2.45}$$

By the reasoning in Section 2.3.7, n and $n^{1/2}$ are the right rates for $[X, X, X, X]^{\mathcal{G}}$ and $[X, X, X]^{\mathcal{G}}$, respectively. Hence U_t and V_t will exist under reasonable regularity conditions. Also, from Section 2.3.7, if the AQVT exists and is absolutely continuous, then so are U_t and V_t. We shall use

$$U_t = \int_0^t u_s ds \text{ and } V_t = \int_0^t v_s ds. \tag{2.46}$$

Triticity is handled in much the same way as quarticity. In analogy to the development in Section 2.3.5, observe that

$$\begin{aligned} d(X_t - X_{t_i})^3 &= 3(X_t - X_{t_i})^2 dX_t + 3(X_t - X_{t_i}) d[X, X]_t \\ &= 3(X_t - X_{t_i})^2 dX_t + \frac{3}{2} d[M, X]_t, \end{aligned}$$

since $d[M, M]_t = 4(X_t - X_{t_i})^2 d[X, X]_t$. It follows that if we set

$$M_t^{(3/2)} = \sum_{t_{i+1} \le t} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i})^3 dX_s + \int_{t_*}^t (X_s - X_{t_*})^3 dX_s$$

we get

$$[X, X, X]_t^{\mathcal{G}} = \frac{3}{2}[M, X]_t + 3M_t^{(3/2)}.$$

In analogy with Proposition 2.17, we hence obtain:

Proposition 2.31 *Assume the conditions of Proposition 2.17. Then*

$$\sup_{0 \le t \le T} \mid [M, X]_t - \frac{2}{3}[X, X, X]_t^{\mathcal{G}} \mid \;=\; o_p(n^{-1/2}) \text{ as } n \to \infty.$$

It follows that unless $V_t \equiv 0$, the condition (2.41) is Theorem 2.28 will not hold. To solve this problem, normalize as in (2.43), and define an auxiliary martingale

$$\tilde{M}_t^n = M_t^n - \int_0^t g_s dX_s,$$

where g is to be determined. We now see that

$$\begin{aligned}[\tilde{M}^n, X]_t &= [M^n, X]_t - \int_0^t g_s d[X,X]_s \\ &\xrightarrow{p} \int_0^t (\frac{2}{3}v_s - g_s\sigma_s^2)ds \text{ and} \\ [\tilde{M}^n, \tilde{M}^n] &= [M^n, M^n] + \int_0^t g_s^2 d[X,X]_s - 2\int_0^t g_s d[M^n, X] \\ &\xrightarrow{p} \int_0^t (\frac{2}{3}u_s + g_s^2\sigma_s^2 - 2\frac{2}{3}g_s v_s)ds.\end{aligned}$$

Hence, if we chose $g_t = 2v_t/3\sigma_t^2$, we obtain that $[\tilde{M}^n, X]_t \xrightarrow{p} 0$ and $[\tilde{M}^n, \tilde{M}^n] \xrightarrow{p} \int_0^t (u_s - v_s\sigma_s^{-2})ds$.

By going through the same type of arguments as above, we obtain:

Theorem 2.32 *Assume Condition 2.26, as well as the conditions of Proposition 2.17. Also assume that (2.45) holds for each $t \in [0,T]$, and that the absolute continuity (2.46) holds. Then (M_t^n) converges stably in law to M_t, given by*

$$M_t = \frac{2}{3}\int_0^t \frac{v_s}{\sigma_s^2} dX_s + \int_0^t \left(\frac{2}{3}u_s - \frac{4}{9}\frac{v_s^2}{\sigma_s^2}\right)^{1/2} dW_s',$$

where W' is independent of $W^{(1)}, ..., W^{(p)}$.

Again as a special case:

Corollary 2.33 *Under the conditions of the above theorem, for fixed t,*

$$\sqrt{n}\left([X,X]_t^{\mathcal{G}_n} - [X,X]_t\right) \xrightarrow{\mathcal{L}} \frac{2}{3}\int_0^t \frac{v_s}{\sigma_s^2} dX_s + U \times \int_0^t (\frac{2}{3}u_s - \frac{4}{9}\frac{v_s^2}{\sigma_s^2})ds,$$

where U is a standard normal random variable independent of $\mathcal{F}_T$.

It is clear from this that the assumption of independent sampling times implies that $v_t \equiv 0$.

A similar result was shown in Li et al. (2009), where implications of this result are discussed further.

2.4.4 Statistical Risk Neutral Measures

We have so far ignored the drift μ_t. We shall here provide a trick to reinstate the drift in any analysis, without too much additional work. It will turn out that stable convergence is a key element in the discussion. Before we go there, we need to introduce the concept of absolute continuity.

We refer to a probability where there is no drift as a "statistical" risk neutral measure. This is in analogy to the use of equivalent measures in asset pricing. See, in particular, Ross (1976), Harrison and Kreps (1979), Harrison and Pliska (1981), Delbaen and Schachermayer (1995), and Duffie (1996).

Absolute Continuity

We shall in the following think about having two different probabilities on the same observables. For example, P can correspond to the system

$$dX_t = \sigma_t dW_t, X_0 = x_0, \tag{2.47}$$

while Q can correspond to the system

$$dX_t = \mu_t dt + \sigma_t dW_t^Q, X_0 = x_0. \tag{2.48}$$

In this case, W_t is a Wiener process under P, and W_t^Q is a Wiener process under Q. Note that since we are modeling the process X_t, this process is the observable quantity whose distribution we seek. Hence, the process X_t does not change from P to Q, but its distribution changes. If we equate (2.47) and (2.48), we get

$$\mu_t dt + \sigma_t dW_t^Q = \sigma_t dW_t,$$

or

$$\frac{\mu_t}{\sigma_t} dt + dW_t^Q = dW_t.$$

As we discussed in the constant μ and σ case, when carrying out inference for observations in a fixed time interval $[0, T]$, the process μ_t cannot be consistently estimated. A precise statement to this effect (Girsanov's Theorem) is given below.

The fact that μ cannot be observed means that one cannot fully distinguish between P and Q, even with infinite data. This concept is captured in the following definition:

Definition 2.34 *For a given σ-field $\mathcal{A}$, two probabilities P and Q are mutually absolutely continuous (or equivalent) if, for all $A \in \mathcal{A}$, $P(A) = 0 <=> Q(A) = 0$. More generally, Q is absolutely continuous with respect to P if, for all $A \in \mathcal{A}$, $P(A) = 0 => Q(A) = 0$.*

We shall see that P and Q from (2.47) and (2.48) are, indeed, mutually absolutely continuous.

The Radon-Nikodym Theorem, and the Likelihood Ratio

Theorem 2.35 *(Radon-Nikodym) Suppose that Q is absolutely continuous under P on a σ-field $\mathcal{A}$. Then there is a random variable ($\mathcal{A}$ measurable) dQ/dP so that for all $A \in \mathcal{A}$,*

$$Q(A) \;=\; E_P\left(\frac{dQ}{dP} I_A\right).$$

For proof and a more general theorem, see Theorem 32.2 (p. 422) in Billingsley (1995).

The quantity dQ/dP is usually called either the Radon-Nikodym derivative or the likelihood ratio. It is easy to see that dQ/dP is unique "almost surely" (in the same way as the conditional expectation).

Example 2.36 *The simplest case of a Radon-Nikodym derivative is where X_1, $X_2, ..., X_n$ are iid, with two possible distributions P and Q. Suppose that X_i has density f_P and f_Q under P and Q, respectively. Then*

$$\frac{dQ}{dP} \;=\; \frac{f_Q(X_1)f_Q(X_2)...f_Q(X_n)}{f_P(X_1)f_P(X_2)...f_P(X_n)}.$$

Likelihood ratios are of great importance in statistical inference generally.

Properties of Likelihood Ratios

- $P(\frac{dQ}{dP} \geq 0) = 1$
- If Q is equivalent to P: $P(\frac{dQ}{dP} > 0) = 1$
- $E_P\left(\frac{dQ}{dP}\right) = 1$
- For all $\mathcal{A}$-measurable Y: $E_Q\left(Y\right) \;=\; E_P\left(Y \frac{dQ}{dP}\right)$
- If Q is equivalent to P: $\frac{dP}{dQ} = \left(\frac{dQ}{dP}\right)^{-1}$

Girsanov's Theorem

We now get to the relationship between P and Q in systems (2.47) and (2.48). To give the generality, we consider the vector process case (where μ is a vector, and σ is a matrix). The superscript "T" here stands for "transpose".

Theorem 2.37 *(Girsanov). Subject to regularity conditions, P and Q are mutually absolutely continuous, and*

$$\frac{dP}{dQ} = \exp\left\{-\int_0^T \sigma_t^{-1}\mu_t dW_t^Q - \frac{1}{2}\int_0^T \mu_t^T(\sigma_t\sigma_t^T)^{-1}\mu_t dt\right\},$$

The regularity conditons are satisfied if $\sigma_- \leq \sigma_t \leq \sigma_+$, and $|\mu_t| \leq \mu_+$, but they also cover much more general situations. For a more general statement, see, for example, Chapter 5.5 of Karatzas and Shreve (1991).

How to get rid of μ: Interface with Stable Convergence

The idea is borrowed from asset pricing theory. We think that the true distribution is Q, but we prefer to work with P since then calculations are much simpler.

Our plan is the following: carry out the analysis under P, and adjust results back to Q using the likelihood ratio (Radon-Nikodym derivative) dP/dQ. Specifically suppose that θ is a quantity to be estimated (such as $\int_0^T \sigma_t^2 dt$, $\int_0^T \sigma_t^4 dt$, or the leverage effect). An estimator $\hat{\theta}_n$ is then found with the help of P, and an asymptotic result is established whereby, say,

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(b, a^2) \text{ stably}$$

under P. It then follows directly from the measure theoretic equivalence that $n^{1/2}(\hat{\theta}_n - \theta)$ also converges in law under Q. *In particular, consistency and rate of convergence are unaffected by the change of measure.* We emphasize that this is due to the finite (fixed) time horizon T.

The asymptotic law may be different under P and Q. While the normal distribution remains, the distributions of b and a^2 (if random) may change.

The technical result is as follows.

Proposition 2.38 *Suppose that Z_n is a sequence of random variables which converges stably to $N(b, a^2)$ under P. By this we mean that $N(b, a^2) = b + aN(0,1)$, where $N(0,1)$ is a standard normal variable independent of $\mathcal{F}_T$, also a and b are $\mathcal{F}_T$ measurable. Then Z_n converges stably in law to $b + aN(0,1)$ under Q, where $N(0,1)$ remains independent of $\mathcal{F}_T$ under Q.*

Proof of Proposition. $E_Q I_A g(Z_n) = E_P \frac{dQ}{dP} I_A g(Z_n) \to E_P \frac{dQ}{dP} I_A g(Z) = E_Q I_A g(Z)$ by uniform integrability of $\frac{dQ}{dP} I_A g(Z_n)$. □

Proposition 2.38 substantially simplifies calculations and results. In fact, the same strategy will be helpful for the localization results that come next in the paper. It will turn out that the relationship between the localized and continuous process can also be characterized by absolute continuity and likelihood ratios.

Remark 2.39 *It should be noted that after adjusting back from P to Q, the process μ_t may show up in expressions for asymptotic distributions. For instances of this, see Sections 2.5 and 4.3 of Mykland and Zhang (2009). One should always keep in mind that drift most likely is present, and may affect inference.*

Remark 2.40 *As noted, our device is comparable to the use of equivalent martingale measures in options pricing theory (Ross (1976), Harrison and Kreps (1979), Harrison and Pliska (1981), see also Duffie (1996)) in that it affords a convenient probability distribution with which to make computations. In our econometric case, one can always take the drift to be zero, while in the options pricing case, this can only be done for discounted securities prices. In both cases, however, the computational purpose is to get rid of a nuisance "dt term".*

The idea of combining stable convergence with measure change appears to go back to Rootzén (1980).

2.4.5 Unbounded σ_t

We have so far assumed that $\sigma_t^2 \leq \sigma_+^2$. With the help of stable convergence, it is also easy to weaken this assumption. One can similarly handle restrictions on μ_t, and on σ_t^2 being bounded away from zero.

The much weaker requirement is that σ_t be *locally bounded*. This is to say that there is a sequence of stopping times τ_m and of constants $\sigma_{m,+}$ so that

$$P(\tau_m < T) \to 0 \text{ as } m \to \infty \text{ and}$$
$$\sigma_t^2 \leq \sigma_{m,+}^2 \text{ for } 0 \leq t \leq \tau_m.$$

For example, this is automatically satisfied if σ_t is a continuous process.

As an illustration of how to incorporate such local boundedness in existing results, take Corollary 2.30. If we replace the condition $\sigma_t^2 \leq \sigma_+^2$ by local boundedness, the corollary continues to hold (for fixed m) with $\sigma_{\tau_n \wedge t}$ replacing σ_t. On the other hand we note that $[X, X]^{\mathcal{G}_n}$ is the same for $\sigma_{\tau_n \wedge t}$ and σ_t on the set $\{\tau_n = T\}$. Thus, the corollary tells us that for any set $A \in \mathcal{F}_T$, and for any bounded continuous function g,

$$\begin{aligned} & EI_{A\cap\{\tau_m=T\}} g\left(\sqrt{n}\left([X,X]_t^{\mathcal{G}_n} - [X,X]_t\right)\right) \\ \to \quad & EI_{A\cap\{\tau_m=T\}} g\left(U \times \left(2T\int_0^t \sigma_s^4 dH_s\right)^{1/2}\right) \end{aligned}$$

as $n \to \infty$ (and for fixed m), where U has the same meaning as in the corollary.

Hence,

$$
\begin{aligned}
&|EI_A g\left(\sqrt{n}\left([X,X]_t^{\mathcal{G}_n} - [X,X]_t\right)\right) - EI_A g\left(U \times \left(2T\int_0^t \sigma_s^4 dH_s\right)^{1/2}\right)| \\
&\quad \leq \Big| EI_{A\cap\{\tau_m = T\}} g\left(\sqrt{n}\left([X,X]_t^{\mathcal{G}_n} - [X,X]_t\right)\right) \\
&\qquad - EI_{A\cap\{\tau_m=T\}} g\left(U \times \left(2T\int_0^t \sigma_s^4 dH_s\right)^{1/2}\right)\Big| \\
&\qquad + 2\max|g(x)|P(\tau_m \neq T) \\
&\quad \to 2\max|g(x)|P(\tau_m \neq T)
\end{aligned}
$$

as $n \to \infty$. By choosing m large, the right hand sice of this expression can be made as small as we wish. Hence, the left hand side actually converges to zero. We have shown:

Corollary 2.41 *Theorem 2.29, Corollary 2.30, and Theorem 2.32 all remain true if the condition $\sigma_t^2 \leq \sigma_+^2$ is replaced by a requirement that σ_t^2 be locally bounded.*

2.5 Microstructure

2.5.1 The Problem

The basic problem is that the semimartingale X_t is actually contaminated by noise. One observes

$$Y_{t_i} = X_{t_i} + \epsilon_i. \tag{2.49}$$

We do not right now take a position on the structure of the ϵ_is.

The reason for going to this structure is that the convergence (consistency) predicted by Theorem 2.10 manifestly does not hold. To see this, in addition to $\mathcal{G}$, we also use subgrids of the form $\mathcal{H}_k = \{t_k, t_{K+k}, t_{2K+k}, ...\}$. This gives rise to the *Average Realized Voltatility (ARV)*

$$ARV(Y, \mathcal{G}, K) = \frac{1}{K}\sum_{k=1}^{K}[Y,Y]^{\mathcal{H}_k}.$$

Note that $ARV(Y, \mathcal{G}, 1) = [Y,Y]^{\mathcal{G}}$ in obvious notation. If one believes Theorem 2.10, then the $ARV(Y, \mathcal{G}, K)$ should be close for small K. In fact, the convergence in the theorem should be visible as K decreases to 1. Figure 2.1 looks at the $ARV(Y, \mathcal{G}, K)$ for Alcoa Aluminun (AA) for January 4, 2001. As can be seen in the figure, the actual data behaves quite differently from what the theory predicts. It follows that the semimartingale assumption does not hold, and we have to move to a model like (2.49).

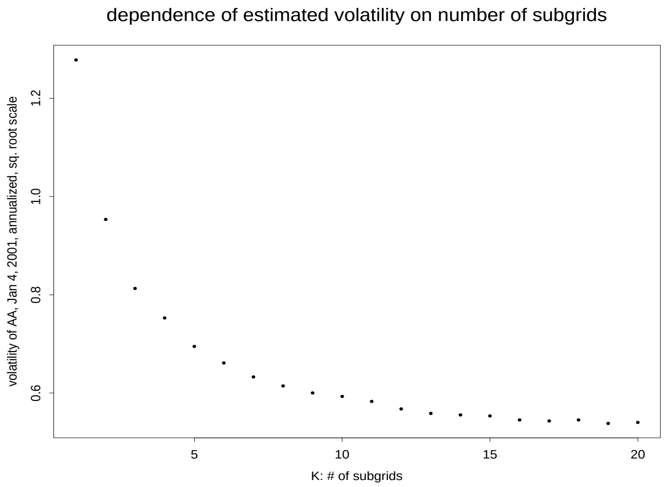

Figure 2.1 *RV as One Samples More Frequently. The plot gives $ARV(Y, \mathcal{G}, K)$ for $K = 1, ..., 20$ for Alcoa Aluminum for the transcations on January 4, 2001. It is clear that consistency does not hold for the quadratic variation. The semimartingale model, therefore, does not hold.*

2.5.2 An Initial Approach: Sparse Sampling

Plots of the type given in Figures 2.1 and 2.2 were first considered by T. G. Andersen, Bollerslev, Diebold, and Labys (2000) and called *signature plots*. The authors concluded that the most correct values for the volatility were the lower ones on the left hand side of the plot, based mainly on the stabilization of the curve in this region. On the basis of this, the authors recommended to estimate volatility using $[Y, Y]^{\mathcal{H}}$, where $\mathcal{H}$ is a sparsely sampled subgrid of $\mathcal{G}$. In this early literature, the standard approach was so subsample about every five minutes.

The philosophy behind this approach is that the size of the noise ϵ is very small, and if there are not too many sampling points, the effect of noise will be limited. While true, this uses the data inefficiently, and we shall see that better methods can be found. The basic subsampling scheme does, however, provide some guidance on how to proceed to more complex schemes. For this reason, we shall analyze its properties.

The model used for most analysis is that ϵ_i is independent of X, and iid. One can still, however, proceed under weaker conditions. For example, if the ϵ_i have serial dependence, a similar analysis will go through.

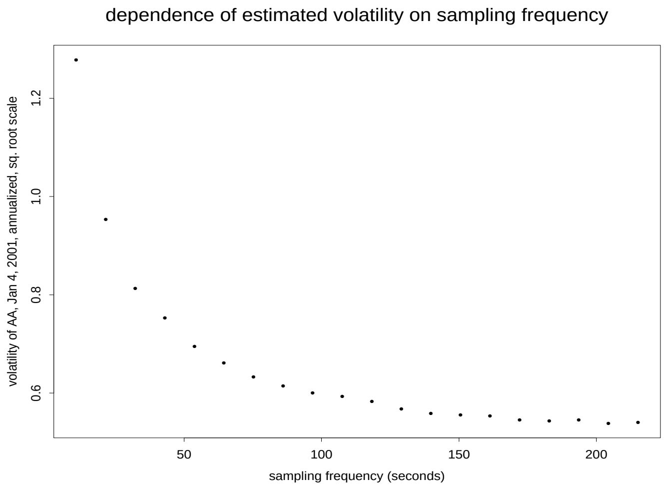

Figure 2.2 *RV as One Samples More Frequently. This is the same figure as Figure 2.1, but the x axis is the average number of observations between each transaction for each* $ARV(Y, \mathcal{G}, K)$*. There is one transaction about each 50 seconds in this particular data.*

The basic decomposition is

$$[Y, Y]^{\mathcal{H}} = [X, X]^{\mathcal{H}} + [\epsilon, \epsilon]^{\mathcal{H}} + 2[X, \epsilon]^{\mathcal{H}},$$

where the cross term is usually (but not always) ignorable. Thus, if the ϵ's are independent of X, and $E(\epsilon) = 0$, we get

$$E([Y, Y]^{\mathcal{H}} | X \text{ process}) = [X, X]^{\mathcal{H}} + E[\epsilon, \epsilon]^{\mathcal{H}}.$$

If the ϵ are identically distributed, then

$$E[\epsilon, \epsilon]^{\mathcal{H}} = n_{sparse} E(\epsilon_K - \epsilon_0)^2,$$

where $n_{sparse} = (\text{number of points in } \mathcal{H}) - 1$. Smaller n_{sparse} gives smaller bias, but bigger variance.

At this point, if you would like to follow this line of development, please consult the discussion in Section 2 in Zhang et al. (2005). This shows that there is an optimal subsampling frequency, given by equation (31) (p. 1399) in the paper. A similar analysis for $ARV(Y, \mathcal{G}, K)$ *is carried out in Section 3.1-3.3 of the paper.*

2.5.3 Two Scales Realized Volatility (TSRV)

To get a consistent estimator, we go to the *two scales realized volatility (TSRV)*. The TRSV is defined as follows.

$$\widehat{[X,X]}_T^{(\text{tsrv})} = a_n ARV(Y,\mathcal{G},K) - b_n ARV(Y,\mathcal{G},J) \tag{2.50}$$

where we shall shortly fix a_n and b_n. It will turn out to be meaningful to use

$$b_n = a_n \times \frac{\bar{n}_K}{\bar{n}_J},$$

where $\bar{n}_K = (n-K+1)/K$. For asymptotic purposes, we can take $a_n = 1$, but more generally will assume that $a_n \to 1$ as $n \to \infty$. Choices with good small sample properties are given in Section 4.2 in Zhang et al. (2005), and equation (4.22) in Aït-Sahalia, Mykland, and Zhang (2011).

This estimator is discussed in Section 4 in Zhang et al. (2005), though only in the case where $J = 1$. In the more general case, J is not necessarily 1, but $J << K$.

One can prove under weak assumptions, that

$$\sum_{i=0}^{n-J}(X_{t_{i+j}} - X_{t_i})(\epsilon_{t_{i+j}} - \epsilon_{t_i}) = O_p(J^{-1/2}).$$

This is important because it gives rise to the sum of squares decomposition

$$ARV(Y,\mathcal{G},J) = ARV(X,\mathcal{G},J) + ARV(\epsilon,\mathcal{G},J) + O_p(J^{-1/2}).$$

Thus, if we look at linear combinations of the form (2.50), one obtains, for $a_n = 1$,

$$\begin{aligned}\widehat{[X,X]}_T^{(\text{tsrv})} \\ = \;& \underbrace{ARV(X,\mathcal{G},K) - \frac{\bar{n}_K}{\bar{n}_J}ARV(X,\mathcal{G},J)}_{\text{signal term}} \\ & + \underbrace{ARV(\epsilon,\mathcal{G},K) - \frac{\bar{n}_K}{\bar{n}_J}ARV(\epsilon,\mathcal{G},J)}_{\text{noise term}} \;+\; O_p(K^{-1/2}),\end{aligned} \tag{2.51}$$

so long as

$$1 \le J \le K \quad \text{and} \quad K = o(n).$$

The noise term behaves as follows:

$$[\epsilon,\epsilon]_T^{(J)} = \frac{1}{J}\sum_{i=0}^{n} c_i^{(J)} \epsilon_{t_i}^2 - \frac{2}{J}\sum_{i=0}^{n-J} \epsilon_{t_i}\epsilon_{t_{i+J}},$$

where $c_i^{(J)} = 2$ for $J \le i \le n - J$, and $= 1$ for other i. By construction

$$\sum_i c_i^{(J)} = 2J\bar{n}_J,$$

so that

$$\text{noise term } \approx -\frac{2}{K}\sum_{i=0}^{n-K}\epsilon_{t_i}\epsilon_{t_{i+K}} + \frac{\bar{n}_K}{\bar{n}_J}\frac{2}{J}\sum_{i=0}^{n-J}\epsilon_{t_i}\epsilon_{t_{i+J}} \tag{2.52}$$

so that (1) the ϵ^2 terms have been removed, and (2) the estimator is unbiased if J is chosen to be bigger than the range of dependence of the ϵ's.

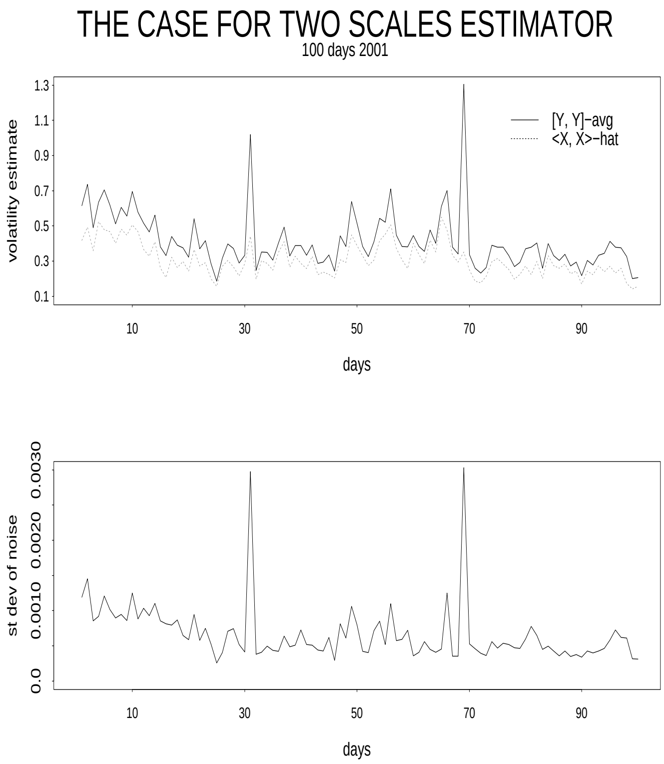

Figure 2.3 $ARV(Y, \mathcal{G}, K)$ *and the two scales estimator for Alcoa Aluminum for the first 100 trading days of 2001. Square root, annualized scale. Also estimated size of the microstructure noise. One can see from the plot that the microstructure has a substantially bigger impact on the* ARV *than on the TSRV.*

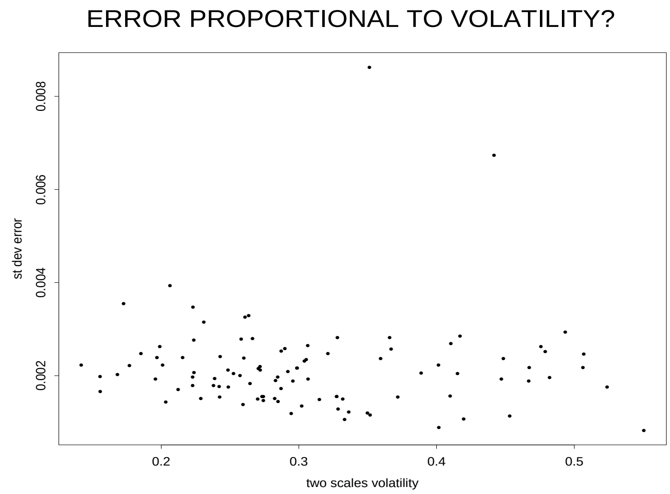

Figure 2.4 *Data as in Figure 2.3. Here size of microstructure noise plotted vs. TSRV. The figure suggests that the size of the microstructure is largely unaffected by the volatility.*

2.5.4 Asymptotics for the TSRV

If the noise is assumed to be independent of the X process, one can deal separately with the signal and noise terms in (2.51). The signal term is analyzed with the kind of technique developed in Sections 2.3-2.4 above. The precise derivation is given in Section 3.4 (p. 1400-1401) and Appendix A.3 (p. 1410-1411) of Zhang et al. (2005). Meanwhile, the noise term (2.52) is a U-statistic which can be handled with methods from discrete process limit theory, based either on martingales or mixing, using the limit theory in Hall and Heyde (1980) (see, in particular, *ibid.*, Theorem 3.2 (p. 58-59)). For concrete implementation, see Zhang et al. (2005) for the case of iid noise. By combining the two sources of error in (2.51), a rate of convergence of the TSRV to the true integrated volatility is $O_p(n^{-1/6})$, and the limit is, again, of mixed normal type.

2.5.5 The emerging literature on estimation of volatility under microstructure

This estimation problem has by now become somewhat of an industry. The following approaches are now in the process of becoming available:

- *Extensions of the two scales approach.* Zhang (2006) studies a multi scale realized volatility (MSRV), and obtains that the estimator of integrated volatility converges at the rate of $O_p(n^{-1/4})$. This rate is optimal, as it also comes up in the case where σ_t is constant and the noise is normal, which is a parametric problem. Also, the conditions on the noise structure in the TSRV has been weakened. Aït-Sahalia et al. (2011) studies noise that is internally dependent but independent of the signal, and Li and Mykland (2007) discusses a formulation where where the noise can depend on the process X.
- *An approach based on autocovariances.* Often called the "kernel" approach. This work has been pioneered by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). We refer to this paper for further discussion.
- *Preaveraging.* The idea here is to try to reduce the noise by averaging observations before computing volatilities. The two main papers for the moment are Jacod, Li, Mykland, Podolskij, and Vetter (2009) and Podolskij and Vetter (2009).
- *Likelihood based methods.* There are here two approaches under development. On the one hand, Xiu (2010) uses the likelihood function for constant σ and normal noise as a quasi-likelihood to estimate $[X, X]$. On the other hand, Mykland and Zhang (2009) show that in sufficiently small neighboorhoods of observations, one can act as if σ_t really is constant. We return to a discussion of this in Section 2.6.

To first order, all of these estimators do similar things. The main difference between them seems to be the handling of end effects. This topic will, no doubt, be the subject of future research.

2.5.6 *A wider look at subsampling and averaging*

We have seen above that subsampling and averaging can help with several problems.

1. It is a first order remedy for microstructure.
2. If an estimator $\widehat{[X, X]}$ is based on the noise being an independent series, one can ameliorate the effects of actual dependence by subsampling every J'th observation, and then average across the J resulting estimators. We have explicitly described this for the TSRV, but the same *rationale* can be used to subsample and then average any of the estimators mentioned in Section 2.5.5.
3. It should be noted that subsampling and averaging is robust to noise that depends on the latent process X, cf. Delattre and Jacod (1997) and Li and Mykland (2007).

4. A further use of subsampling is that it seems in some instances to regularize time. This is further discussed in Section 2.7.2.

2.6 Methods based on Contiguity

We have seen in Section 2.4.4 that measure changes can be a powerful tool in high frequency data problems. We here pursue this matter further, by considering measure changes that are asymptotically absolutely continuous. This is closely related to the concept of *contiguity*, which is discussed further in Section 2.6.1. This first section mainly abstracts the results in Mykland and Zhang (2009), which should be consulted for details and proofs. The later sections is new material.

2.6.1 Block Discretization

In addition to the grid $\mathcal{G}_n = \{0 = t_{n,0} < t_{n,1} < ... < t_{n,n} = T\}$, which we again take to be independent of the underlying process to be observed, we consider a subgrid

$$\mathcal{H}_n = \{0 = \tau_{n,0} < \tau_{n,1} < ... < \tau_{n,K_n} = T\} \subseteq \mathcal{G}_n.$$

We shall now define a new measure, *on the observations $X_{t_{n,j}}$ only*, for which the volatility is constant on each of the K_n blocks $(\tau_{n,i-1}, \tau_{n,i}]$. Specifically, consider the approximate measure, called Q_n, satisfying $X_0 = x_0$ and

$$\text{for each } i = 1, ..., K_n : \Delta X_{t_{j+1}} = \sigma_{\tau_{n,i-1}} \Delta W^Q_{t_{j+1}} \text{ for } t_{n,j+1} \in (\tau_{n,i-1}, \tau_{n,i}]. \tag{2.53}$$

Formally, we define the approximation Q_n recursively (with $\Delta t_{n,j} = t_{n,j} - t_{n,j-1}$).

Definition 2.42 *(Block approximation). Define the probability Q_n recursively by:*
(i) U_0 has same distribution under Q_n as under P;
(ii) The conditional Q_n-distribution of $U^{(1)}_{t_{n,j+1}}$ given $U_0, ..., U_{t_{n,j}}$ is given by (2.53), where $\Delta W^Q_{t_{j+1}}$ is conditionally normal $N(0, \Delta t_{n,j+1})$, and
(iii) The conditional P^*_n-distribution of $U^{(2)}_{t_{n,j+1}}$ given $U_0, ..., U_{t_{n,j}}, U^{(1)}_{t_{n,j+1}}$ is the same as under P.

Note that we often drop the subscript "n" on $\Delta t_{n,j}$, and write Δt_j.

Denote by $M_{n,i} = \#\{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]\}$. We shall require that $\max_i M_{n,i} = O(1)$ as $n \to \infty$, from which it follows that K_n is of exact order $O(n)$. To measure the extent to which we hold the volatility constant, we define the following

"Asymptotic Decoupling Delay" (ADD) by

$$K(t) = \lim_{n\to\infty} \sum_i \sum_{t_{n,j}\in(\tau_{n,i-1},\tau_{n,i}]\cap[0,t]} (t_{n,j} - \tau_{n,i-1}), \tag{2.54}$$

provided the limit exists. In the case of equidistant observations and equally sized blocks of M observations, the ADD takes the form $K(t) = \frac{1}{2}(M-1)t$.

It is shown in Theorems 1 and 3 in Mykland and Zhang (2009) that, subject to regularity conditions, P and Q_n are mutually absolutely continuous on the σ-field $\mathcal{X}_{n,n}$ generated by $U_{t_{n,j}}, j = 0, ..., n$. Furthermore, let $(dP/dQ_n)(U_{t_0}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}})$ be the likelihood ratio (Radon-Nikodym derivative) on $\mathcal{X}_{n,n}$. Then,

$$\frac{dP}{dQ_n}(U_{t_0}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}}) \xrightarrow{\mathcal{L}} \exp\{\Gamma^{1/2} N(0,1) - \frac{1}{2}\Gamma\} \tag{2.55}$$

stably in law, under Q_n, as $n \to \infty$. The asymptotic variance is given by $\Gamma = \Gamma_0 + \Gamma_1$, where

$$\begin{aligned} \Gamma_0 &= \frac{3}{8}\int_0^T \left(\frac{1}{\sigma_t^2}\frac{d}{dt}[\sigma^2, W]_t\right)^2 dt = \frac{3}{2}\int_0^T \left(\frac{1}{\sigma_t}\frac{d}{dt}[\sigma, W]_t\right)^2 dt, \text{ and} \\ \Gamma_1 &= \frac{1}{2}\int_0^T \frac{1}{\sigma_t^4}\left(\frac{d}{dt}[\sigma^2,\sigma^2]_t\right) dK(t) = 2\int_0^T \frac{1}{\sigma_t^2}\left(\frac{d}{dt}[\sigma,\sigma]_t\right) dK(t). \end{aligned} \tag{2.56}$$

Hence, Γ_0 is related to the leverage effect, while Γ_1 is related to the volatility of volatility.

The important consequence of this result is that that P and the approximation Q_n are *contiguous* in the sense of Hájek and Sidak (1967) (Chapter VI), LeCam (1986), LeCam and Yang (2000), and Jacod and Shiryaev (2003) (Chapter IV). This is to say that for a sequence A_n of sets, $P(A_n) \to 0$ if and only if $Q_n(A_n) \to 0$. This follows from (2.55) since dP/dQ_n is uniformly integrable under Q_n (since the sequence dP/dQ_n is nonnegative, also the limit integrates to one under Q_n). In it follows that consistency and orders of convergence are maintained from one measure to the other. In particular, a martingale Z^n is $O_p(1)$ under P if and only if it has the same order of convergence under Q_n. Hence consistency and order of convergence are maintained from one measure to the other. The result in Mykland and Zhang (2009) also covers the multivariate case. Also consult *ibid.*, Section 3.3, for connections to Hermite polynomials. Finally note that contiguity also holds if the sequence $\frac{dP}{dQ_n}$ is tight, which requires even weaker conditions. (For example, K need only exist through subsequences, which is assured under weak conditions by Helly's Theorem, (see, for example, p. 336 in Billingsley (1995)).

The good news: This means that one can construct estimators *as if* σ_t^2 is locally constant. If the resulting estimator $\hat{\theta}_n$ is such that $n^\alpha(\hat{\theta}_n - \theta) = O_p(1)$ under

Q_n (where local constancy is satisfied), then $n^\alpha(\hat{\theta}_n - \theta) = O_p(1)$ also under P. In other words, the change from P to Q_n has much the same simplifying function as the measure change in Section 2.4.4.

Remark 2.43 *The (potentially) bad news: This change of measure is not completely innocuous. A sequence Z^n of martingales may not have exactly the same limit distribution under P and Q_n. The reason is that the Q_n martingale part of $\log dP/dQ_n$ may have nonzero asymptotic covariation with Z^n. This is the same phenomenon which occurs (in a different context) in Section 2.4.3. An adjustment then has to be carried out along the lines of "LeCam's Third Lemma" (Hájek and Sidak (1967), Chapter VI.1.4., p. 208). We refer to Section 2.4 and 3.4 of Mykland and Zhang (2009) for the methodology for adjusting the limit distribution. Ibid., Section 4.3, provides an example where such an adjustment actually has to be made. Other parts of Section 4 of the paper provides examples of how the methodology can be used, and where adjustment is not necessary.*

The approximation above depends on the following Itô process structure on σ:

$$d\sigma_t = \tilde{\sigma}_t dt + f_t dW_t + g_t dB_t, \tag{2.57}$$

where B a Brownian motion independent of W. (It is an open question what happens in, say, the long range dependent case). We also require $\inf_{0\leq t\leq T} \sigma^2 > 0$. Note that in the representation (2.57), (2.56) becomes

$$\Gamma_0 \;=\; \frac{3}{2}\int_0^T \left(\frac{f_t}{\sigma_t}\right)^2 dt \;\text{ and }\; \Gamma_1 \;=\; 2\int_0^T \left(\frac{f_t^2 + g_t^2}{\sigma_t^2}\right) dt.$$

Example 2.44 *In the case of a Heston model (Section 2.2.2), we obtain that*

$$\Gamma_0 = \frac{3}{8}(\rho\gamma)^2 \int_0^T \sigma_t^{-2} dt \;\text{ and }\; \Gamma_1 = \frac{1}{4}\gamma^2(M-1)\int_0^T \sigma_t^{-2} dt.$$

Remark 2.45 *(One step discretization). Let P_n^* be the measure Q_n which arises when the block length is $M = 1$. Observe that even with this one-step discretization, dP/dP_n^* does not necessarily converge to unity. In this case, $\Gamma_1 = 0$, but Γ_0 does not vanish when there is leverage effect.*

2.6.2 Moving windows

The paper so far has considered chopping n data up into non-overlapping windows of size M each. We here show by example that the methodology can be adapted to the moving window case. We consider the estimation of $\theta = \int_0^T |\sigma_t|^p dt$, as in Section 4.1 of Mykland and Zhang (2009). It should be noted that the moving window is close to the concept of a moving kernel, and this may be a promising avenue of further investigation. See, in particular, Linton (2007).

We use block length M, and equidistant times with spacing $\Delta t_n = T/n$. Also, we use for simplicity

$$\tilde{\sigma}^2_{\tau_{n,i}} = \frac{1}{\Delta t_n M_n} \sum_{t_{n,j} \in (\tau_{n,i}, \tau_{n,i+1}]} (\Delta X_{t_{n,j+1}})^2,$$

as estimator of $\sigma^2_{\tau_{n,i}}$. The moving window estimate of θ is now

$$\tilde{\theta}_n^{MW} = (\Delta t) \sum_{i=0}^{n-M} \widetilde{|\sigma_{t_{n,i}}|^r}.$$

It is easy to see that

$$\tilde{\theta}_n^{MW} = \frac{1}{M} \sum_{m=1}^{M} \tilde{\theta}_{n,m} + O_p(n^{-1}),$$

where $\tilde{\theta}_{n,m}$ is the non-overlapping block estimator, with block number one starting at $t_{n,m}$. In view of this representation, it is once again clear from sufficiency considerations that the moving window estimator will have an asymptotic variance which is smaller (or, at least, no larger) than the estimator based on non-overlapping blocks. We now carry out the precise asymptotic analysis.

To analyze this estimator, let $\mathcal{M} > M$, and let

$$A_n = \{i = 0, ..., n - M : \ [t_{n,i}, t_{n,i+M}] \subseteq [k\mathcal{M}, (k+1)\mathcal{M}] \text{ for some } k\},$$

with $B_n = \{0, ..., n - M\} - A_n$. Write

$$\begin{aligned} & n^{1/2}(\tilde{\theta}_n^{MW} - \theta) \\ = \ & n^{1/2} \Delta t \sum_k \sum_{i:[t_{n,i}, t_{n,i+M}] \subseteq [k\mathcal{M}/n, (k+1)\mathcal{M}/n]} (\widetilde{|\sigma_{t_{n,i}}|^r} - |\sigma_{t_{k\mathcal{M}}}|^r) \\ & + n^{1/2} \Delta t \sum_{i \in B_n} (\widetilde{|\sigma_{t_{n,i}}|^r} - |\sigma_{t_{n,i}}|^r) + O_p(n^{-1/2}). \end{aligned} \tag{2.58}$$

Now apply our methodology from Section 2.6.1, *with block size* $\mathcal{M}$, to the first term in (2.58). Under this block approximation, the inner sum in the first term is based on conditionally i.i.d. observations, in fact, for $[t_{n,i}, t_{n,i+M}] \subseteq [k\mathcal{M}/n, (k+1)\mathcal{M}/n]$, $\tilde{\sigma}^2_{t_{n,i}} = \sigma^2_{k\mathcal{M}/n} S_i$, in law, where

$$S_i = M^{-1} \sum_{j=i}^{i+M-1} U_j^2, \quad U_0, U_1, U_2, ... \text{ iid standard normal.} \tag{2.59}$$

As in Section 4.1 of Mykland and Zhang (2009), there is no adjustment (*à* *la* Remark 2.43) due to covariation with the asymptotic likelihood ratios, and so the first term in (2.58) converges stably to a mixed normal with random variance as the limit of $n\Delta t_n^2 \sum_k |\sigma|^r_{k\mathcal{M}/n} \mathrm{Var}\left(\mathrm{c}^{-1}_{\mathrm{M,r}} \sum_{\mathrm{i}=0}^{\mathcal{M}-\mathrm{M}} \mathrm{S}_{\mathrm{i}}^{\mathrm{r}/2}\right)$, which is

$$T c^{-2}_{M,r} \frac{1}{\mathcal{M}} \mathrm{Var}\left(\sum_{\mathrm{i}=0}^{\mathcal{M}-\mathrm{M}} \mathrm{S}_{\mathrm{i}}^{\mathrm{r}/2}\right) \int_0^{\mathrm{T}} |\sigma|^{\mathrm{r}}_{\mathrm{t}} \mathrm{dt}. \tag{2.60}$$

Similarly, one can apply the same technique to the second term in (2.58), but now with the k'th block ($k \geq 2$) starting at $k\mathcal{M} - M$. This analysis yields that the second term is also asymptotically mixed normal, but with a variance that is of order $o_p(1)$ as $\mathcal{M} \to \infty$. (In other words, once again, first send n to infinity, and then, afterwards, do the same to $\mathcal{M}$). This yields that, overall, and in the sense of stable convergence,

$$n^{1/2}(\tilde{\theta}_n^{MW} - \theta) \xrightarrow{\mathcal{L}} N(0,1) \times \left(c^{-2}_{M,r} \alpha_{M,r} T \int_0^T |\sigma|^r_t dt\right)^{1/2},$$

where, from (2.60), $\alpha_{M,r} = \lim_{\mathcal{M}\to\infty} \mathrm{Var}\left(\sum_{\mathrm{i}=0}^{\mathcal{M}-\mathrm{M}} \mathrm{S}_{\mathrm{i}}^{\mathrm{r}/2}\right)/\mathcal{M}$, *i.e.*,

$$\alpha_{M,r} = \mathrm{Var}(\mathrm{S}_0^{\mathrm{r}/2}) + 2\sum_{\mathrm{i}=1}^{\mathrm{M}-1} \mathrm{Cov}(\mathrm{S}_0^{\mathrm{r}/2}, \mathrm{S}_{\mathrm{i}}^{\mathrm{r}/2}),$$

where the S_i are given in (2.59).

2.6.3 Multivariate and Asynchronous data

The results discussed in Section 2.6.1 also apply to vector processes (see Mykland and Zhang (2009) for details). Also, for purposes of analysis, asynchronous data does not pose any conceptual difficulty when applying the results. One includes all observation times when computing the likelihood ratios in the contiguity theorems. It does not matter that some components of the vector are not observed at all these times. In a sense, they are just treated as missing data. Just as in the case of irregular times for scalar processes, this does not necessarily mean that it is straightforward to write down sensible estimators.

For example, consider a bivariate process $(X_t^{(1)}, X_t^{(2)})$. If the process $(X_t^{(r)})$ is observed at times:

$$\mathcal{G}_n^{(r)} = \{0 \leq t_{n,0}^{(r)} < t_{n,1}^{(r)} < \ldots < t_{n,n_r}^{(r)} \leq T\}, \tag{2.61}$$

one would normally use the grid $\mathcal{G}_n = \mathcal{G}_n^{(1)} \cup \mathcal{G}_n^{(2)} \cup \{0, T\}$ to compute the likelihood ratio dP/dQ_n.

To focus the mind with an example, consider the estimation of covariation under asynchronous data. It is shown in Mykland (2010a) that the Hayashi-Yoshida estimator (Hayashi and Yoshida (2005)) can be seen as a nonparametric maximum likelihood estimator (MLE). We shall here see that blocking induces an additional class of local likelihood based MLEs. The difference between the former and the latter depends on the continuity assumptions made on the volatility process, and is a little like the difference between the Kaplan-Meier (Kaplan and Meier (1958)) and Nelson-Aalen (Nelson (1969), Aalen (1976, 1978)) estimators in survival analysis. (Note that the variance estimate for the Haysahi-Yoshida estimator from Section 5.3 of Mykland (2010a) obviously also remains valid in the setting of this paper).

For simplicity, work with a bivariate process, and let the grid $\mathcal{G}_n$ be given by (2.61). For now, let the block dividers τ be any subset of $\mathcal{G}_n$. Under the approximate measure Q_n, note that for

$$\tau_{n,i-1} \le t^{(1)}_{n,j-1} < t^{(1)}_{n,j} \le \tau_{n,i} \text{ and } \tau_{n,i-1} \le t^{(2)}_{k-1} < t^{(2)}_{k} \le \tau_{n,i} \tag{2.62}$$

the set of returns $X^{(1)}_{t^{(1)}_{n,j}} - X^{(1)}_{t^{(1)}_{n,j-1}}$ and $X^{(2)}_{t^{(2)}_{n,k}} - X^{(2)}_{t^{(2)}_{n,k-1}}$ are conditionally jointly normal with mean zero and covariances

$$\begin{aligned}&\mathrm{Cov}_{\mathrm{Q_n}}((\mathrm{X}^{(\mathrm{r})}_{\mathrm{t}^{(\mathrm{r})}_{\mathrm{n,j}}} - \mathrm{X}^{(\mathrm{r})}_{\mathrm{t}^{(\mathrm{r})}_{\mathrm{n,j-1}}}, \mathrm{X}^{(\mathrm{s})}_{\mathrm{t}^{(\mathrm{s})}_{\mathrm{n,k}}} - \mathrm{X}^{(\mathrm{s})}_{\mathrm{t}^{(\mathrm{s})}_{\mathrm{n,k-1}}}) \mid \mathcal{F}_{\tau_{\mathrm{n,i-1}}})\\ &= \quad (\zeta_{\tau_{n,i-1}})_{r,s} d\{(t^{(r)}_{n,j-1}, t^{(r)}_{n,j}) \cap (t^{(s)}_{n,k-1}, t^{(s)}_{n,k})\}\end{aligned}$$

where d is length (Lebesgue measure). Set

$$\kappa_{r,s;j,k} = \zeta d\{(t^{(r)}_{n,j-1}, t^{(r)}_{n,j}) \cap (t^{(s)}_{n,k-1}, t^{(s)}_{n,k})\}.$$

The Q_n log likelihood ratio based on observations fully in block $(\tau_{n,i-1}, \tau_{n,i}]$ is therefore given as

$$\begin{aligned}\ell(\zeta) \quad = \quad & -\frac{1}{2}\ln\det(\kappa)\\ & -\frac{1}{2}\sum_{r,s,j,k} \kappa^{r,s;j,k}(X^{(r)}_{t^{(r)}_{n,j}} - X^{(r)}_{t^{(r)}_{n,j-1}})(X^{(s)}_{t^{(s)}_{n,k}} - X^{(s)}_{t^{(s)}_{n,k-1}})\\ & -\frac{N_i}{2}\ln(2\pi),\end{aligned}$$

where $\kappa^{r,s;j,k}$ are the elements of the matrix inverse of $(\kappa_{r,s;j,k})$, and N_i is a measure of block sample size. The sum in (j,k) is over all intersections $(t^{(r)}_{n,j-1}, t^{(r)}_{n,j}) \cap (t^{(s)}_{n,k-1}, t^{(s)}_{n,k})$ with positive length satisfying (2.62). Call the number of such terms

$$\begin{aligned}m^{(r,s)}_{n,i} \quad = \quad & \#\text{ nonempty intersections } (t^{(r)}_{n,j-1}, t^{(r)}_{n,j}) \cap (t^{(s)}_{n,k-1}, t^{(s)}_{n,k})\\ & \text{satisfying (2.62) .}\end{aligned}$$

The "parameter" ζ corresponds to $\zeta_{\tau_{n,i-1}}$. The block MLE is thus given as

$$\hat{\zeta}^{(r,s)}_{\tau_{n,i-1}} = \frac{1}{m^{(r,s)}_{n,i}} \sum_{j,k} \frac{(X^{(r)}_{t^{(r)}_{n,j}} - X^{(r)}_{t^{(r)}_{n,j-1}})(X^{(s)}_{t^{(s)}_{n,k}} - X^{(s)}_{t^{(s)}_{n,k-1}})}{d\{(t^{(r)}_{n,j-1}, t^{(r)}_{n,j}) \cap (t^{(s)}_{n,k-1}, t^{(s)}_{n,k})\}} \tag{2.63}$$

where the sum is over j, k satisfying (2.62) for which the denominator in the summand is nonzero. The overall estimate of covariation is thus

$$\widehat{\langle X^{(r)}, X^{(s)} \rangle}_T = \sum_i \hat{\zeta}^{(r,s)}_{\tau_{n,i-1}} (\tau_{n,i} - \tau_{n,i-1}).$$

We suppose, of course, that each block is large enough for $m^{(r,s)}_{n,i}$ to be always greater than zero.

Under Q_n, $E_{Q_n}(\hat{\zeta}_{\tau_{n,i-1}} | \mathcal{F}_{\tau_{n,i-1}}) = \zeta_{\tau_{n,i-1}}$, and

$$\begin{aligned}
&\mathrm{Var}_{\mathrm{Q_n}}(\hat{\zeta}^{(\mathrm{r,s})}_{\tau_{\mathrm{n,i-1}}} | \mathcal{F}_{\tau_{\mathrm{n,i-1}}}) \\
&= \left(\frac{1}{m^{(r,s)}_{n,i}}\right)^2 \left(\zeta^{(r,r)}_{\tau_{n,i-1}} \zeta^{(s,s)}_{\tau_{n,i-1}} \sum_{j,k} \frac{(t^{(r)}_{n,j} - t^{(r)}_{n,j-1})(t^{(s)}_{n,k} - t^{(s)}_{n,k-1})}{d\{(t^{(r)}_{n,j-1}, t^{(r)}_{n,j}) \cap (t^{(s)}_{n,k-1}, t^{(s)}_{n,k})\}^2} \right. \\
&\quad + (\zeta^{(r,s)}_{\tau_{n,i-1}})^2 \sum_{j_1, j_2, k_1, k_2} \left(\frac{d\{(t^{(r)}_{n,j_1-1}, t^{(r)}_{n,j_1}) \cap (t^{(s)}_{n,k_2-1}, t^{(s)}_{n,k_2})\}}{d\{(t^{(r)}_{n,j_1-1}, t^{(r)}_{n,j_1}) \cap (t^{(s)}_{n,k_1-1}, t^{(s)}_{n,k_1})\}} \right. \\
&\quad \left.\left. \times \frac{d\{(t^{(r)}_{n,j_2-1}, t^{(r)}_{n,j_2}) \cap (t^{(s)}_{n,k_1-1}, t^{(s)}_{n,k_1})\}}{d\{(t^{(r)}_{n,j_2-1}, t^{(r)}_{n,j_2}) \cap (t^{(s)}_{n,k_2-1}, t^{(s)}_{n,k_2})\}} \right) \right).
\end{aligned} \tag{2.64}$$

The first sum is over the same (j, k) as in (2.63), and the second sum is over all j_1, j_2, k_1, k_2 satisfying (2.62), again for which the denominator in the summand is nonzero.

It is therefore easy to see that subject to conditions on the observation times $t^{(r)}_{n,i}$ and $t^{(s)}_{n,i}$, $n^{1/2}(\widehat{\langle X^{(r)}, X^{(s)} \rangle}_T - \langle X^{(r)}, X^{(s)} \rangle_T)$ converges stably (under Q_n), to a mixed normal distribution with variance as the limit of

$$n \sum_i \mathrm{Var}_{\mathrm{Q_n}}(\hat{\zeta}^{(\mathrm{r,s})}_{\tau_{\mathrm{n,i-1}}} | \mathcal{F}_{\tau_{\mathrm{n,i-1}}})(\tau_{\mathrm{n,i}} - \tau_{\mathrm{n,i-1}})^2. \tag{2.65}$$

It is straightforward to see that there is no adjustment from Q_n to P. A formal asymptotic analysis would be tedious, and has therefore been omitted. In any case, to estimate the asymptotic variance, one would use (2.64)-(2.65), with $\hat{\zeta}_{\tau_{n,i-1}}$ replacing $\zeta_{\tau_{n,i-1}}$ in (2.64).

Remark 2.46 *An important difference from the Hayashi-Yoshida estimator is that (2.63) depends on the observation times. This is in many instances undesirable, and the choice of estimator will depend on the degree to which these times are trusted. The Hayashi-Yoshida estimator is also aesthetically more* *pleasing. We note, however, that from likelihood considerations, the estimator (2.63) will have an asymptotic variance which, as the block size tends to infinity, converges to a limit which corresponds to the efficient minimum for constant volatility matrix.*

This phenomenon can be best illustrated for a scalar process (so there is no asynchronicity). In this case, our estimator (2.63) of $\langle X, X\rangle_T$ becomes (for block size M fixed)

$$\widehat{\langle X, X\rangle}_T = \sum_i (\tau_{n,i} - \tau_{n,i-1}) \frac{1}{M} \sum_{j:\ \tau_{n,i-1} < t_{n,j} \leq \tau_{n,i}} \frac{\Delta X^2_{t_{n,j}}}{\Delta t_{n,j}}. \tag{2.66}$$

It is easy to see, by the methods in this paper, or directly, that for this estimator, the asymptotic variance is $2T\int_0^T \sigma_t^4 dt$, while for the standard realized volatility, the corresponding expression is $2T\int_0^T \sigma_t^4 H'(t)dt$, where $H(t)$ is the asymptotic quadratic variation of time (2.39). It is always the case that $H'(t) \geq 1$, and when observations are sufficiently irregular (under, say, Poisson sampling), the inequality is strict, cf. Section 2.7.2 below. Thus, (2.66) is more efficient than regular realized volatility, but since the times can in many cases not be trusted, the realized volatility remains a main tool for estimating volatility.

2.6.4 More complicated data generating mechanisms

Jumps

We only consider the case of finitely many jumps (compound Poisson processes, and similar). The conceptually simplest approach is to remove these jumps using the kind of procedure described in Mancini (2001) and Lee and Mykland (2006). The procedure will detect all intervals $(t_{n,j-1}, t_{n,j}]$, with probability tending to one (exponentially fast) as $n \to \infty$. If one simply removes the detected intervals from the analysis, it is easy to see that our asymptotic results go through unchanged.

The case of infinitely many jumps is more complicated, and beyond the scope of this paper.

Note that there is a range of approaches for estimating the continuous part of volatility in such data. Methods include bi- and multi-power (Barndorff-Nielsen and Shephard (2004b)). Other devices are considered by Aït-Sahalia (2004), and Aït-Sahalia and Jacod (2007). One can use our method of analysis for all of these approaches.

Microstructure noise

The presence of noise does not alter the analysis in any major way. Suppose one observes

$$Y_{t_{n,j}} \;=\; X_{t_{n,j}} + \epsilon_{n,j}$$

where the $\epsilon_{n,j}$'s are independent of the (X_t) process. The latter still follows (2.25). We take the σ-field $\mathcal{X}_{n,n}$ to be generated by $\{X_{t_{n,j}}, \epsilon_{n,j},\ 0 \le j \le n\}$. Suppose that P_1 and P_2 are two measures on $\mathcal{X}_{n,n}$ for which: (1) the variables $\{\epsilon_{n,j},\ 0 \le j \le n\}$ are independent of $\{X_{t_{n,j}},\ 0 \le j \le n\}$, and (2) the variables $\{\epsilon_{n,j},\ 0 \le j \le n\}$ have the same distribution under P_1 and P_2. Then, from standard results in measure theory,

$$\frac{dP_2}{dP_1}\left((X_{t_{n,j}}, \epsilon_{n,j}),\ 0 \le j \le n\right) \;=\; \frac{dP_2}{dP_1}\left(X_{t_{n,j}},\ 0 \le j \le n\right).$$

The results in our theorems are therefore unchanged in the case of microstructure noise (unless one also wants to change the probability distribution of the noise). We note that this remains the case irrespective of the internal dependence structure of the noise.

The key observation which leads to this easy extension is that it is not required for our results to work that the observables $Y_{t_{n,j}}$ generate the σ-field $\mathcal{X}_{n,n}$. It is only required that the observables be measurable with respect to this σ-field. The same principle was invoked in Section 2.6.3.

The extension does not, obviously, solve all problems relating to microstructure noise, since this type of data generating mechanism is best treated with an asymptotics where $M \to \infty$ as $n \to \infty$. This is currently under investigation. For one approach, see Mykland and Zhang (2011).

2.7 Irregularly spaced data

2.7.1 A second block approximation.

The approximation in Section 2.6.1 reduces the problem (in each block) to a case of independent (but not identically distributed) increments. Can we do better than this, and go to iid observations? We here give a criterion for this to be the case.

Consider yet another approximate probability measure R_n, under which $X_0 = x_0$, and

$$\text{for each } i = 1, ..., K_n: \quad \Delta X_{t_{j+1}} = \sigma_{\tau_{n,i-1}} \left(\frac{\Delta\tau_i}{\Delta t_{j+1} M_i}\right)^{1/2} \Delta W^*_{t_{j+1}}$$
$$\text{for } t_{n,j+1} \in (\tau_{n,i-1}, \tau_{n,i}]. \tag{2.67}$$

Formally, we define the approximation as follows.

Definition 2.47 *R_n is defined as Q_n in Definition 2.42, but with (2.67) replacing (2.53).*

The crucial fact will be that under R_n, the observables $\Delta X_{t_{j+1}}$ are conditionally iid $N(0, \zeta_{\tau_{n,i-1}} \Delta\tau_i / M_i)$ for $t_{n,j+1} \in (\tau_{n,i-1}, \tau_{n,i}]$.

So that the following can be used together with the results in Mykland and Zhang (2009), we will in the following let the process X be multivariate, and we make the following assumption

Condition 2.48 *(Structure of the instantaneous volatility). We assume that the matrix process σ_t is itself an Itô processes, and that if $\lambda_t^{(p)}$ is the smallest eigenvalue of σ_t, then $\inf_t \lambda_t^{(p)} > 0$ a.s.*

The contiguity question is then addressed as follows. Let P_n^* be the measure from Remark 2.45 (corresponding to block length $M = 1$). Recall that

$$\log \frac{dR_n}{dP} = \log \frac{dR_n}{dQ_n} + \log \frac{dQ_n}{dP_n^*} + \log \frac{dP_n^*}{dP}.$$

Define

$$B_{n,j} = \left(\Delta t_{n,j+1} \left(\frac{\Delta\tau_{n,i}}{M_i} \right)^{-1} - 1 \right).$$

Theorem 2.49 *(Asymptotic relationship between P_n^*, Q_n and R_n). Assume the conditions of Theorem 4 in Mykland and Zhang (2009), and let $Z_n^{(1)}$ and $M_n^{(1)}$ be as in that theorem (see (2.71) and (2.74) in Section 2.7.3). Assume that the following limits exist:*

$$\Gamma_2 = \frac{p}{2} \lim_{n\to\infty} \sum_j B_{n,j}^2 \text{ and } \Gamma_3 = \frac{p}{2} \lim_{n\to\infty} \sum_j \log(1 + B_{n,j}).$$

Set

$$\begin{aligned} Z_n^{(2)} &= \frac{1}{2} \sum_i \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} \Delta X_{t_{n,j}}^T ((\sigma\sigma^T)_{\tau_{n,i-1}}^{-1}) \Delta X_{t_{n,j}} \\ &\qquad \times \left(\Delta t_{n,j+1}^{-1} - \left(\frac{\Delta\tau_{n,i}}{M_i} \right)^{-1} \right), \end{aligned}$$

and let $M_n^{(2)}$ be the end point of the martingale part of $Z_n^{(2)}$ (see (2.72) and (2.74) in Section 2.7.3 for the explicit formula). Then, as $n \to \infty$, $(M_n^{(1)}, M_n^{(2)})$ converges stably in law under P^* to a normal distribution with mean zero and diagonal variance matrix with diagonal elements Γ_1 and Γ_2. Also, under P^*,

$$\log \frac{dR_n}{dQ_n} = M_n^{(2)} + \Gamma_3 + o_p(1).$$

The theorem can be viewed from the angle of contiguity:

Corollary 2.50 *Under regularity conditions, the following statements are equivalent, as $n \to \infty$:*
(i) R_n is contiguous to P.
(ii) R_n is contiguous to Q_n.
(iii) The following relationship holds:

$$\Gamma_3 = -\frac{1}{2}\Gamma_2. \tag{2.68}$$

As we shall see, the requirement (2.68) is a substantial restriction. Corollary 2.50 says that unlike the case of Q_n, inference under R_n may not give rise to desired results. Part of the probability mass under Q_n (and hence P) is not preserved under R_n.

To understand the requirement (2.68), note that

$$\frac{p}{2}\sum_j \log(1 + B_{n,j}) = -\frac{p}{4}\sum_j B_{n,j}^2 + \frac{p}{6}\sum_j B_{n,j}^3 - ...$$

since $\sum_j B_{n,j} = 0$. Hence, (2.68) will, for example, be satisfied if $\max_j |B_{n,j}| \to 0$ as $n \to \infty$. One such example is

$$t_{n,j} = f(j/n) \text{ and } f \text{ is continuously differentiable.} \tag{2.69}$$

However, (2.69) will not hold in more general settings, as we shall see from the following examples.

Example 2.51 (POISSON SAMPLING.) *Suppose that the sampling time points follow a Poisson process with parameter λ. If one conditions on the number of sampling points n, these points behave like the order statistics of n uniformly distributed random variables (see, for example, Chapter 2.3 in S. Ross (1996)). Consider the case where $M_i = M$ for all but (possibly) the last interval in $\mathcal{H}_n$. In this case, K_n is the smallest integer larger than or equal to n/M. Let Y_i be the M-tuple $(B_j, \tau_{i-1} \le t_j < \tau_i)$.*

We now obtain, by passing between the conditional and unconditional, that $Y_1, ..., Y_{K_n - 1}$ are iid, and the distribution can be described by

$$Y_1 = M(U_{(1)}, U_{(2)} - U_{(1)}, ..., U_{(M-1)} - U_{(M-2)}, 1 - U_{(M-1)}) - 1,$$

where $U_{(1)}, ..., U_{(M-1)}$ is the order statistic of $M - 1$ independent uniform random variables on $(0, 1)$. It follows that

$$\sum_j B_{n,j}^2 = \frac{n}{M}(M^2 E U_{(1)}^2 - 1) + o_p(n)$$

$$\sum_j \log(1 + B_{n,j}) = \frac{n}{M} E \log(M U_{(1)}) + o_p(n)$$

since $EU_{(1)}^2 = 2/(M+1)(M+2)$. Hence, both Γ_2 and Γ_3 are infinite. The contiguity between R_n and the other probabilities fails. On the other hand all our assumptions up to Section 2.6 are satisfied, and so P, P_n^* and Q_n are all contiguous. The AQVT (equation (2.39)) is given by $H(t) = 2t$. Also, if the block size is constant (size M), the ADD is $K(t) = (M-1)t$.

Example 2.52 (SYSTEMATIC IRREGULARITY.) *Let ϵ be a small positive number, and let $\Delta t_{n,j} = (1+\epsilon)T/n$ for odd j and $\Delta t_{n,j} = (1-\epsilon)T/n$ for even j (with $\Delta t_{n,n} = T/n$ for odd n). Again, all our assumptions up to Section 2.6 are satisfied. The AQVT is given by $H(t) = t(1+\epsilon^2)$. If we suppose that all $M_i = 2$, the ADD becomes $K(t) = t$. On the other hand, $B_{n,j} = \pm\epsilon$, so that, again, both Γ_2 and Γ_3 are infinite. The contiguity between R_n and the other probabilities thus fails in the same radical fashion as in the case of Poisson sampling.*

2.7.2 Irregular Spacing and Subsampling

We here return to a more direct study of the effect of irregular spacings. We put ourselves in the situation from Section 2.4.3, where observation times are independent of the process. As stated in equation (2.44), the limit law for the realized volatility (for $\sqrt{n}\left([X,X]_t^{\mathcal{G}_n} - [X,X]_t\right)$) is mixed normal with (random) variance

$$2T\int_0^t \sigma_s^4 dH_s,$$

where H is the asymptotic quadratic variation of time (AQVT). When observations are equidistant, $H'(t) \equiv 1$. From the preceeding section, we also know that if times are on the form (2.69), the asymptotic variance is unaffected. It is worth elaborating on this in direct computation. Set

$$F(t) = \lim_{n\to\infty} \frac{1}{n}\#\{t_{n,i+1} \le t\}.$$

This quantity exists, if necessary by going through subsequences (Helly's Theorem, see, for example, p. 336 in Billingsley (1995)). Set

$$u_{n,i} = F(t_{n,i}). \qquad (2.70)$$

Asymptotically, the $u_{n,i}$ are equispaced:

$$\frac{1}{n}\#\{u_{n,i+1} \le t\} = \frac{1}{n}\#\{t_{n,i+1} \le F^{(-1)}(t)\} \to F(F^{(-1)}(t)) = t$$

Inference is invariant to this transformation: Observing the process X_t at times $t_{n,i}$ is the same as observing the process $Y_t = X_{F^{(-1)}(t)}$ at times $u_{n,i}$. If we set $\mathcal{U} = \{u_{n,j},\ j = 0, ..., n\}$, then $[X,X]_T^{\mathcal{G}} = [Y,Y]_T^{\mathcal{U}}$. Also, in the limit,

$[X, X]_T = [Y, Y]_T$. Finally, the asymptotic distribution is the same in these two cases.

If the $u_{n,i}$ have AQVT $U(t)$, the mixed normal variance transforms

$$2T \int_0^T H'(u)((\langle X, X\rangle_t')^2 dt = 2 \int_0^1 U'(u)((\langle Y, Y\rangle_t')^2 dt.$$

The transformation (2.70) regularizes spacing. It means that without loss of generality, one can take $T = 1$, $F' = 1$ and $U = H$. Also, the transformation (2.70) regularizes spacing defined by (2.69), and in this case, $U'(t) \equiv 1$.

Example 2.53 *On the other hand, it is clear from Example 2.51 that it is possible for $U'(t)$ to take other values than 1. The example shows that for Poisson distributed observation times, $H' = U' \equiv 2$, while, indeed $F'(t) \equiv 1/T$.*

The general situation can be expressed as follows:

Proposition 2.54 *Assume that F exists and is monotonely increasing. Also assume that H exists. Then U exists. For all $s \leq t$, $U(t) - U(s) \geq t - s$. In particular, if $U'(t)$ exists, then $U'(t) \geq 1$. The following statements are equivalent:*
(i) $U(1) = 1$
(ii) $U' \equiv 1$
(iii) $\sum_{j=0}^{n} \left(u_{n,j+1} - u_{n,j} - \frac{1}{n}\right)^2 = o_p(n^{-1})$.

Proof of Proposition 2.54. The first statement uses a standard property of the variance: if $\Delta t_{n,j+1} = t_{n,j+1} - t_{n,j}$, and $\overline{\Delta}_n = T/n$, then

$$\begin{aligned}
&\frac{n}{T} \sum_{t_{n,j+1} \leq t} (\Delta t_{n,j+1})^2 \\
&= \frac{n}{T} \sum_{t_{n,j+1} \leq t} (\Delta t_{n,j+1} - \overline{\Delta}_n)^2 + \frac{n}{T} \#\{t_{n,i+1} \leq t\}(\overline{\Delta}_n)^2 \\
&\geq \frac{n}{T} \#\{t_{n,i+1} \leq t\}(\overline{\Delta}_n)^2.
\end{aligned}$$

By taking limits as $n \to \infty$ under $F'(t) \equiv 1/T$, we get that $H(t) - H(s) \geq t - s$. In particular, the same will be true for U.

The equivalence between (i) and (iii) follows from the proof of Lemma 2 (p. 1029) in Zhang (2006). (The original lemma uses slightly different assumptions). □

The implication of the proposition is that under the scenario $U(1) = 1$, observation times are "almost" equidistant. In particular, subsampling does not change the structure of the spacings. On the other hand, when $U(1) > 1$, there is scope for subsampling to regularize the times.

Example 2.55 *Suppose that the times are Poisson distributed. Instead of picking every observation, we now pick every M'th observation. By the same methods as in Example 2.51, we obtain that*

$$U(t) = \frac{M+1}{M} t.$$

Hence the sparser the subsampling, the more regular the times will be. This is an additional feature of subsampling that remains to be exploited.

2.7.3 Proof of Theorem 2.49

We begin by describing the relationship between R_n and P_n^*. In analogy with Proposition 2 of Mykland and Zhang (2009), we obtain that

Lemma 2.56

$$\log \frac{dR_n}{dP_n^*}(U_{t_0}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}}) = \sum_i \sum_{\tau_{i-1} \le t_j < \tau_i} \left\{ \ell(\Delta X_{t_{j+1}}; \zeta_{\tau_{n,i-1}} \Delta \tau_i / M_i) - \ell(\Delta X_{t_{j+1}}; \zeta_{t_{n,j}} \Delta t_{j+1}) \right\}.$$

Now set $\zeta_t = \sigma_t \sigma_t^T$ (superscript "T" meaning transpose).

PROOF OF THEOREM 2.49. Let $Z_n^{(1)}$ and $Z_n^{(2)}$ be as in the statement of the theorem. Set

$$\begin{aligned} \Delta Z^{(1)}_{n,t_{n,j+1}} &= \frac{1}{2} \Delta X^T_{t_{n,j+1}} (\zeta^{-1}_{t_{n,j}} - \zeta^{-1}_{\tau_{n,i-1}}) \Delta X_{t_{n,j+1}} \Delta t^{-1}_{n,j+1} \\ \Delta Z^{(2)}_{n,t_{n,j+1}} &= \frac{1}{2} \Delta X^T_{t_{n,j+1}} (\zeta^{-1}_{\tau_{n,i-1}}) \Delta X_{t_{n,j+1}} \left(\Delta t^{-1}_{n,j+1} - \left(\frac{\Delta \tau_{n,i}}{M_i} \right)^{-1} \right) \end{aligned} \tag{2.71}$$

and note that $Z_n^{(v)} = \sum_j \Delta Z^{(v)}_{n,t_{n,j+1}}$ for $v = 1, 2$. Set $A_j = \zeta^{1/2}_{t_{n,j}} \zeta^{-1}_{\tau_{n,i-1}} \zeta^{1/2}_{t_{n,j}} - I$ and $B_j = \left(\Delta t_{n,j+1} \left(\frac{\Delta \tau_{n,i}}{M_i} \right)^{-1} - 1 \right)$ (the latter is a scalar). Set $C_j = \zeta^{1/2}_{t_{n,j}} \zeta^{-1}_{\tau_{n,i-1}} \zeta^{1/2}_{t_{n,j}} \left(\Delta t_{n,j+1} \left(\frac{\Delta \tau_{n,i}}{M_i} \right)^{-1} - 1 \right) = (I + A_j) B_j$.

Since $\Delta X_{t_{n,j}}$ is conditionally Gaussian, we obtain (under P_n^*)

$$\begin{aligned} E_{P_n^*}(\Delta Z^{(1)}_{n,t_{n,j+1}} | \mathcal{X}_{n,t_{n,j}}) &= -\frac{1}{2} \mathrm{tr}(A_j) \\ E_{P_n^*}(\Delta Z^{(2)}_{n,t_{n,j+1}} | \mathcal{X}_{n,t_{n,j}}) &= -\frac{1}{2} \mathrm{tr}(C_j) = -\frac{1}{2}(p + \mathrm{tr}(A_j)) B_j \end{aligned} \tag{2.72}$$

and

$$\begin{array}{c}\text{conditional covariance of} \\ \Delta Z^{(1)}_{n,t_{n,j+1}} \text{ and } \Delta Z^{(2)}_{n,t_{n,j+1}}\end{array} = \frac{1}{2}\begin{pmatrix} \operatorname{tr}(A_j^2) & \operatorname{tr}(A_jC_j) \\ \operatorname{tr}(A_jC_j) & \operatorname{tr}(C_j^2)\end{pmatrix}. \qquad (2.73)$$

Finally, let $M_n^{(v)}$ be the (end point of the) martingale part (under P^*) of $Z_n^{(v)}$ ($v = 1, 2$), so that

$$M_n^{(1)} = Z^{(1)} + (1/2)\sum_j \operatorname{tr}(A_j) \text{ and } M_n^{(2)} = Z^{(2)} + (1/2)\sum_j \operatorname{tr}(C_j). \qquad (2.74)$$

If $\langle \cdot, \cdot \rangle^{\mathcal{G}}$ represents discrete time predictable quadratic variation on the grid $\mathcal{G}$, then equation (2.73) yields

$$\begin{pmatrix} \langle M_n^{(1)}, M_n^{(1)}\rangle^{\mathcal{G}} & \langle M_n^{(1)}, M_n^{(2)}\rangle^{\mathcal{G}} \\ \langle M_n^{(1)}, M_n^{(2)}\rangle^{\mathcal{G}} & \langle M_n^{(2)}, M_n^{(2)}\rangle^{\mathcal{G}} \end{pmatrix} = \frac{1}{2}\sum_j \begin{pmatrix} \operatorname{tr}(A_j^2) & \operatorname{tr}(A_jC_j) \\ \operatorname{tr}(A_jC_j) & \operatorname{tr}(C_j^2)\end{pmatrix}. \qquad (2.75)$$

The following is shown in *ibid.*, Appendix B:

$$\langle M_n^{(1)}, M_n^{(1)}\rangle^{\mathcal{G}} = \Gamma_1 + o_p(1), \qquad (2.76)$$

where K is the ADD given by equation (2.54),

$$\sup_j \operatorname{tr}(A_j^2) \to 0 \text{ as } n \to \infty, \qquad (2.77)$$

$$\text{for } r > 2,\ |\operatorname{tr}(A_j^r)| \leq \operatorname{tr}(A_j^2)^{r/2}, \qquad (2.78)$$

and

$$\log \frac{dQ_n}{dP_n^*} = M_n^{(1)} - \frac{1}{2}\langle M_n^{(1)}, M_n^{(1)}\rangle^{\mathcal{G}} + o_p(1).$$

Now observe that by (2.76)-(2.78),

$$\begin{aligned}\sum_j \operatorname{tr}(C_j^2) &= \sum_j \operatorname{tr}(I_p)B_j^2 + \sum_j \operatorname{tr}(A_j)B_j^2 + \sum_j \operatorname{tr}(A_j^2)B_j^2 \\ &= p\sum_j B_j^2 + o_p(1) \\ &= 2\Gamma_2 + o_p(1),\end{aligned}$$

and

$$\begin{aligned}\sum_j \operatorname{tr}(A_jC_j) &= \sum_j \operatorname{tr}(A_j)B_j + \sum_j \operatorname{tr}(A_j^2)B_j \\ &= \sum_j \operatorname{tr}(A_j)B_j + o_p(1) \\ &= o_p(1)\end{aligned}$$

where the last transition follows by Condition 2.48. Meanwhile, since

$$\log \frac{dR_n}{dP_n^*} = \log \frac{dR_n}{dQ_n} + \log \frac{dQ_n}{dP_n^*},$$

we obtain similarly that

$$\begin{aligned} \log \frac{dR_n}{dQ_n} &= Z_n^{(2)} + \frac{p}{2} \sum_j \log(1 + B_j) \\ &= M_n^{(2)} + \Gamma_3 + o_p(1). \end{aligned}$$

At this point, let $\langle M_n, M_n \rangle$ be the quadratic variation of the continuous martingale that coincides at points $t_{n,j}$ with the discrete time martingale leading up to the end point $M_n^{(1)}$. By a standard quarticity argument (as in the proof of Remark 2 in Mykland and Zhang (2006)), (2.75)-(2.76)-(2.78) and the conditional normality of $(\Delta Z^{(1)}_{n,t_{n,j+1}}, \Delta Z^{(2)}_{n,t_{n,j+1}})$ yield that $\langle M_n, M_n \rangle = \langle M_n, M_n \rangle^{\mathcal{G}} + o_p(1)$. The stable convergence to a normal distribution with covariance matrix

$$\begin{pmatrix} \Gamma_1 & 0 \\ 0 & \Gamma_2 \end{pmatrix}$$

then follows by the same methods as in Zhang et al. (2005). The result is thus proved. □

Acknowledgments

We would like to thank Mathieu Kessler, Alexander Lindner, and Michael Sørensen, and the anonymous referee, for encouragement and constructive comments. We would also like to thank Eric Renault and D. Christina Wang for a careful reading of the manuscript. We are also grateful to seminar participants in Hong Kong and Bergen for terrific and useful feedback, in particular Knut Aase, Jonas Andersson, Lancelot James, Bingyi Jing, Jostein Lillestøl, Albert Lo, Trygve Nilsen, and Bård Støve. Financial support from the National Science Foundation under grants DMS 06-04758, SES 06-31605, and SES 11-24526 is gratefully acknowledged. We would like to thank Hong Kong University of Science and Technology, where part of the manuscript was written.

References

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scand. J. Statist.*, *3*, 15-27.

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.*, *6*, 701-726.

Aït-Sahalia, Y. (2004). Disentangling diffusion from jumps. *J. Finan. Econ.*, *74*, 487-528.

Aït-Sahalia, Y., & Jacod, J. (2007). Volatility estimators for discretely sampled Lévy processes. *Ann. Statist.*, *35*, 335-392.

Aït-Sahalia, Y., Mykland, P. A., & Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *J. Econometrics*, *160*, 160-165.

Aldous, D. J., & Eagleson, G. K. (1978). On mixing and stability of limit theorems. *Ann. Probab.*, *6*, 325-331.

Andersen, P. K., Borgan, Ø., Gill, R. D., & Keiding, N. (1992). *Statistical Models Based on Counting Processes*. New York: Springer.

Andersen, T., Bollerslev, T., & Meddahi, N. (2005). Correcting the errors: Volatility forecast evaluation using high frequency data and realized volatilities. *Econometrica*, *73*, 279-296.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *Int. Econ. Rev.*, *39*, 885-905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2000). Great realizations. *Risk*, *13*, 105-108.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *J. Amer. Statistical Assoc.*, *96*, 42-55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*, 579-625.

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics*, *162*, 149-169.

Barndorff-Nielsen, O. E., & Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, *63*, 167-241.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, *64*, 253-280.

Barndorff-Nielsen, O. E., & Shephard, N. (2004a). Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. *Econometrica*, *72*, 885-925.

Barndorff-Nielsen, O. E., & Shephard, N. (2004b). Power and bipower variation with stochastic volatility and jumps (with discussion). *J. Finan. Econ.*, *2*, 1-48.

Bibby, B. M., Jacobsen, M., & Sørensen, M. (2009). Estimating functions for discretely sampled diffusion-type models. In Y. Aït-Sahalia & L. P. Hansen (Eds.), *Handbook of Financial Econometrics* (pp. 203–268). North Holland, Oxford.

Bibby, B. M., Skovgaard, I. M., & Sørensen, M. S. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli*, *11*, 191-220.

Billingsley, P. (1995). *Probability and Measure* (Third ed.). New York: Wiley.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *J. Polit. Economy*, *81*, 637-654.

Cochran, W. (1934). The distribution of quadratic forms in a normal system, with applications to the analysis of variance. *Proc. Camb. Phil. Soc.*, *30*, 178-191.

Comte, F., & Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Math. Finance*, *8*, 291-323.

Dacorogna, M. M., Gençay, R., Müller, U., Olsen, R. B., & Pictet, O. V. (2001). *An Introduction to High-Frequency Finance*. San Diego: Academic Press.

Dambis, K. (1965). On the decomposition of continuous sub-martingales. *Theory Probab. Appl.*, *10*, 401-410.

Delattre, S., & Jacod, J. (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli*, *3*, 1-28.

Delbaen, F., & Schachermayer, W. (1995). The existence of absolutely continuous local martingale measures. *Ann. Appl. Probab.*, *5*, 926-945.

Dellacherie, C., & Meyer, P. (1982). *Probabilities and Potential B*. Amsterdam: North-Holland.

Ditlevsen, P., Ditlevsen, S., & Andersen, K. (2002). The fast climate fluctuations during the stadial and interstadial climate states. *Annals of Glaciol-*

ogy, *35*, 457-462.

Ditlevsen, S., & Sørensen, M. (2004). Inference for observations of integrated diffusion processes. *Scand. J. Statist.*, *31*, 417-429.

Dubins, L., & Schwartz, G. (1965). On continuous martingales. *Proc. Nat. Acad. Sci. U.S.A.*, *53*, 913-916.

Duffie, D. (1996). *Dynamic Asset Pricing Theory*. Princeton, N.J.: Princeton University Press.

Elerian, O., Siddhartha, C., & Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, *69*, 959-993.

Feller, W. (1951). Two singular diffusion problems. *Ann. Math.*, *54*, 173-182.

Fleming, J., Kirby, C., & Ostdiek, B. (2001). The economic value of volatility timing. *J. Finance*, *56*, 329-352.

Fleming, J., Kirby, C., & Ostdiek, B. (2003). The economic value of volatility timing using realized volatility. *J. Finan. Econ.*, *67*, 473-509.

Foster, D., & Nelson, D. (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica*, *64*, 139-174.

Genon-Catalot, V., & Jacod, J. (1994). Estimation of the diffusion coefficient of diffusion processes: Random sampling. *Scand. J. Statist.*, *21*, 193-221.

Genon-Catalot, V., Jeantheau, T., & Larédo, C. (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, *5*, 855-872.

Genon-Catalot, V., Jeantheau, T., & Larédo, C. (2000). Stochastic volatility models as hidden markov models and statistical applications. *Bernoulli*, *6*, 1051-1079.

Gikhman, I. I., & Skorohod, A. V. (1969). *Introduction to the Theory of Random Processes*. Philadelphia, P.A.: W. B. Sauders Company.

Gloter, A. (2000). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM Probab. Statist.*, *4*, 205-227.

Gloter, A., & Jacod, J. (2001a). Diffusions with measurement errors. II - optimal estimators. *ESAIM Probab. Statist.*, *5*, 243-260.

Gloter, A., & Jacod, J. (2001b). Diffusions with measurement errors. I - local asymptotic normality. *ESAIM Probab. Statist.*, *5*, 225-242.

Hájek, J., & Sidak, Z. (1967). *Theory of Rank Tests*. New York: Academic Press.

Hall, P., & Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Boston: Academic Press.

Harrison, M., & Kreps, D. (1979). Martingales and arbitrage in multiperiod securities markets. *J. Econ. Theory*, *20*, 381-408.

Harrison, M., & Pliska, S. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Process. Appl.*, *11*, 215-260.

Hayashi, T., & Mykland, P. (2005). Evaluating hedging errors: An asymptotic approach. *Math. Finance*, *15*, 1931-1963.

Hayashi, T., & Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, *11*, 359-379.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bonds and currency options. *Rev. Finan. Stud.*, *6*, 327-343.

Hoffmann, M. (2002). Rate of convergence for parametric estimation in a stochastic volatility model. *Stochastic Process. Appl.*, *97*, 147-170.

Jacobsen, M. (2001). Discretely observed diffusions: Classes of estimating functions and small δ-optimality. *Scand. J. Statist.*, *28*, 123-150.

Jacod, J. (1994). *Limit of random measures associated with the increments of a Brownian semimartingale* (Tech. Rep.). Université de Paris VI.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., & Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.*, *119*, 2249-2276.

Jacod, J., & Protter, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.*, *26*, 267-307.

Jacod, J., & Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes* (Second ed.). New York: Springer.

Kaplan, E., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statistical Assoc.*, *53*, 457-481.

Karatzas, I., & Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*. New York: Springer.

LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. New York: Springer.

LeCam, L., & Yang, G. (2000). *Asymptotics in Statistics: Some Basic Concepts* (Second ed.). New York: Springer.

Lee, S. Y., & Mykland, P. A. (2006). Jumps in financial markets: A new nonparametric test and jump dynamics. *Rev. Finan. Stud.*, *21*, 2535-2563.

Lehmann, E. (1983). *Theory of Point Estimation*. New York: Wiley.

Li, Y., Mykland, P., Renault, E., Zhang, L., & Zheng, X. (2009). *Realized volatility when endogeniety of time matters* (Tech. Rep.). University of Chicago.

Li, Y., & Mykland, P. A. (2007). Are volatility estimators robust with respect to modeling assumptions? *Bernoulli*, *13*, 601-622.

Linton, O. (2007). *Notes about Gaussian calculus.* (Presented in Imperical College Financial Econometrics Conference)

Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. *Giornale dell'Istituto Italiano degli Attuari*, *LXIV*, 19-47.

Mardia, K. V., Kent, J., & Bibby, J. (1979). *Multivariate Analysis*. London: Academic Press.

Meddahi, N. (2001). *An eigenfunction approach for volatility modeling* (Tech. Rep.). Université de Montréal.

Merton, R. C. (1973). The theory of rational option pricing. *Bell J. Econ. Manag. Sci.*, *4*, 141-183.

Mykland, P. A. (1995). Embedding and asymptotic expansions for martingales. *Probab. Theory Rel. Fields*, *103*, 475-492.

Mykland, P. A. (2000). Conservative delta hedging. *Ann. Appl. Probab.*, *10*, 664-683.

Mykland, P. A. (2003a). Financial options and statistical prediction intervals. *Ann. Statist.*, *31*, 1413-1438.

Mykland, P. A. (2003b). The interpolation of options. *Finance Stochastics*, *7*, 417-432.

Mykland, P. A. (2005). *Combining statistical intervals and market proces: The worst case state price distribution* (Tech. Rep.). University of Chicago.

Mykland, P. A. (2010a). A Gaussian calculus for inference from high frequency data. *Ann. Finance (to appear)*.

Mykland, P. A. (2010b). Option pricing bounds and statistical uncertainty: Using econometrics to find an exit strategy in derivatives trading. In Y. Aït-Sahalia & L. P. Hansen (Eds.), *Handbook of Financial Econometrics* (Vol. II, p. 135-196). North Holland, Oxford.

Mykland, P. A., & Zhang, L. (2006). ANOVA for diffusions and Itô processes. *Ann. Statist.*, *34*, 1931-1963.

Mykland, P. A., & Zhang, L. (2008). Inference for volatility type objects and implications for hedging. *Statistics and its Interface*, *1*, 255-278.

Mykland, P. A., & Zhang, L. (2009). Inference for continuous semimartingales observed at high frequency. *Econometrica*, *77*, 1403-1455.

Mykland, P. A., & Zhang, L. (2011). The double Gaussian approximation for high frequency data. *Scand. J. Statist.*, *38*, 215-236.

Neftci, S. N. (2000). *Introduction to the Mathematics of Financial Derivatives* (Second ed.). Amsterdam: Academic Press.

Nelson, W. (1969). Hazard plotting for incomplete failure data. *J. Qual. Tech.*, *1*, 27-52.

Øksendal, B. (2003). *Stochastic Differential Equations: An Introduction with Applications* (Sixth ed.). Berlin: Springer.

Podolskij, M., & Vetter, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, *15*, 634-658.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer.

Protter, P. (2004). *Stochastic Integration and Differential Equations* (Second ed.). New York: Springer.

Rényi, A. (1963). On stable sequences of events. *Sankyā Series A*, *25*, 293-302.

Rice, J. (2006). *Mathematical Statistics and Data Analysis* (Third ed.). Duxbury Press.

Rootzén, H. (1980). Limit distributions for the error in approximations of stochastic integrals. *Ann. Probab.*, *8*, 241-251.

Ross, S. (1996). *Stochastic Processes* (Second ed.). New York, N.Y.: Wiley.

Ross, S. M. (1976). The arbitrage theory of capital asset pricing. *J. Econ. Theory*, *13*, 341-360.

Shreve, S. E. (2004). *Stochastic Calculus for Finance II*. New York: Springer-Verlag.

Sørensen, H. (2001). Discretely observed diffusions: Approximation of the continuous-time score function. *Scand. J. Statist.*, *28*, 113-121.

Valdés-Sosa, P. A., Bornot-Sánchez, J. M., Melie-García, L., Lage-Castellanos, A., & Canales-Rodriguez, E. (2007). Granger causality on spatial manifolds: Applications to neuroimaging. In B. Schelter, M. Winterhalder, & J. Timmer (Eds.), *Handbook of Time Series Analysis: Recent Theoretical Developments* (p. 461-485). New York: Wiley.

Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *J. Econometrics*, *159*, 235-250.

Zhang, L. (2001). *From martingales to ANOVA: Implied and realized volatility*. Unpublished doctoral dissertation, The University of Chicago, Department of Statistics.

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

Zhang, L. (2009). Implied and realized volatility: empirical model selection. *Ann. Finance (to appear)*.

Zhang, L. (2011). Estimating covariation: Epps effect and microstructure noise. *J. Econometrics*, *160*, 33-47.

Zhang, L., Mykland, P. A., & Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statistical Assoc.*, *100*, 1394-1411.