The Estimation of Continuous and Discontinuous Leverage Effects

Yacine Aït-Sahalia\textsuperscript{1}, Jianqing Fan\textsuperscript{1}, Roger J.A. Laeven\textsuperscript{2}, D. Christina Wang\textsuperscript{1}, and Xiye Yang\textsuperscript{2}

\textsuperscript{1}Princeton University
\textsuperscript{2}University of Amsterdam

This draft, June 2013

Abstract

This paper examines leverage effect under a general setup that allows the log-price and volatility processes to be Itô semi-martingales and decomposes it into the continuous and the discontinuous part. The challenges in estimation consist in the various biases due to the latency of volatility and mutual interference between the Brownian and jump parts of returns. We remove these potential biases by employing blocking and truncation techniques. Applying the estimation strategy to high-frequency data, we find positive evidence for the presence of the two leverage effects in financial market, especially the discontinuous one.

Keywords: Leverage effect, Itô semimartingale, quadratic covariation, high-frequency data

1 Introduction

The leverage effect, an enduring yet controversial phenomenon in financial market, refers to the generally negative dependence between asset prices and volatilities. In the discrete time (also low frequency) framework, the main concern is its economic source(s). Black (1976) proposed two possible explanations for this phenomenon: “direct causation effect,” referring to the causal relation from returns to volatility
changes ¹; and “reverse causation effect,” namely, the other way around ². Christie (1982) attributed this phenomenon to changes in financial leverages, or debt-to-equity ratios. However, the validity of this explanation has been questioned by some other empirical studies. For instance, Figlewski and Wang (2000) noted that there is no apparent effect on volatility when financial leverage changes, but only when stock prices change. Hasanhodzic and Lo (2011) found just as strong an inverse relationship between returns and the subsequent volatility changes for all-equity-financed companies as for their debt-financed counterparts. The other explanation, which rests on a time varying risk premium, hence named volatility feedback effect later on, was further discussed by, for example, Pindyck (1984), French, Schwert, and Stambaugh (1987), and Campbell and Hentschel (1992). In an empirical study, Bekaert and Wu (2000) argued that the volatility feedback effect dominates the financial leverage effect.

Although theoretically, such a difference in causality is very fundamental and has distinct economic consequences, the above two causal relationships may both exist hence, from an empirical perspective, are hardly distinguishable using low-frequency data. Moreover, such economic theory oriented studies do not attach sufficient importance to rigorously estimating the leverage effect (also limited by available data and statistical tools), a step which is essential to any conclusive claim. Hence, in this paper, we focus on the statistical properties of the leverage effect and provide valid estimator(s) to identify it in practice, and leave its economic implications for future study.

In the continuous time framework, with more powerful analytical tools available, we can allow more general dynamics of the log-price and volatility processes, by including different stochastic components like Brownian motion and jumps. Consequently, the leverage effect can take more sophisticated form(s) and exhibit more dynamic properties. Hence, the first question we are going to answer in this paper is that what kind(s) of leverage effect (in a statistical sense) can we have in a continuous time model. We categorize the (statistical) leverage effects according to the correlated stochastic components and define: (1) continuous leverage effect (CLE) as the quadratic covariation between the continuous parts of the log-price and volatility processes; (2) discontinuous leverage effect (DLE) as the quadratic variation of their

¹A drop in the value of the firm’s equity will cause a negative return on its stock and will increase the leverage of the firms (i.e., its debt/equity ratio), a rise which further leads to a rise in the volatility of the stock.
²When pricing volatility, an anticipated increase would raise the required rate of return, in turn necessitating an immediate stock-price decline to allow for higher future returns.
discontinuous parts; and (3) generalized leverage effect (GLE) as the sum of the above two.

In most of the literature, as a direct extension to the discrete time leverage effect, CLE is controlled by a constant, namely the leverage parameter. In Bandi and Renò (2012), the authors allowed leverage to be a function of the state of the firm (hence time-varying), which is summarized by the spot variance (or spot volatility). In our paper, we impose no such structural assumption and allow CLE to be a very general stochastic process, which may have additional source(s) of randomness other than those in log-prices and volatilities. Wang and Mykland (2011) defined the leverage effect in the same way as CLE and provided non-parametric estimators, with a particular interest in eliminating the impacts from market micro-structure noises. They assumed both return and volatility processes to be continuous, whereas we allow both of them to have discontinuous parts, hence introduce the second leverage effect. DLE comes from co-jumping of log-prices and volatilities and has not been investigated rigorously so far. Eraker, Johannes, and Polson (2003) proposed a parametric model which permits the presence of co-jumps. However, the jump process was assumed to have finite activity and the parametric form is also quite restrictive. Additionally, the authors did not explicitly define a concept like DLE. Bandi and Renò (2012) did discuss the concept “co-jump leverage”, which is the conditional expectation of our DLE. But their model is nested in ours and they did not provide any estimator of the “co-jump leverage” while we do. Since Brownian motion and jumps have distinct stochastic properties, CLE and DLE would have completely different economic implications, for example, having different impacts on pricing and risk management. Compared with small continuous changes, jumps could be much bigger and most importantly, co-jumps can induce very large risks by contagion. Furthermore, Jacod and Todorov (2010) found that in approximately forty percent of the sample weeks from January 1997 to June 2007, there is strong evidence for common price and volatility jumps. Therefore, it would be very important and helpful to study DLE, which provides a meaningful measurement of co-jumping. Because of their distinctive features, it is more interesting to study CLE and DLE separately rather than together. In other words, GLE may have limited economic implications, hence is less interesting in practice. As to be shown later, there is no Central Limit Theorem result associated to GLE in general.

Given more sources of disturbances (noise, jumps etc.) and various kinds of leverage effects, it may be harder to identify each than in discrete time case. Therefore,
the second and main question is about the existence and identification of the above statistical leverage effects. Relying on high frequency (five-minute) absolute returns as a simple volatility proxy, Bollerslev, Litvinova, and Tauchen (2006) revealed a new and striking prolonged negative correlation between the volatility and the current and lagged returns, which lasts for several days. On the other hand, the authors also observed a very strong contemporaneous correlation between the high-frequency returns and their absolute value. These findings support the dual presence of prolonged leverage effects and an almost instantaneous volatility feedback effect.

However, most recently, a puzzle in Aït-Sahalia, Fan, and Li (2012) casted doubt on the presence of (continuous) leverage effects in high frequency data. Using the well-known Heston model, they found that, at high frequency and over short horizons, the estimated “leverage parameter” $\rho$ is close to zero instead of an expected strong negative value. At longer horizons, the effect is present, especially when using option-implied volatility. The authors further provided theoretical results to disentangle biases involved in estimating the “leverage parameter,” and isolated the sources one by one: discretization bias, smoothing bias, estimation error and noise correction error. Additionally, they also considered the influence of price jumps on estimating the correlation between Brownian shocks to price and volatility. The most interesting and striking part was, as shown in simulation study, they successfully recovered the “leverage parameter” using a simple OLS regression.

Bandi and Renò (2012) found significant volatility-varying leverage effect, but there was limiting bias in their estimators. We are going to show that one main reason their estimators are biased is that they include variance of estimated spot volatility increments in their estimators. In contrast, our estimators for both CLE and DLE are asymptotically unbiased since our definitions are based on quadratic covariation only, hence the estimators do not involve squared estimated volatility changes. In our simulation study, we compare the performance of leverage parameter estimators constructed from either variance and covariance (the usual definition), or quadratic variation and covariation. We found that the bias in the latter one is

---

3This is not a rigorous estimator. But the authors also estimated both one- and two- factor stochastic volatility models, using the Efficient Method of Moments (EMM), to study which one is more capable to reproduce the prolong leverage effect.

4Both of them defined the leverage effect by using a correlation function, where variances are in the denominator. Intuitively, the variance of estimated spot volatility increments would be larger than the true one due to estimation error. Therefore, with potentially over-estimated denominators, their leverage effect estimators are biased.

5It should be biased since that same as variance, the quadratic variation of estimated spot volatility
much smaller than the former one. This may imply that although variance, covariance and correlation are very basic concepts, especially in stationary discrete time case, it would be better to use quadratic covariation and variation in continuous time.

Additionally, this paper also has contributions in terms of statistical or econometric theory. In the past decade, in estimating integrated volatility, a lot of efforts have been made to find the statistical limits of even (or globally even \(^6\)) function(s) of the increments of stochastic (in most cases, Itô semimartingale) process(es) under various circumstances. One can refer to Jacod (2008, 2009) and Jacod and Protter (2011) among others to see various asymptotic limits of such functionals of stochastic processes. In this paper, the proposed estimators involve the asymptotic properties of (globally) odd functions \(^7\) of Itô semi-martingale’s increments, as a complement to the existing results. The study of odd functionals may not be interesting at first glance, since, for example, the expectation of odd powers of Brownian increments are zero, implying that its odd functionals may not converge to a non-trivial limit. However, as shown in the following, after appropriate arrangement, the sum of certain odd functionals of the observed return process do converge to something meaningful, namely, the continuous (discontinuous or generalized) leverage effect. Moreover, it is worth pointing out that in our estimates the number of successive increments in each summand of the functionals goes to infinity, as time step goes to 0. To the best of our knowledge, in most of the existing papers except for “pre-averaging” related ones [e.g., Jacod, Li, Mykland, Podolskij, and Vetter (2009), Jacod, Podolskij, and Vetter (2010) and Chapter 12, 16 in Jacod and Protter (2011)], the studied functions are of finite number of increments. For instance, there are two adjacent increments in each summand in the bi-power variation case.

The rest of the paper proceeds as follows. Section 2 details model setup, assumptions and the definitions of three leverage effects. We introduce the blocking scheme and estimators in Section 3. We analyze biases in our estimator of CLE, investigate the puzzle in Aït-Sahalia, Fan, and Li (2012) and present the Law of Large Number.

\(^6\)For instance, \(f(x_1, x_2) = x_1^r x_2^s\) is globally even if and only if \(r + s\) is even, but not even in either \(x_1\) nor \(x_2\) if both \(r\) and \(s\) are odd integers.

\(^7\)To be specific, we use functions like \(f(x_1, x_2) = x_1 x_2^2\) and its truncated versions. Observe that this function is odd with respect to \(x_1\), but even in the second argument \(x_2\). However, if we treat \(x = (x_1, x_2)\) as a whole, we have \(f(-x) = -f(x)\). The use of odd functions here is due to the fact that the volatility process is latent hence needs to be estimated from returns. Otherwise, if both the return and volatility series are observable, then this case does not have an essential difference from estimating quadratic co-variation between two (or more) asset prices.
(LLN) results. Section 5 gives the corresponding Central Limit Theorems (CLTs) and section 6 discusses testing issues. We demonstrate the performance of our estimators by Monte-Carlo simulations in Section 7, where we also compare two different estimators of the leverage parameter and show the source of bias which accounts for the puzzle. Section 8 applies our nonparametric estimation method to real financial data. We further analyze the financial implications of these two effects in pricing and hedging in section 9. All proofs are given in Section 10.

2 Model Setup and Definitions

2.1 Setup

We start with a filtered space \((\Omega, \mathcal{F}, \mathcal{F}_t, t \geq 0, \mathcal{P})\), on which various processes are defined. More detailed assumption about the return and its volatility processes is given as follows:

Assumption (H): Assume the both the return process and volatility process are Itô semimartingales, which could be represented in terms of Wiener process(es) and Poisson random measure (Grigelsonis decomposition):

\[
X_t = X_0 + \int_0^t a_s ds + \int_0^t \sigma_s dW_s + (\delta 1_{|\delta| \leq \kappa}) \star (\mu - \nu)_t + (\delta 1_{|\delta| > \kappa}) \star \mu_t, \tag{2.1}
\]

\[
\sigma_t = \sigma_0 + \int_0^t \tilde{a}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \tilde{b}_s dB_s + (\tilde{\delta} 1_{|\tilde{\delta}| \leq \kappa}) \star (\tilde{\mu} - \tilde{\nu})_t + (\tilde{\delta} 1_{|\tilde{\delta}| > \kappa}) \star \tilde{\mu}_t. \tag{2.2}
\]

Here, \(W\) and \(B\) are independent standard Brownian motions; \(\mu\) is a Poisson random measure (independent with the Brownian measure of \(W\)) on \((0, \infty) \times E\) with intensity measure \(\nu(dt, dx) = dt \otimes \lambda(dx)\), where \(\lambda\) is a \(\sigma\)-finite measure without atom on an auxiliary measurable set \((E, \mathcal{E})\); \(\delta(\omega, t, x)\) is a predictable function on \(\Omega \times \mathbb{R}_+ \times E\); so are \(\tilde{\mu}, \tilde{\nu}, \tilde{\delta}\); and \(\kappa\) is a positive constant; finally, the symbol \(\star\) denotes the (possibly stochastic) integral w.r.t a random measure. More specifically we have

\[
(\delta 1_{|\delta| \leq \kappa}) \star (\mu - \nu)_t = \int_0^t \int_{\mathbb{R}} (\delta(\omega, s, x) 1_{|\delta(\omega, s, x)| \leq \kappa})(\mu - \nu)(ds, dx),
\]

\[
(\delta 1_{|\delta| > \kappa}) \star \mu_t = \int_0^t \int_{\mathbb{R}} (\delta(\omega, s, x) 1_{|\delta(\omega, s, x)| > \kappa}) \mu(ds, dx).
\]

Additionally, we assume:
(a) The processes $a_t(\omega), \tilde{a}_t(\omega), \tilde{\sigma}_t(\omega), \tilde{b}_t(\omega)$, $\sup_{x \in E} \|\delta(\omega,t,x)\|_{\gamma(x)}$ and $\sup_{x \in E} \|\tilde{\delta}(\omega,t,x)\|_{\tilde{\gamma}(x)}$ are locally bounded, where $\gamma(x)$ and $\tilde{\gamma}(x)$ are nonnegative functions satisfying $\int_E (\gamma(x)^2 \land 1) \lambda(dx) < \infty$ and $\int_E (\tilde{\gamma}(x)^2 \land 1) \tilde{\lambda}(dx) < \infty$ respectively.

(b) All paths $t \mapsto a_t(\omega), t \mapsto \tilde{a}_t(\omega), t \mapsto \tilde{\sigma}_t(\omega), t \mapsto \tilde{b}_t(\omega)$ are càdlàg (left-continuous with right limits). $\tilde{\sigma}_+ \ (\text{right limit of } \tilde{\sigma})$ is also an Itô semimartingale with a Grigelionis decomposition similar to (2.2);

(c) The processes $\sigma^2$ and $\sigma^2_-$ (left limit of $\sigma^2$) are everywhere invertible. 8

Remark 2.1. We could rewrite the volatility jump measure $\tilde{\mu}$ as a linear combination of another two Poisson random measures $\mu_1$ and $\mu_2$, where $\mu_1$ has the same jumping times as $\mu$ (but the size part could be different), and $\mu_2$ is independent of $\mu$. In other words, $\mu_1$ represents the co-jump part and $\mu_2$ is the disjoint jump part.

2.2 Definitions of various leverage effects

A natural measure of the co-movement of two stochastic processes is their quadratic covariation. Denote $\Delta X_S = X_S - X_{S-}$ as the jump size of any stochastic process $X$ at time $S$. The quadratic covariation of $X$ and $\sigma^2$ can be decomposed into two parts:

$$[X, \sigma^2]_T = \langle X, \sigma^2 \rangle_T + \sum_{S \leq T} \Delta X_S \Delta \sigma^2_S. \quad (2.3)$$

We shall call the term on left hand side as the generalized leverage effect, and the first and second term on the right hand side as the continuous and discontinuous leverage effect respectively.

The continuous leverage effect could be rewritten as follows:

$$\langle X, \sigma^2 \rangle_T := \int_0^T 2 \sigma^2_{t-} \tilde{\sigma}_t dt. \quad (2.4)$$

In some applications, one may be interested in some functional of $\sigma^2$, say $F(\sigma^2)$, instead of itself. Then the continuous leverage effect becomes:

$$\langle X, F(\sigma^2) \rangle_T = \int_0^T \nabla F(\sigma^2_{t-}) 2 \sigma^2_{t-} \tilde{\sigma}_t dt. \quad (2.5)$$

8As pointed out in Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006), under this condition, the process $\sigma^2$ has a similar form of (2.2) with the same assumptions on the coefficients.
An example is the geometric OU model, with $F(x^2) = \log x^2$, $\tilde{\mu} \equiv 0$ and

$$d \log \sigma_t^2 = -\kappa \log \sigma_t^2 \, dt + \rho dW_t + \sqrt{1 - \rho^2} \, dV_t.$$  

But for notation simplicity, we mainly focus on the $\sigma^2$ case. The results in subsequent sections can be extended to the $F(\sigma^2)$ case if the function $F \in \mathbb{C}^2$.

On the other hand, we denote discontinuous leverage effect by

$$\Delta(X, \sigma^2)_T := \sum_{S \leq T} \Delta X_S \Delta \sigma^2_S.$$  

Additionally, we also consider a truncated version, where only co-jumps induced by large price jumps are selected:

$$\Delta(X, \sigma^2)_T(\epsilon) := \sum_{S \leq T} \Delta X_S \Delta \sigma^2_S \mathbf{1}_{|\Delta X_S| > \epsilon}.$$  

We could name this the tail discontinuous leverage effect.

**Remark 2.2.** The randomness of jumps comes from two sources: jump time and size. The first one could be characterized by jump intensity, which may or may not depend on the state of the underlying stochastic process(es), and the second one is described by jump size distribution.

In the definition of co-jumping leverage in Bandi and Renò (2012), jump intensity is further related to the level of volatility (or variance) and the random jump sizes (in return and volatility series) is replaced with their correlation. As a result, the first source of randomness is restricted, although such an assumption embeds an interesting driving force in the dynamics of the jump intensity process, while the second one is eliminated.

While in our definition, we preserve both without integrating any of them out, because of the using of quadratic covariation. Consequently, our setting allows more flexibility. For example, there could be alternative source(s) of jump intensity rather than contemporaneous volatility alone. And the jump size distribution could depend on time hence may have time-varying mean, variance, etc.

**Remark 2.3.** One point should be mentioned here is that, similar to the case of integrated volatility, the estimation here is not a statistical problem in the usual sense since the quantity to estimate is a random variable, instead of a parameter. Take the discontinuous leverage as an example. This means “estimating” the random variables $\langle X, \sigma^2 \rangle_T(\omega)$ and $\sum_{S \leq T} \Delta X_S(\omega) \Delta \sigma^2_S(\omega)$, for the observed $\omega$, although of
course ω is indeed not “fully” observed. Therefore, the quality of the estimator we are going to present is something which fundamentally depends on ω as well.

From this point of view, for example, in the co-jumping leverage proposed by Bandi and Renò (2012), the conditional correlation maps the two random variables, namely the sizes of return and volatility jumps, into one parameter, their correlation, hence turning the problem into a usual one.

3 Construction of the Estimators

Suppose the data is observed every $\Delta_n = \frac{T_n}{n}$ units of time without any measurement error. The full grid containing all of the observation points is given by:

$$U = \{0 = t_0^n < t_1^n < t_2^n < \cdots < t_n^n = \lfloor T/\Delta_n \rfloor \Delta_n\},$$

where $t_j^n = j\Delta_n$ for each $j$. Then the increment of log-prices over $j$-th interval could be denoted as $\Delta_j^n X := X_{t_j^n} - X_{t_{j-1}^n}$.

3.1 The continuous leverage effect (CLE)

Rather than directly describing the way the estimator to CLE to be constructed, we are going to show in several steps what the difficulties are and how they can be overcome. For simplicity, we assume the log-price process has no jumps for a while and give priority to the problem of latent volatility.

To begin with, note that both of the two causal explanations of continuous leverage effect imply the dependence of current return and future volatility, which could be approximated by normalized squared (future) return. This prompts us to study the following expression

$$\eta_{j,k}^n(X) = (\Delta_j^n X)(\Delta_k^n X)^2, \quad (3.1)$$

for any $j, k$ such that $t_j^n, t_k^n \in U$. Its conditional expectation of which can be summarized as

$$\mathbb{E}[\eta_{j,k}^n(X) \mid \mathcal{F}_{j \land k-1}] = \begin{cases} a'_j (\sigma_{t_{j-1}^n}^2) \Delta_n^2 + o_p(\Delta_n^2) & \text{if } j > k; \\
(\sigma_{t_{j-1}^n}^2)(\sigma_{t_{j-1}^n}^2) + 2\sigma_{t_{j-1}^n}^2 \Delta_n^2 + o_p(\Delta_n^2) & \text{if } j < k.
\end{cases}$$

As we expected, when $j < k$ (corresponds to current return and future volatility), $\eta_{j,k}^n(X)$ contains the information about instantaneous continuous leverage effect, i.e.
2(\sigma_{j-1}^2 - \hat{\sigma}_j^2 - \sigma_{j-1}^2). But there is another bias term, which equals its conditional expectation in the case \( j > k \). This observation suggests that we could consider the following candidate
\[
\sum_{j=1}^{[T/3\Delta_n] - 1} \left( \eta_{3j+1,3j+2}^n(X) - \eta_{3j+1,3j}^n(X) \right) / \Delta_n.
\]
This candidate may cancel the bias in the conditional expectation. And a careful calculation could verify this is indeed the case. However, \( (\Delta_{j+1}^n X)^2 / \Delta_n \) is not a consistent estimator of spot volatility. So instead, we will replace \( (\Delta_{j+1}^n X)^2 / \Delta_n \) by the consistent estimator:
\[
\hat{\sigma}_t^2 = \frac{1}{k_n \Delta_n} \sum_{j=i+1}^{i+j+k_n} (\Delta_{j}^n X \cdot 1_{\{\|\Delta_{j}^n X\| \leq \alpha_n\}})^2,
\]
where \( \alpha_n = \alpha \Delta_n^\varpi \), for some \( \alpha > 0, \varpi \in (0, \frac{1}{2}) \), and \( k_n \) is an integer and satisfies the condition below with some positive constant \( K \),
\[
\frac{1}{K} \leq k_n \Delta_n^b \leq K \quad \text{with } 0 < b < 1.
\]
Let \( I^-_n(j) = \{j - k_n - 1, j - k_n, \ldots, j - 1\} \) if \( j > k_n \) and \( I^+_n(j) = \{j + 1, j + 2, \ldots, j + k_n + 1\} \) define two local windows in time of length \( k_n \Delta_n \) just before and after time \( j \Delta_n \). Denote the truncated increment of \( X \) by \( \Delta_{j}^n X_{\alpha_n} = \Delta_{j}^n X \cdot 1_{\{\|\Delta_{j}^n X\| \leq \alpha_n\}} \). Then we can define
\[
< X, \sigma^2 >_T^C = \sum_{i=k_n+1}^{[T/\Delta_n] - k_n} \Delta_{i,j}^n X_{\alpha_n} (\hat{\sigma}_i^2 - \hat{\sigma}_{i-1}^2) := \sum_{i=k_n+1}^{[T/\Delta_n] - k_n} \beta_i^n (X)
\]
\[
\hat{\sigma}_i^2 = \frac{1}{k_n \Delta_n} \sum_{j \in I^-_n(i)} \Delta_{j}^n X_{\alpha_n}, \quad \hat{\sigma}_{i-1}^2 = \frac{1}{k_n \Delta_n} \sum_{j \in I^-_n(i)} \Delta_{j}^n X_{\alpha_n}.
\]
It is interesting to compare the estimation of CLE to that of integrated volatility. The bi-power variation is the sum of even function of the increment of log-price process \( X \). Here in our case, \( \eta_{j,k}^n(X) \) takes the functional form \( f(x_1, x_2) = x_1 x_2 \), which is odd instead of even. Additionally, \( \beta_i^n (X) \) is an odd function of increasing number of increments, the asymptotic properties of which has not been studied in the literature.
3.2 The discontinuous leverage effect (DLE)

When estimating the discontinuous leverage effect, we need to preserve jumps, instead of eliminating them as in the CLE case. Therefore, define
\[ \Delta_n^j X^\alpha_n = \Delta_j^n X \cdot 1_{\{\|\Delta_j^n X\| > \alpha_n\}} \]
and the way to estimate DLE is similar to that of CLE. We will use a centered blocking scheme as in Jacod and Todorov (2010):

\[
\hat{(X, \sigma^2)}_T^D = \sum_{i=k_n+1}^{[T/\Delta_n] - k_n} \Delta_i^n X^\alpha_n \cdot \Delta \tilde{\sigma}(k_n)_i,
\] (3.4)

where

\[
\tilde{\sigma}(k_n)_i = \frac{1}{k_n \Delta_n} \sum_{j=1}^{k_n} |\Delta_{i+j}^n X^\alpha_n|^2, \quad \text{and} \quad \Delta \tilde{\sigma}(k_n)_i = \tilde{\sigma}(k_n)_i - \tilde{\sigma}(k_n)_{i-k_n-1}. \quad (3.5)
\]

Remark 3.1. In Jacod and Todorov (2010), the authors studied a very general limit functional \( U(F)_T \) (with minor changes in notations in what follows)

\[
U(F)_T = \sum_{S \leq T} F(\Delta X_S, \sigma^2_S, \sigma^2_S) 1_{\{\Delta X_S \neq 0\}},
\] (3.6)

together with its estimator

\[
U(F, k_n)_T = \sum_{i=k_n+1}^{[T/\Delta_n] - k_n} F(\Delta_i^n X, \tilde{\sigma}(k_n)_{i-k_n-1}, \tilde{\sigma}(k_n)_i) 1_{\{\|\Delta_i^n X\| > \alpha_n\}}.
\] (3.7)

Here \( F \) is a function on \( \mathbb{R} \times \mathbb{R}^*_+ \times \mathbb{R}^*_+ \) where \( \mathbb{R}^*_+ = (0, \infty) \). Observe that with \( F(x, y, z) = x(z - y) \), the above equations are equivalent to (2.6) and (3.4) (the discontinuous leverage and its estimator), respectively.

We are going to show that both \( \hat{(X, \sigma^2)}_T^D \) converges to the discontinuous leverage effect in probability. However, except for finite jump activity case and another one where volatility process has no Brownian components, we could not have CLT result in general. Instead, we consider the tail discontinuous leverage effect and accordingly define

\[
\hat{(X, \sigma^2)}_T^{D, \epsilon} = \sum_{i=w_n+1}^{[T/\Delta_n] - w_n} \Delta_i^n X^{\alpha_n \wedge \epsilon} \cdot \Delta \tilde{\sigma}(w_n)_i,
\]
3.3 The generalized leverage effect (GLE)

It is a well-known property of convergence in probability that if \( X_n \xrightarrow{p} X \) and \( Y_n \xrightarrow{p} Y \), then \( X_n + Y_n \xrightarrow{p} X + Y \). Therefore, for example, \( \Delta(X, \sigma^2)_{T} + \Delta(X, \sigma^2)_{T} \) is a consistent estimator of \([X, \sigma^2]_{T}\).

4 Convergence in probability

4.1 Errors in estimation of continuous leverage effect

For \( 0 \leq r < 1 \), we can rewrite \( X \) as \( X_t = X'_t + X''_t \), where

\[
X''_t = \delta \ast \mu_t = \int_{0}^{t} \int_{\mathbb{R}} \delta(x) \mu(ds, dx),
\]

\[
X'_t = X_t + \int_{0}^{t} a'_s ds + \int_{0}^{t} \sigma_s dW_s \quad \text{with} \quad a'_s = a_s - \int_{|\delta(t, x)| \leq \kappa} \delta(t, x) \lambda(dx).
\]

Then we can decompose the overall estimation error in estimating continuous leverage effect as follows:

\[
\langle X, \sigma^2 \rangle^C_T - \langle X', \sigma^2 \rangle^C_T = \sum_{i=k_n+1}^{[T/\Delta_n]-k_n} (X'_{i+1} - X'_{i}) (\sigma^2_{i+1} - \sigma^2_{i}) - \int_{0}^{T} 2 \sigma^2_t \bar{\sigma}_t dt
\]

\[
\begin{align*}
\text{Discretization Error} & \quad + \quad \sum_{i=k_n+1}^{[T/\Delta_n]-k_n} \left( \Delta^n_i X' (\tilde{\sigma}^2_{i+} - \tilde{\sigma}^2_{i-}) - \Delta^n_i X' \Delta^n_i \sigma^2 \right) \\
\text{Volatility Estimation Error} & \quad + \quad \sum_{i=k_n+1}^{[T/\Delta_n]-k_n} \left( \Delta^n_i X (\sigma^2_{i+} - \sigma^2_{i}) - \Delta^n_i X' (\tilde{\sigma}^2_{i+} - \tilde{\sigma}^2_{i-}) \right) \\
\text{Truncation Error} & \quad + \quad \sum_{i=k_n+1}^{[T/\Delta_n]-k_n} \left( \Delta^n_i X (\sigma^2_{i+} - \sigma^2_{i}) - \Delta^n_i X' (\tilde{\sigma}^2_{i+} - \tilde{\sigma}^2_{i-}) \right)
\end{align*}
\]

Here, we merge the “smoothing bias” and “estimation error” in Aït-Sahalia, Fan, and Li (2012) into one term “volatility estimation error”. Instead of “noise correction error”, we study truncation bias, which is due to the presence of price jumps. As discussed in the introduction section, it can be further decomposed into two parts, namely, truncation bias in volatility estimation and truncation bias in returns. If return and volatility co-jump, then by effectively truncating out the return jumps out at such time points, we can eliminate the effect of co-jumps on the estimation of continuous leverage effect. This shares the same idea behind the test function.
proposed in Jacod and Todorov (2010). The only difference is that they truncated the continuous part of returns out to construct a co-jump testing statistics.

4.2 LLN: continuous leverage effect

Theorem 4.1. Under (H), condition (3.2) and either one of the following assumptions:

(a) $X$ is continuous and $\sigma^2$ has finite variation;

(b) $X$ is discontinuous with $\int (\gamma(x)^r \wedge 1)\lambda(dx) < \infty$ for some $r \in [0, 1)$ and $\alpha_n = \alpha \Delta_n^\omega$, for some $\alpha > 0$ and $\omega \in \left[\frac{1}{2(2-r)}, \frac{1}{2}\right]$. And disjoint jump part of $\sigma^2$ has finite variation.

then we have

$$\langle \hat{X}, \sigma^2 \rangle^C_T \xrightarrow{\text{u.c.p.}} \langle X, \sigma^2 \rangle^C_T = \int_0^T 2\sigma^2_t \tilde{\sigma}_t dt, \quad (4.2)$$

where $Z^n_T \xrightarrow{\text{u.c.p.}} Z_T$ means the sequence of random processes $Z^n_T$ “convergence in probability, locally uniformly in time” to a limit $Z_T$: that is, $\sup_{S \leq T} |Z^n_S - Z_S|_{\mathbb{F}} \xrightarrow{\text{p}} 0$ for all $T$ finite.

Remark 4.1. To compare with the estimation of integrated volatility, observe that in the Theorem 6.3 of Jacod (2009), the assumption on the jump activity index, that is, $r \in [0, 2)$, is less restrictive. The reason for this is price jumps affect the measurement of continuous leverage not only through volatility estimation, but also through its cross product with spot volatility, a way which will generate a bias term even without co-jumping\(^9\). In short, price jumps cause more problems in continuous leverage estimation than in integrated volatility estimation. Therefore, we could only allow for less active price jumps. Otherwise, their effect can not be effectively eliminated.

4.3 LLN: discontinuous leverage effect

Theorem 4.2. Assume (H) and $\mu \equiv \tilde{\mu}$, which is not vanishing and condition (3.2) as well.

\(^9\)Note that the conditional expectation of $(\Delta_n^\gamma X'')\sigma^2_n \Delta_n$ is of order $\Delta_n^2$, the same as $(\Delta_n^\gamma X')\sigma^2_t \Delta_n$. 

13
(i) If \( r \in [0, 1] \), then \( \langle X, \sigma^2 \rangle_T \) converges in probability, for the Shorokhod topology, to \( \Delta(X, \sigma^2)_T \).

(ii) Given some \( \epsilon > 0 \), for \( r \in [0, 2) \), we have \( \langle X, \sigma^2 \rangle_{T, \epsilon} \), converges in probability, for the Shorokhod topology, to \( \Delta(X, \sigma^2)_T(\epsilon) \).

Remark 4.2. These are direct results from Theorem 3.1 in Jacod and Todorov (2010). In case (i), we require \( r \in [0, 1] \) to satisfy the condition (c) of their theorem with our particular choice of \( F \). Note that case (ii) satisfies the condition (a) of that theorem.

What interesting here is the tradeoff between jump activity and jump size in the assumptions. That is, if all co-jumps are considered, then we must put a stronger restriction on the jump activity, as in case (i). While if we are only interested in co-jumps where price jumps are large (bigger than \( \epsilon \)), then the restriction on jump activity could be relaxed. Intuitively speaking, we need to put certain restrictions on summability of the process to be estimated. Since the probability of having large price jumps is relatively low, by considering co-jumps only induced by large price jumps, we implicitly set such restriction.

5 The central limit theorems

5.1 The main results

Theorem 5.1. Under \((H)\), for some finite and positive constant \( c \) let \( k_n = cn^b \) with \( 0 < b < 1 \), or equivalently \( k_n = cT^b \Delta_n^{-b} \), and either one of the following assumptions:

(a) \( X \) is continuous, and \( \sigma^2 \) has finite variation;

(b) \( X \) is discontinuous with \( \int (\gamma(x)^r \wedge 1)\lambda(dx) < \infty \) for some \( r \in [0, 1/2] \) and \( \alpha_n = \alpha \Delta_n^\omega \), for some \( \alpha > 0 \) and \( \omega \in \left[ \frac{3}{4(2-r)}, \frac{1}{2} \right] \), and disjoint jump part of \( \sigma^2 \) has finite variation.

then \( \sqrt{n^{b\wedge(1-b)}} \left( \langle X, \sigma^2 \rangle_T^C - \langle X, \sigma^2 \rangle_T^C \right) \) converges stably in law to a limiting variable defined on an extension of the original space. More formally,

\[
\sqrt{n^{b\wedge(1-b)}} \left( \langle X, \sigma^2 \rangle_T^C - \langle X, \sigma^2 \rangle_T^C \right) \overset{\mathcal{L}}{\rightarrow} \int_0^T \rho_s dB_s, \tag{5.1}
\]
where $B$ is a standard Wiener process independent of $\mathcal{F}$ and $\rho_s$ is the square root of:

$$
\int_0^T \rho_s^2 ds = \begin{cases} 
4 \int_0^T \sigma_t^6 \, dt, & \text{if } b < 1/2; \\
\frac{4}{c} \int_0^T \sigma_t^6 \, dt + \frac{2cT}{3} \int_0^T \sigma_t^2 \, d\langle \sigma^2, \sigma^2 \rangle_t, & \text{if } b = 1/2; \\
\frac{2T}{3} \int_0^T \sigma_t^2 \, d\langle \sigma^2, \sigma^2 \rangle_t, & \text{if } 1/2 < b < 1.
\end{cases}
$$

(5.2)

It is easy to see that the optimal convergence rate we can achieve here is $n^{1/4}$ when $\rho = 1/2$.

The feasible version is

$$
\frac{\sqrt{n}^{1-b}}{\sqrt{\hat{V}_T^n(X, \alpha_n)}} \left( B^n_T(X, \alpha_n) - \langle X, \sigma^2 \rangle_T \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),
$$

(5.3)

where the standard normal random variable is independent of $\mathcal{F}$, and

$$
\hat{V}_T^n(X, \alpha_n) = 4 \sum (\Delta^n_i X_{\alpha_n})^6 + 2/3 * \langle \sigma^2, \sigma^2 \rangle_{\text{soph}}.
$$

(5.4)

Remark 5.1. It worth to point out that the spot volatility estimates determine the convergence rate and the limiting process. There are two sources of errors from estimation of spot volatility: one is controlled by the number of local observations $k_n$ and the other is determined by the length of local window $k_n \Delta_n$. The larger $b$, the more number of local observations are available hence the first type of error will decrease. On the other hand, however, a large $b$ also lead to a long local window within which the volatility process varies. Consequently, the second type of error becomes more important. To sum up, as we see in the above theorem, when $b < 1/2$, the first one dominates and otherwise when $1/2 < b < 1$. These two have the same convergence rate when $b = 1/2$.

Remark 5.2. Here, when there is no price jumps but only volatility jumps (with finite variation), the convergence rate does not change. But in Bandi and Renò (2012), volatility jumps will slow down the convergence rate. This difference arises from definition. They define (continuous) leverage effect as a function of spot volatility, hence related to the state space of the latter one. Then intuitively, the dispersion of volatility, reflected by bandwidth and local time, will have a non-trivial impact on the nonparametric estimation of such function. However, our definition is solely based on time domain properties of the underlying stochastic processes and not a function of
spot volatility. Therefore, volatility jumps do not change the convergence rate of the estimate here.

**Theorem 5.2.** Assume (H) for some \( r < 2, \mu \equiv \tilde{\mu} \) (not vanishing) and condition (3.2) with

\[
b \wedge (1 - b) < (2\varpi(2 - r))
\]  

(i) If \( r = 0 \), then both \( \sqrt{u_n}(\hat{\langle X, \sigma^2 \rangle}_T^D - \Delta(X, \sigma^2)_T^D) \) converges stably in law to the process

\[
D_T = \sum_{p \geq 1} \Delta X_{T_p}(V_p^+ - V_p^-)1_{\{T_p \leq T\}}.
\]

where \( V_p^+ \) and \( V_p^- \) are independent normal variables with conditional variance given in (A.14).

(ii) Given some \( \epsilon > 0 \), for \( r \in [0, 2) \), the two processes

\[
\sqrt{u_n}(\hat{\langle X, \sigma^2 \rangle}_T^{D,\epsilon} - \Delta(X, \sigma^2)_T(\epsilon))
\]

converges stably in law to the process

\[
D_{T}^{\epsilon} = \sum_{p \geq 1} \Delta X_{T_p}1_{\{\Delta X_{T_p} > \epsilon\}}(V_p^+ - V_p^-)1_{\{T_p \leq T\}}.
\]

(iii) Assume \( \sigma \) has no Brownian parts, i.e. \( \bar{\sigma} \equiv \bar{b} \equiv 0 \). Then for \( r \in [0, 2) \) and \( b \) satisfying

\[
b \wedge (1 - b) < (2\varpi(2 - r)) \wedge \frac{2}{2 + r},
\]

we have the same conclusion as (i).

**Remark 5.3.** These are direct results from Theorem 3.2 in Jacod and Todorov (2010). Observe that when \( r > 0 \), \( F(x, y, z) = x(z - y) \) fails to satisfy the conditions in that theorem. Hence, except for the case of finite jump activity \( r = 0 \) and another one where volatility process has no Brownian components, there are no CLT results associated with \( \hat{\langle X, \sigma^2 \rangle}_T^D \). Instead, as in case (ii), we can consider the tail discontinuous leverage effect when volatility has Brownian parts and infinity jump activity.
6 Monte-Carlo Simulations

6.1 Continuous leverage effect without jumps

The tasks in this subsection contain: (1) showing the finite sample performance of the CLE estimator; (2) comparing different estimators of the leverage parameter.

Below, we use the following Heston model to simulate log-price process $X_t$ and volatility process $\sigma^2_t$:

\[
\begin{align*}
    dX_t &= (\mu - \sigma^2_t/2)dt + \sigma_t dW_t \\
    d\sigma^2_t &= \kappa(\theta - \sigma^2_t) + \eta\sigma_t(\rho dW_t + \sqrt{1-\rho^2}dV_t), 
\end{align*}
\]

with the following parameter values: $\theta = 0.1, \eta = 0.5, \kappa = 5, \rho = -0.8$ and $\mu = 0.05$.

Finite sample performance  Throughout this section, time is measured in years, hence the time span of one trading day is $T_d = 1/252$. We simulate data using the following three different choices of time span, namely one week (5 trading days), one month (20 trading days) and one year (252 trading days). Within each trading day, the number of observation is 405, corresponding to sampling every one minute for a 6 hours and 45 minutes trading day, the same as the empirical data we are going to use in next section. The simulation is conducted 5000 times.

Figure 1 presents the simulation results. The first row gives the densities of true (red solid line) and estimated (blue dashed line) continuous leverage effects. In the weekly case, where the time span $T$ is relatively small, the variance of the estimates are much higher than that of the true values. As the time span increases, these two densities become more comparable.

To measure the estimation precision, we draw the density of ratios of estimated values to the corresponding true ones. As expected, it increases with time span and sample frequency. Although in the weekly case, the estimation precision is very low, the mean of estimation errors (or order $10^{-6}$) is very small compared to that of the true values (of order $10^{-3}$). And the mean value of the ratios is around 1. Both of these results suggest that our estimator is indeed unbiased.

The graphs in the last two rows show the finite sample normality: in the third row are the densities of standardized estimation errors (blue dashed line), and the standard normal density (red solid line) as benchmark; the last row presents the corresponding QQ plots. Contrary to the first two rows, results in the last two are not very different from each other. In all the three cases, both densities and quantiles
are quite close to those of a standard normal variable. These results further verify that the poor finite sample performance in the weekly case is due to large variance not bias.

We further plot the rejection rate of the test for continuous leverage effect against significant levels in Figure 2. Since that we assume the present of CLE in the simulated model, this shows the finite sample power of the test. In the weekly case, the rejection rates of both one-sided and two sided tests are just above the corresponding significant levels, indicating the performance of the tests is not good enough. However, the tests become much more effective in rejecting the false null hypotheses in the monthly case. Even when significant level is small, we could have reasonably large rejection rate. This is not only because we obtain more observations in the monthly case, but also because that the magnitude of the latent true values become larger compared with the estimated standard error. Finally, in the yearly case, as the significant level increasing, the rejection rate goes to 1 very rapidly. The results here will serve as a guide to choose a suitable time span to test the presence of continuous leverage effect in the empirical study.

Comparison of different leverage parameter estimators  Here, we simulate one year of data at one minute frequency. In Figure 3, we plot the densities of $\rho_{est}$ (using true volatilities) and $\rho_{true}$ (using estimated spot volatilities) based on variation-covariation (left panel) and variance-covariance (right panel). As we discussed in Section 4.1, the ones using estimated volatility introduce large bias while those using true spot volatility do not. The biases in $\rho_{est}^2$s are extremely large and the leverage parameter is not significantly different from zero, hence the puzzle arises. In contrast, although also exist, the biases in $\rho_{est}^1$s are much smaller. More importantly, the mean value of those estimates is around -0.5, which is significantly different from zero. 

<table>
<thead>
<tr>
<th>Table 1: Biases in $\rho_{est}^1$ and $\rho_{est}^2$</th>
<th>max</th>
<th>mean</th>
<th>min</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{est}^1/\rho_{true}^1$</td>
<td>0.9111</td>
<td>0.6451</td>
<td>0.4053</td>
</tr>
<tr>
<td>$\rho_{est}^2/\rho_{true}^2$</td>
<td>0.3801</td>
<td>0.0749</td>
<td>-0.1143</td>
</tr>
<tr>
<td>$\rho_{est}^1 - \rho_{true}^1$</td>
<td>0.8271</td>
<td>0.2812</td>
<td>0.0325</td>
</tr>
<tr>
<td>$\rho_{est}^2 - \rho_{true}^2$</td>
<td>0.9088</td>
<td>0.7377</td>
<td>0.4941</td>
</tr>
</tbody>
</table>

Table 1 gives some quantified measurement of biases in $\rho_{est}^1$ and $\rho_{est}^2$. Under both
Figure 1: First three rows are densities of: (1) true (red and solid) and estimated (blue and dashed) continuous leverage effects; (2) ratios of estimated CLEs and the true values; (3) standardized estimation errors (blue and dashed) and standard normal variable. The last row are quantile-quantile plot of the sample quantiles of standardized estimation errors versus theoretical quantiles from a normal distribution. Each column corresponds to one sampling scheme.
Figure 2: Testing powers with different choices of time span. The red solid curve corresponds to the one-sided test power, and the blue dashed curve to the two-sided test.

Figure 3: Densities of estimated $\rho$ using true (red and solid) and estimated (blue and dashed) volatilities. The true value of $\rho$ is -0.8 (black dash-dot line).
measurement, the biases in the former one are systematically smaller than those in the latter one, implying that the estimate based on variation and covariation outperform the one constructed by variance and covariance in terms of unbiasedness. Besides, as we expected, \( \rho_{\text{est}}^1 \) and \( \rho_{\text{true}}^1 \) have larger degrees of dispersion than their counterparts since they are constructed from random quantities hence should be more sensitive to the realized path of the underlying stochastic process. Furthermore, the mean square errors of \( \rho_{\text{est}}^1 \) and \( \rho_{\text{est}}^2 \) (with respect to the true \( \rho \)) are 0.1040 and 0.5502 respectively. Therefore we can conclude that, without any further correction techniques, the estimate of the leverage parameter based on variation and covariation works better.

6.2 Continuous leverage effect with jumps

Now we consider the following model, which incorporates jumps. Parameters in the diffusion part are the same as before.

\[
\begin{align*}
\frac{dX_t}{dt} &= \left(\mu - \frac{\sigma_t^2}{2}\right)dt + \sigma_t dW_t + J_t^X dN_t \\
\frac{d\sigma_t^2}{dt} &= \kappa(\theta - \sigma_t^2) + \eta \sigma_t (\rho dW_t + \sqrt{1-\rho^2} dV_t) + J_t^\sigma dN_t.
\end{align*}
\]

where \( N_t \) is a Poisson process with intensity \( \lambda \), \( J_t^X \) and \( J_t^\sigma \) are the jump sizes of log-price and volatility processes at time \( t \), respectively.

The density of price jumps is

\[
f_X(x) = \begin{cases} 
\frac{p}{\gamma_d} \exp\left(-\frac{-x}{\gamma_d}\right), & -\infty < x \leq 0; \\
\frac{1-p}{\gamma_u} \exp\left(-\frac{x}{\gamma_u}\right), & 0 < x < \infty,
\end{cases}
\]

where \( \gamma_u, \gamma_d > 0 \) and \( 0 \leq p \leq 1 \). And the one of volatility jumps is

\[
f_\sigma(x) = \frac{1}{\gamma_\sigma} \exp\left(-\frac{x}{\gamma_\sigma}\right), \ x \geq 0.
\]

In our simulation, we set \( \gamma_u = 0.008, \gamma_d = 0.018, p = 0.6 \) and \( \lambda = 300 \). We vary the values of \( \gamma_\sigma \) to study the effects of volatility jumps on the performance of our estimator \( B^n_T(X, \alpha_n) \) and the test power. Recall the result in Theorem 5.1, volatility jumps do not change the convergence rate but increase the asymptotic variance. Based on the results in previous subsection, we choose \( T = 20/252 \) (one trading month) so that it is easier to see the differences in test power with various choices of volatility jump size.

\[^{10}\text{Var}(\rho_{\text{true}}^1) = 0.0410 > \text{Var}(\rho_{\text{true}}^2) = 0.0005 \text{ and } \text{Var}(\rho_{\text{est}}^1) = 0.0174 > \text{Var}(\rho_{\text{est}}^2) = 0.0020.\]
Figure 4: Standardized estimation error densities and testing powers. First row: red solid line corresponds to standard normal density, blue dashed line to standardized estimation error. Second row: red solid line to one-sided test and blue dashed line to two-sided test.
Figure 4 presents the simulation results. Again, the standardized estimation error density approximate standard normal density quite well. When volatility jump sizes are small (the first two cases), the decrease in test power is no very large. While in the third case, when volatility jumps are more comparable the $\theta$, it becomes more significant.

6.3 Discontinuous leverage effect

In this section, we work with a different stochastic volatility model

$$dX_t = \sqrt{V_t^2 + V_t^2}dW_t + \int_{\mathbb{R}} x\mu(dt, dx, dy),$$

$$dV_t^1 = \kappa_1(\theta - V_t^1)dt + \sigma\sqrt{V_t^1}dW'_t,$$

$$dV_t^2 = -\kappa_2V_t^2dt + \int_{\mathbb{R}} y\mu(dt, dx, dy),$$

where $W$ and $W'$ are two independent Brownian motions; the Poisson measures $\mu$ has compensator

$$\nu(dt, dx, dy) = \frac{\lambda}{(h - 1)(u - d)}1_{\{x \in [-h, -l]\}}1_{\{y \in [d, u]\}}$$

for $0 < l < h$ and $0 < d < u$. The parameters used in simulation study is given in Table 2.

<table>
<thead>
<tr>
<th>Case</th>
<th>$\kappa_1$</th>
<th>$\theta$</th>
<th>$\sigma$</th>
<th>$\kappa_2$</th>
<th>$\lambda$</th>
<th>$l$</th>
<th>$h$</th>
<th>$d$</th>
<th>$u$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-j</td>
<td>5.04</td>
<td>0.4</td>
<td>0.2</td>
<td>126</td>
<td>126</td>
<td>0.1</td>
<td>1.0420</td>
<td>0.04</td>
<td>0.76</td>
</tr>
<tr>
<td>II-j</td>
<td>5.04</td>
<td>0.4</td>
<td>0.2</td>
<td>126</td>
<td>252</td>
<td>0.1</td>
<td>0.7197</td>
<td>0.04</td>
<td>0.36</td>
</tr>
<tr>
<td>III-j</td>
<td>5.04</td>
<td>0.4</td>
<td>0.2</td>
<td>126</td>
<td>1008</td>
<td>0.1</td>
<td>0.3275</td>
<td>0.04</td>
<td>0.06</td>
</tr>
</tbody>
</table>

These settings are almost the same with the cases I-j, II-j and III-j in Jacod and Todorov (2010), except that we only include negative return jumps and some parameters are “annualized” the match the setting that one year corresponds to $T = 1$. We set $T = 5/252$ (one trading week) and simulate 5000 replications. On each day, we consider sampling $n = 405$ and $n = 2030$ times. For the calculation of the local volatility estimators we use a window $k_n = [\{(T/\Delta_n)^{0.49}\}]^{11}$. We choose the

\[\text{In the above setting, the choice of } k_n = [2 \ast (T/\Delta_n)^{0.49}] \text{ may generate small bias.}\]
truncation parameters $\alpha = 5\sqrt{BV_T}$, where $BV_T$ stands for the bipower variation in the selected time span $T$, and $\varpi = 0.49$.

The first row in Figure 5 shows the densities of standardized estimation errors. In all cases, there are very close to the standard normal density, showing that the asymptotic normality even holds quite well in finite sample. The send row give testing power in each cases. As we can see, the power increases with sampling frequency, and decreases as the number of jumps gets larger but the jump size becomes smaller.

![Figure 5: Standardized estimation error densities and testing powers. First row: red solid line corresponds to standard normal density, blue dashed line to $n = 1560$, and black dotted line to $n = 7800$. Second row: red solid line to $n = 7800$ and blue dashed line to $n = 1560$.]

6.4 The estimation of leverage parameter $\rho$

In this subsection, we study the estimation of correlation $\rho$, which is also the correlation between the two Wiener processes appearing in $X_t$ and $\sigma_t^2$. We will study everything in the simulation only.

We will consider the following Heston model:

$$dX_t = (\mu - \sigma_t^2/2)dt + \sigma_t dB_t$$
\[ d\sigma_t^2 = \kappa (\alpha - \sigma_t^2) + \gamma \sigma_t dW_t, \]

where \( B \) and \( W \) are two standard Brownian motions with \( \mathbb{E}(dB_t dW_t) = \rho dt \). Note that

In Heston model

\[
\rho = \frac{\text{Cov}(X, \sigma^2)}{\sqrt{\text{Var}(X) \text{Var}(\sigma^2)}} = \frac{\langle X, \sigma^2 \rangle}{\sqrt{\langle X, X \rangle \langle \sigma^2, \sigma^2 \rangle}}.
\]

But in general, the second equality may not hold. We will compare the results with different ways to estimate volatility of volatility. Smoothed naive estimators:

\[
\langle \sigma^2, \sigma^2 \rangle_{\text{naive}}^{S} = \frac{1}{k_n} \sum_{j=k_n+1}^{[t/\Delta_n] - k_n} \left( \tilde{\sigma}_j^+ - \tilde{\sigma}_j^- \right)^2,
\]

where

\[
\tilde{\sigma}_j^+ = \frac{1}{k_n \Delta_n} \sum_{l \in I_t^+(j)} (\Delta_n^X)^2; \quad \tilde{\sigma}_j^- = \frac{1}{k_n \Delta_n} \sum_{l \in I_t^-(j)} (\Delta_n^X)^2.
\]

Smoothed sophisticated estimators:

\[
\langle \sigma^2, \sigma^2 \rangle_{\text{soph}}^{S} = \frac{1}{k_n} \sum_{j=k_n+1}^{[t/\Delta_n] - k_n} \left( \frac{3}{2} (\tilde{\sigma}_j^+ - \tilde{\sigma}_j^-)^2 - \frac{6}{k_n^4} \tilde{\sigma}_j^4 \right),
\]

where

\[
\tilde{\sigma}_j^4 = \frac{1}{6 k_n \Delta_n^2} \sum_{l \in I_t^+(j)} (\Delta_n^X)^4.
\]

Smoothed ratio estimators:

\[
\langle \sigma^2, \sigma^2 \rangle_{\text{ratio}}^{S} = \frac{1}{k_n} \sum_{j=k_n+1}^{[t/\Delta_n] - k_n} \frac{3}{2} (\tilde{\sigma}_j^+ - \tilde{\sigma}_j^-)^2 \frac{1 + \sum_{i=1}^{\infty} r^i}{1 + \sum_{i=1}^{\infty} r^i},
\]

where

\[
r = \frac{\sum_{j=k_n+1}^{[t/\Delta_n] - k_n} \frac{6}{k_n} \tilde{\sigma}_j^4}{\sum_{j=k_n+1}^{[t/\Delta_n] - k_n} \frac{3}{2} (\tilde{\sigma}_j^+ - \tilde{\sigma}_j^-)^2}.
\]
We study the following 7 estimators in simulation.

\[
\begin{align*}
\rho_1 & = \frac{\langle X, \sigma^2 \rangle}{\sqrt{\langle X, X \rangle \langle \sigma^2, \sigma^2 \rangle}} \\
\rho_2 & = \frac{\langle X, \sigma^2 \rangle_T}{\sqrt{\langle X, X \rangle \langle \sigma^2, \sigma^2 \rangle}} \\
\rho_3 & = \frac{\langle X, \sigma^2 \rangle_T^C}{\sqrt{\langle X, X \rangle \langle \sigma^2, \sigma^2 \rangle_{\text{naive}}}} \\
\rho_4 & = \frac{\langle X, \sigma^2 \rangle_T^C}{\sqrt{\langle X, X \rangle \langle \sigma^2, \sigma^2 \rangle_{\text{soph}}}} \\
\rho_5 & = \frac{\langle X, \sigma^2 \rangle_T^C}{\sqrt{\langle X, X \rangle \langle \sigma^2, \sigma^2 \rangle_{\text{naive}}}} \\
\rho_6 & = \frac{\langle X, \sigma^2 \rangle_T^C}{\sqrt{\langle X, X \rangle \langle \sigma^2, \sigma^2 \rangle_{\text{soph}}}} \\
\rho_7 & = \frac{\langle X, \sigma^2 \rangle_T^C}{\sqrt{\langle X, X \rangle \langle \sigma^2, \sigma^2 \rangle_{\text{ratio}}}}
\end{align*}
\]

We will examine how the choice of tuning parameter \( c \) will impact the MSE of the correlation estimation. In the estimation, the parameterization for Heston model is set as: \( \theta = 0.06, \eta = 1, \kappa = 10, \rho = -0.8 \) and \( \mu = 0.05 \). \( T = 22/252 \) (monthly), \( \Delta_n = 1/252/4860 \) (5-sec) or \( \Delta_n = 1/252/23400 \) (1-sec). Simulation is conducted 5000 times.

![Figure 6: Change of MSE along with the increase of tuning parameter \( c \) in the case every 5 sec sampling.](image1)

![Figure 7: Change of MSE along with the increase of tuning parameter \( c \) in the case every 1 sec sampling.](image2)

In the plots above, we only studied estimators \( \rho_3 \) to \( \rho_6 \). Clearly, the MSE will first drop and then increase as the tuning parameter \( c \) increase from 1 to 12.

The estimation of \( \rho \) depends on the estimation of volatility of volatility. But the estimator of volatility of volatility in \( \rho_2 \) and \( \rho_4 \) often produces negative value. This negative estimation for a positive quantity has a big impact on the estimation
of $\rho$. So we studied the adjusted version of the estimator of volatility and volatility as shown in equation (6.1) and hence the correlation estimator $\rho_\tau$.

Here we didn’t study the unsmoothed estimators, since usually the smoothed version has smaller variation. Clearly, $\rho_\tau$ is very close to $\rho_6$, but without any negative estimation of volatility of volatility.

We also studied how accurate the estimator of $\rho$ is relative to the true value by checking the relative error of the estimators. The relative errors of $\rho_1 - \rho_6$ are presented in Figure 10 and Figure 11. The relative errors of $\rho_7$ displays almost the same results as that of $\rho_6$.

The simulation study of estimating $\rho$ suggest that a bigger sample size can provide better estimation and the adjusted estimator of volatility of volatility should be applied in the empirical study. The study also shows that the tuning parameter should be big enough to achieve a better MSE. Of course, it can be optimized by minimizing the asymptotic variance.

Figure 8: MSE of $\rho$ estimation: monthly 5 sec case
The estimation of rho: rhomse monthly 1 sec

Figure 9: MSE of $\rho$ estimation: monthly 1 sec case

Figure 10: Relative error (Est/True-1) of $\rho$ estimation: monthly 5 sec case
Figure 11: Relative error (Est/True-1) of $\rho$ estimation: monthly 1 sec case
7 Empirical application

Microstructure noise It is a well-known fact that in high frequency financial applications, the presence of microstructure noise in the prices is noneligible. To deal with the microstructure noise, we will apply pre-averaging method.

The contaminated log return process $Y_t$ is observed every $\Delta t_{n,i} = \frac{T}{n}$ units of time, at times $0 = t_{n,0} < t_{n,1} < t_{n,2} < \ldots < t_{n,n} = T$.

Assumption

$Y_t = X_t + \epsilon_t$, where $\epsilon_t's$ are i.i.d. $N(0, a^2)$ and $\epsilon_t \perp \perp$ the $W$ and $B$ processes, for all $t$. \hfill (7.1)

We also assume that $\epsilon_t$'s have finite fourth moment, and are independent of both return and volatility processes.

Blocks are defined on a much less dense grid of $\tau_{n,i}$, also spanning $[0, T]$, so that

block # $i = \{t_{n,j} : \tau_{n,i} \leq t_{n,j} < \tau_{n,i+1}\}$ \hfill (7.2)

(the last block, however, includes $T$). We define the block size by

$M_{n,i} = \#\{j : \tau_{n,i} \leq t_{n,j} < \tau_{n,i+1}\}$. \hfill (7.3)

In principle, the block size $M_{n,i}$ can vary across the trading period $[0, T]$, but for this development we take $M_{n,i} = M$: it depends on the sample size $n$, but not on the block index $i$.

We then use as an estimated value of the efficient price in the time period $[\tau_{n,i}, \tau_{n,i+1})$:

$\hat{X}_{\tau_{n,i}} = \frac{1}{M_n} \sum_{t_{n,j} \in [\tau_{n,i}, \tau_{n,i+1})} Y_{t_{n,j}}$

Let $I_n^-(j) = \{j - k_n M - 1, j - k_n M, \ldots, j - M\}$ if $j > k_n M$ and $I_n^+(j) = \{j + M, j + 2M, \ldots, j + k_n M + 1\}$ define two local windows in time of length $k_n M \Delta_n$ just before and after time $j \Delta_n$. Denote the truncated increment of $X$ by $\Delta_j^n \hat{X}_{\alpha_n} = \Delta_j^n \hat{X} \cdot 1_{\{\|\Delta_j^n \hat{X}\| \leq \alpha_n\}}$. Then we can define
\begin{align*}
\langle \hat{X}, \sigma^2 \rangle_T^C &= \frac{3}{2} \sum_{i=k_n*M+1}^{n-k_n*M} \Delta_i^n \hat{X}_{\alpha_n}(\hat{\sigma}^2_{i^+} - \hat{\sigma}^2_{i^-}) \\
\hat{\sigma}^2_{i^+} &= \frac{1}{k_n \Delta_n} \sum_{j \in I_n^+(i)} \Delta_j^n \hat{X}_{\alpha_n}^2, \\
\hat{\sigma}^2_{i^-} &= \frac{1}{k_n \Delta_n} \sum_{j \in I_n^-(i)} \Delta_j^n \hat{X}_{\alpha_n}^2.
\end{align*}

\textbf{Theorem 7.1.} Under the same condition as in Theorem 4.1

\begin{equation}
\langle \hat{X}, \sigma^2 \rangle_T^C \overset{u.c.p.}{\longrightarrow} \langle X, \sigma^2 \rangle_T^C = \int_0^T 2 \sigma_t^2 \tilde{\sigma}_t dt,
\end{equation}

Here we only give the law of large number for the estimation instead of the CLT due to the significantly more involved mathematical approach. Also since we are more interested in the estimation value of correlation \( \rho \), the LLN itself is enough for this purpose. The slower convergence rate with the pre-averaging method is not a big concern, as in the previous simulation study we already conclude that a larger sample size is necessary for a better estimation of \( \rho \). In practice, monthly estimation of \( \rho \) will well service the purpose and monthly second data should provide a big enough sample size in the empirical study.

For the estimation of integrated volatility and volatility of volatility, we will also apply the pre-averaging methods. For the details of those estimations, we will refer to Jacod, etc (2009) and Vetter(2012).

\textbf{Choices of time span} \hspace{1em} As shown in the simulation study, we need enough sample size to better estimate the correlation \( \rho \). In the empirical study, we estimate the monthly correlation.

\textbf{Window size selection} \hspace{1em} Since Dow almost has every second transaction, this frequency provides proxies to second data case in the simulation. In the simulation, the sample size is 23,400 and the window for pre averaging is around \( c\sqrt{23,400} \) which corresponding to every 3-minute averaging (if \( c \sim 1 \)) and every 5-minute (if \( c \sim 2 \)). Furthermore, in the second step estimation of all three quantities in \( \rho_6 \) or \( \rho_7 \), we take the tuning parameter as 4 for the every 3-min pre averaging case, and 3 for every 5-min pre averaging case. This is also suggested by the simulation study where usually the minimum MSE is achieved when the tuning parameter is between 3 and 5.
7.1 Empirical results about CLE and DLE

Again, the truncation parameters are given by: \( \alpha = 5 \times \sqrt{BV_T} \) and \( \varpi = 0.49 \). In Table 3, we can see that with weekly data, the rejection rates (of the one-sided test) are not high, while we reject the null hypothesis of the absence of CLE more often with monthly data. These are consistent with what we have found in the simulation study, that is, the testing power is quite low with weekly data. Moreover, the testing power with monthly data is also limited. However, had there been no continuous leverage effect within each week, we would not find more supportive evidence for its present with monthly data. Therefore, we would like to conclude that there is positive evidence for the presence of continuous leverage effect.

<table>
<thead>
<tr>
<th>Rejection rate</th>
<th>( \alpha=1% )</th>
<th>( \alpha=5% )</th>
<th>( \alpha=10% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c=1 )</td>
<td>4.58%</td>
<td>15.34%</td>
<td>21.71%</td>
</tr>
<tr>
<td>( c=2 )</td>
<td>4.98%</td>
<td>15.94%</td>
<td>25.50%</td>
</tr>
<tr>
<td>4-week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( c=1 )</td>
<td>39.68%</td>
<td>60.32%</td>
<td>67.46%</td>
</tr>
<tr>
<td>( c=2 )</td>
<td>36.51%</td>
<td>55.56%</td>
<td>65.08%</td>
</tr>
</tbody>
</table>

Table 4: Testing for the absence of DLE

<table>
<thead>
<tr>
<th>Jump size</th>
<th># of weeks</th>
<th>DLE</th>
<th>Rejection rate</th>
<th>( \alpha=1% )</th>
<th>( \alpha=5% )</th>
<th>( \alpha=10% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>any size</td>
<td>495</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td></td>
<td></td>
<td>62.83%</td>
<td>71.31%</td>
<td>74.95%</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td>87.68%</td>
<td>91.52%</td>
<td>92.93%</td>
</tr>
<tr>
<td>all</td>
<td></td>
<td></td>
<td></td>
<td>57.01%</td>
<td>61.99%</td>
<td>67.42%</td>
</tr>
<tr>
<td>&gt; 0.20%</td>
<td>221</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td></td>
<td></td>
<td>35.29%</td>
<td>41.18%</td>
<td>43.44%</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td>46.15%</td>
<td>49.77%</td>
<td>53.39%</td>
</tr>
<tr>
<td>all</td>
<td></td>
<td></td>
<td></td>
<td>51.79%</td>
<td>58.93%</td>
<td>62.50%</td>
</tr>
<tr>
<td>&gt; 0.30%</td>
<td>112</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td></td>
<td></td>
<td>24.11%</td>
<td>28.57%</td>
<td>30.36%</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td>39.29%</td>
<td>39.29%</td>
<td>42.86%</td>
</tr>
<tr>
<td>all</td>
<td></td>
<td></td>
<td></td>
<td>47.54%</td>
<td>52.46%</td>
<td>55.74%</td>
</tr>
<tr>
<td>&gt; 0.40%</td>
<td>61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td></td>
<td></td>
<td></td>
<td>18.03%</td>
<td>22.95%</td>
<td>26.23%</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td>34.43%</td>
<td>39.34%</td>
<td>39.34%</td>
</tr>
</tbody>
</table>

As for the discontinuous leverage effect case, we first employed the procedure introduced in Aït-Sahalia and Jacod (2009) to test whether the returns jump or
not within each week. Then we applied our estimation and testing procedure for DLE to those weeks containing return jumps. Table 7.1 gives the testing results for the absence of DLE. Here, we further classify the DLE according to the sign of the return jump. The letters ‘p’ and ‘n’ in the above table stand for the discontinuous leverage effects from positive and negative return jumps respectively, while ‘all’ means all return jumps together. Meanwhile, we also consider the tail DLE with three thresholds, namely 0.2%, 0.3% and 0.4%. It is easy to see that the rejection rate increases drastically as \( c \) goes from 1 to 2. This may either due to more precise estimates of local volatilities, or the small bias we found in simulation study, or both. To be conservative, we mainly rely on the \( c = 1 \) case. Even so, the evidence for the presence of various types of DLE is pretty strong.

Additionally, we found two very interesting phenomena from the above results. First, at any threshold level, negative return jumps are more likely to be accompanied by volatility jumps than positive ones. Second, given any type of DLE, similar conclusion holds for large return jumps.

### 7.2 Empirical results about \( \rho \)

In the empirical study, we are interested in estimating the correlation \( \rho \). We will apply \( \rho_6 \) (or \( \rho_7 \) when the estimation of volatility of volatility is negative in \( \rho_6 \)). We apply the Dow data from year 2007 to 2012, which covers the period of the recent financial crises. We collect the trades from 9:30am to 4:00 pm. We pre-average data in two different ways, every 5 minutes and every 3 minutes. The estimation of \( \rho \) is given in the following table:
Table 5: Estimation of monthly $\rho$ in year 2007-2012.

<table>
<thead>
<tr>
<th>Month</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>07, 5mn</td>
<td>-0.46</td>
<td>-0.23</td>
<td>-0.32</td>
<td>-0.49</td>
<td>0.09</td>
<td>-0.81</td>
<td>-0.76</td>
<td>-0.40</td>
<td>-0.62</td>
<td>-0.43</td>
<td>-0.50</td>
<td>-0.63</td>
</tr>
<tr>
<td>07, 3mn</td>
<td>-0.49</td>
<td>-0.32</td>
<td>-0.29</td>
<td>0.03</td>
<td>-0.21</td>
<td>-0.49</td>
<td>-0.31</td>
<td>-0.79</td>
<td>-0.23</td>
<td>-0.70</td>
<td>-0.18</td>
<td>-0.36</td>
</tr>
<tr>
<td>08, 5mn</td>
<td>-0.64</td>
<td>-0.24</td>
<td>-0.44</td>
<td>-0.11</td>
<td>-0.06</td>
<td>-0.25</td>
<td>-0.42</td>
<td>-0.95</td>
<td>-0.47</td>
<td>-0.76</td>
<td>-0.65</td>
<td>-0.34</td>
</tr>
<tr>
<td>08, 3mn</td>
<td>-0.46</td>
<td>-0.33</td>
<td>-0.37</td>
<td>-0.02</td>
<td>-0.19</td>
<td>-0.55</td>
<td>-0.32</td>
<td>-0.93</td>
<td>-0.47</td>
<td>-0.84</td>
<td>-0.49</td>
<td>-0.26</td>
</tr>
<tr>
<td>09, 5mn</td>
<td>-0.33</td>
<td>-0.38</td>
<td>-0.43</td>
<td>-0.23</td>
<td>-0.25</td>
<td>-0.05</td>
<td>-0.40</td>
<td>-0.03</td>
<td>-0.26</td>
<td>-0.46</td>
<td>-0.71</td>
<td>-0.72</td>
</tr>
<tr>
<td>09, 3mn</td>
<td>-0.07</td>
<td>-0.66</td>
<td>-0.55</td>
<td>-0.22</td>
<td>-0.32</td>
<td>-0.41</td>
<td>-0.45</td>
<td>-0.02</td>
<td>-0.38</td>
<td>-0.53</td>
<td>-0.6</td>
<td>-0.78</td>
</tr>
<tr>
<td>10, 5mn</td>
<td>-0.19</td>
<td>-0.86</td>
<td>-0.32</td>
<td>-0.49</td>
<td>-0.49</td>
<td>-0.9</td>
<td>-0.66</td>
<td>-0.35</td>
<td>0.01</td>
<td>-0.22</td>
<td>-0.30</td>
<td>-0.245</td>
</tr>
<tr>
<td>10, 3mn</td>
<td>-0.22</td>
<td>-0.78</td>
<td>-0.34</td>
<td>-0.39</td>
<td>-0.56</td>
<td>-0.74</td>
<td>-0.58</td>
<td>-0.31</td>
<td>-0.16</td>
<td>-0.25</td>
<td>-0.28</td>
<td>-0.19</td>
</tr>
<tr>
<td>11, 5mn</td>
<td>-0.37</td>
<td>-0.25</td>
<td>-0.02</td>
<td>-0.25</td>
<td>-0.36</td>
<td>-0.79</td>
<td>-0.34</td>
<td>-0.36</td>
<td>-0.64</td>
<td>-0.60</td>
<td>-0.36</td>
<td>-0.10</td>
</tr>
<tr>
<td>11, 3mn</td>
<td>-0.44</td>
<td>-0.30</td>
<td>-0.05</td>
<td>-0.35</td>
<td>-0.48</td>
<td>-0.66</td>
<td>-0.31</td>
<td>-0.49</td>
<td>-0.68</td>
<td>-0.58</td>
<td>-0.34</td>
<td>-0.09</td>
</tr>
<tr>
<td>12, 5mn</td>
<td>-0.002</td>
<td>-0.33</td>
<td>-0.93</td>
<td>-0.45</td>
<td>-0.40</td>
<td>-0.07</td>
<td>-0.68</td>
<td>-0.34</td>
<td>-0.14</td>
<td>-0.69</td>
<td>-0.72</td>
<td>-0.24</td>
</tr>
<tr>
<td>12, 3mn</td>
<td>-0.04</td>
<td>-0.30</td>
<td>-0.55</td>
<td>-0.57</td>
<td>-0.45</td>
<td>-0.17</td>
<td>-0.46</td>
<td>-0.47</td>
<td>0.01</td>
<td>-0.49</td>
<td>-0.41</td>
<td>-0.31</td>
</tr>
</tbody>
</table>

Clearly, in year 2008, the correlation $\rho$ was the highest in August and kept the trend until the end of the year. Christina: Here I just played with Dow data and it seems $\rho$ estimation looks not bad. I guess there must be more empirical evidence to support the finding above. Please feel free to provide more insights!

8 Financial Implications

In this section, we are going to briefly discuss the financial implications of the continuous leverage effect and co-jumping. To keep notions simple, consider the following jump-diffusion model instead of the general one in Section 2,

$$dX_t = a_t dt + \sqrt{V_t} dW_t + J_x dN_t,$$

$$dV_t = \tilde{a}_t dt + \tilde{\sigma}_t dW_t + \tilde{b}_t dB_t + J_v d\tilde{N}_t,$$

where $N_t$ and $\tilde{N}_t$ are two counting processes with intensity $\lambda$, $J_x$ and $J_v$ are jump sizes with probability density functions $p_x$ and $p_v$ respectively. According to the presence of continuous leverage effect and/or co-jumping, we classify the above model under three cases:

- **Case 1.** $\tilde{\sigma}_t = 0$, $N_t$ and $\tilde{N}_t$ are independent.
- **Case 2.** $\tilde{\sigma}_t \neq 0$, $N_t$ and $\tilde{N}_t$ are independent.
- **Case 3.** $\tilde{\sigma}_t \neq 0$, $N_t \equiv \tilde{N}_t$ \textsuperscript{12}.

\textsuperscript{12}For simplicity, we assume the jump sizes, $J_x$ and $J_v$, are independent
8.1 Asset Pricing

Assume \( X_t \) is the return process of an underlying asset. Then value of any derivative or option which depends on this underlying asset could be written as a function \( U(t, X, V) \). Assume the risk free interest rate is a constant \( r \). Following the no-arbitrage assumption, we should have that the present value equals to the discounted one. Mathematically, we get the following equation:

\[
U(t, x, v) \equiv e^{-r(T-t)} \mathbb{E}_{t,x,v}[U(T, X_T, V_T)], \text{ for any } t < T.
\]

Consequently, the instantaneous gains from saving and investing on this derivative or option (with the same amount of money) should be the same, that is:

\[
\lim_{T \downarrow t} e^{-r(T-t)} - 1 \mathbb{E}_{t,x,v}[U(T, X_T, V_T)] - U(t, x, v) = \lim_{T \downarrow t} \frac{\mathbb{E}_{t,x,v}[U(T, X_T, V_T)] - U(t, x, v)}{T - t},
\]

which gives the following formula:

\[
rU(t, x, v) = \mathcal{L}U(t, x, v),
\]

where \( \mathcal{L} \) is the infinitesimal generator of the two dimensional process \((X_t, V_t)\). In what follows, we denote \( \mathcal{L}_i \) the corresponding generator in Case \( i \) for \( i = 1, 2, 3 \).

\[
\mathcal{L}_1 U(t, x, v) = \frac{\partial U}{\partial t} + a_t \frac{\partial U}{\partial X} + \tilde{a}_t \frac{\partial U}{\partial V} + \frac{1}{2} v \frac{\partial^2 U}{\partial X^2} + \frac{1}{2} \tilde{v}_t \frac{\partial^2 U}{\partial V^2} + \lambda \int_\mathbb{R} U(t, x + J_x, v)p_x(J_x)dJ_x
\]

\[
+ \lambda \int_\mathbb{R} U(t, x, v + J_v)p_v(J_v)dJ_v - 2\lambda U(t, x, v),
\]

\[
\mathcal{L}_2 U(t, x, v) = \mathcal{L}_1 U(t, x, v) + \tilde{\sigma}_t \sqrt{v} \frac{\partial^2 U}{\partial X \partial V},
\]

\[
\mathcal{L}_3 U(t, x, v) = \mathcal{L}_2 U(t, x, v) + \lambda \int \int_{\mathbb{R}^2} \left[ U(t, x + J_x, v + J_v) + U(t, x, v) - U(t, x + J_x, v) - U(t, x, v + J_v) \right] p_x(J_x)p_v(J_v)dJ_xdJ_v.
\]

By Taylor expansion, we could simplify the integrand as

\[
U(t, x + J_x, v + J_v) + U(t, x, v) - U(t, x + J_x, v) - U(t, x, v + J_v)
\]

\[
= J_xJ_v \frac{\partial^2 U}{\partial X \partial V} + R_{t,x,v}(J_x, J_v).
\]

And assuming the remainder \( R_{t,x,v}(J_x, J_v) \) is negligible, we get

\[
\mathcal{L}_3 U(t, x, v) = \mathcal{L}_1 U(t, x, v) + \tilde{\sigma}_t \sqrt{v} \frac{\partial^2 U}{\partial X \partial V} + \lambda \mu_x \mu_v \frac{\partial^2 U}{\partial X \partial V},
\]
where \( \mu_x \) and \( \mu_v \) are the means of \( J_x \) and \( J_v \) respectively.

In general, it is not easy to solve equation (8.1). Hence, we can not directly compare the solutions to (8.1) in the three cases. However, it is readily to see that the second term in equation (8.2) arises from the continuous leverage effect, while the third one captures the co-jump effect. As long as \( \frac{\partial^2 U}{\partial X \partial V} \neq 0 \), that is, \( U(t, x, v) \) does not take the form \( U_1(t, x) + U_2(t, v) \), then the last two terms in (8.2) will not vanish. Consequently, the continuous leverage effect and co-jump will have non-trivial impacts on asset pricing.

### 8.2 Hedging

Denote \( U_i(t, x, v) \) the solution to equation (8.1) with \( \mathcal{L} \) replaced by \( \mathcal{L}_i \). Let \( i < j \) and assume \( \mathcal{L}_j \) is the true infinitesimal generator but we mistakenly set it as \( \mathcal{L}_i \). Therefore, the error, for example, in delta hedging ratio could be written as

\[
\frac{\partial U_j}{\partial X} - \frac{\partial U_i}{\partial X} = \frac{1}{r} \left[ \frac{\partial \mathcal{L}_j(U_j - U_i)}{\partial X} + \frac{\partial (\mathcal{L}_j - \mathcal{L}_i)U_i}{\partial X} \right] \\
\text{or} = \frac{1}{r} \left[ \frac{\partial \mathcal{L}_i(U_j - U_i)}{\partial X} + \frac{\partial (\mathcal{L}_j - \mathcal{L}_i)U_j}{\partial X} \right].
\]

On the right hand side of each line, the first term could be interpreted as the error arising from pricing error, which is further transformed by a specified infinitesimal generator (right or wrong). While the second one represents the error from the misspecification of the generator, which operates on a given pricing function (wrong or right). Since the differences in \( \mathcal{L}_i \) and \( \mathcal{L}_j \), hence \( U_i \) and \( U_j \), come from either continuous leverage effect or co-jump or both, we could conclude that these two effects also play non-negligible roles in hedging.
Appendix: Proofs

A Preliminary Results

Recall that $X_t = X'_t + X''_t$, where

$$X''_t = \delta \star \mu_t = \int_0^t \int \delta(x) \mu(ds, dx),$$

$$X'_t = X_t + \int_0^t a'_s ds + \int_0^t \sigma_s dW_s \quad \text{with} \quad a'_s = a_s - \int_{\delta(t,x) \leq \kappa} \delta(t, x) \lambda(dx).$$

Observe that $X'$ contains no jump hence has continuous paths almost surely, while $X''$ is a pure jump process.

We first prove that the above LLN and CLT with $B^n_T(X, \alpha_n)$ replaced by $B^n_T(X')$ and next show that the difference $B^n_T(X, \alpha_n) - B^n_T(X')$ is asymptotically negligible.

A.1 Localization

As shown in, for example, Jacod and Protter (2011), localization is a simple but very important tool for proving limit theorems for discretized processes over a finite time interval. With the localization procedure, we can strengthen assumption (H) (and (HF)) by replacing locally boundedness conditions by boundedness, which is much stronger. More precisely, we set

Assumption (SH): We have (H) and for some constant $\Lambda$ and all $(\omega, t, x)$:

$$
\begin{align*}
\|a_t(\omega)\| &\leq \Lambda, \quad \|\sigma_t^2(\omega)\| \leq \Lambda, \quad \|X_t(\omega)\| \leq \Lambda, \quad \|\tilde{a}_t(\omega)\| \leq \Lambda; \\
\|\tilde{\sigma}_t(\omega)\| &\leq \Lambda, \quad \|\tilde{b}_t(\omega)\| \leq \Lambda, \quad \|\tilde{\delta}(\omega, t, x)\| \leq \Lambda(\tilde{\gamma}(x) \wedge 1);
\end{align*}
$$

(A.1)

Coefficients of $\tilde{\sigma}$ are also bounded by $\Lambda$.

Note if all the these are satisfied, we can choose $\tilde{\gamma} < 1$. Assume (SH), if we further choose the truncation parameter $\kappa = 2\Lambda$, then (2.1) and (2.2) can be written in more concise forms:

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t \sigma_s dW_s + \delta \star (\mu - \nu)_t, \quad (A.2)$$

$$\sigma_t = \sigma_0^2 + \int_0^t \tilde{a}_s ds + \int_0^t \tilde{\sigma}_s dW_s + \int_0^t \tilde{b}_s dB_s + \tilde{\delta} \star (\tilde{\mu} - \tilde{\nu})_t, \quad (A.3)$$

37
and consequently
\[
\sigma^p_t = \int_0^t \tilde{a}(p)ds + \int_0^t p \sigma^{p-1}_s \left( \tilde{\sigma}_sdW_s + \tilde{b}_sdB_s \right) + g(p) \ast (\tilde{\mu} - \tilde{\nu})_t. \tag{A.4}
\]

Next, for some \(t_0, s \geq 0\) and integer \(n\), then apply Itô’s lemma to \(Y_s = X'_{t_0+s} - X'_{t_0}\)
\[
Y^n_s = \int_0^s \left( nY^{n-1}_u a'_{t_0+u} + \frac{n(n-1)}{2} Y^{n-2}_u (\sigma^2_{t_0+u-}) \right) du + n \int_0^s Y^{n-1}_u (\sigma_{t_0+u-}) dW_{t_0+u}. \tag{A.5}
\]

In what follows, we will frequently use the above two equations.

A.2 Auxiliary lemmas

First we introduce a very useful Lemma in proving convergence in probability.

**Lemma A.1.** Let \(\{\zeta^n_i\}\) be an array of random variables and each \(\zeta^n_i\) is \(\mathcal{F}_{t^n_i}\)-measurable and \(N_t = \left\lfloor \frac{t}{k_n \Delta_n} \right\rfloor\) for any \(t > 0\). If we have
\[
\sum_{i=1}^{N_t} \mathbb{E}_{t^n_i-1} \left\| \zeta^n_i \right\| \xrightarrow{p} 0, \quad \forall t > 0, \tag{A.6}
\]
or we have
\[
\sum_{i=1}^{N_t} \mathbb{E}_{t^n_i-1} (\zeta^n_i) \xrightarrow{u.c.p.} 0, \tag{A.7}
\]
\[
\sum_{i=1}^{N_t} \mathbb{E}_{t^n_i-1} (\left\| \zeta^n_i \right\|^2) \xrightarrow{p} 0, \quad \forall t > 0, \tag{A.8}
\]
for some continuous adapted process of finite variation \(A\), and if further \(\zeta^n_i\) is \(\mathcal{F}_{t^n_i}\)-measurable, then the following conclusion holds,
\[
\sum_{i=1}^{N_t} \zeta^n_i \xrightarrow{u.c.p.} 0.
\]

Next one is about the stable convergence of triangular arrays.

**Lemma A.2.** If we have (A.7) some continuous adapted process of finite variation \(A\), and also the following conditions
\[
\sum_{i=1}^{N_t} \left( \mathbb{E}_{t^n_i-1} (\left\| \zeta^n_i \right\|^2) - \left( \mathbb{E}_{t^n_i-1} (\zeta^n_i) \right)^2 \right) \xrightarrow{p} C_t, \quad \forall t > 0, \tag{A.9}
\]

38
\[
\sum_{i=1}^{N_t} \mathbb{E}_{\tau_{i-1}} \left( \| \zeta_{n_i}^{n} \|^4 \right) \xrightarrow{p} 0 \quad \forall \ t > 0, \quad \text{(A.10)}
\]

\[
\sum_{i=1}^{N_t} \mathbb{E}_{\tau_{i-1}} \left( \zeta_{n_i}^{n} \Delta_{n_i}^{n} N \right) \xrightarrow{p} 0 \quad \forall \ t > 0, \quad \text{(A.11)}
\]

where \( C \) is a continuous adapted process and \( N \) is either \( W \) or a bound martingale orthogonal to \( W \). Then we have

\[
\sum_{i=1}^{N_t} \zeta_{n_i}^{n} \xrightarrow{L^2} A_t + B_t,
\]

where \( B \) is a continuous process defined on an extension \((\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})\) of the space \( \Omega, \mathcal{F}, \mathbb{P} \) and which, conditionally on the \( \sigma \)-field \( \mathcal{F} \), is a centered Gaussian process with independent increments satisfying \( \tilde{\mathbb{E}}(B_t^2 | \mathcal{F}) = C_t \).

For more discussions, one can refer to Jacod (2009), where a multi-dimension version is provided.

**Lemma A.3.** \( Y \) is an Itô semimartingale represented by (2.2) with \( \sigma \) replaced by \( Y \). Furthermore, assume its coefficients satisfy (SH), hence it has form (A.3) (with \( \sigma \) replaced by \( Y \)), then the process \( Y_{\tau + u}^{J} = \tilde{\delta} \ast (\tilde{\mu} - \tilde{\nu}) \) is a locally square integrable martingale and for any finite stopping times \( T, s > 0 \) and \( q \geq 2 \) we have

\[
\mathbb{E}\left( \sup_{0 \leq u \leq s} \| Y_{T+u}^{J} - Y_T^{J} \|^q \middle| \mathcal{F}_T \right) \leq K_q s \quad \text{and} \quad \mathbb{E}\left( \sup_{0 \leq u \leq s} \| Y_{T+u} - Y_T \|^q \middle| \mathcal{F}_T \right) \leq K_q s,
\]

as \( s \) approaching zero.

Finally, we introduce Lemma 13.2.6 in Jacod and Protter (2011) with some minor changes of notations. Define a \( d \)-dimensional vector

\[
\overline{X}_{j,k} = \left( \frac{\Delta_j X}{\sqrt{\Delta_n}}, \ldots, \frac{\Delta_{j+k-1} X}{\sqrt{\Delta_n}}, \frac{\Delta_{j+k} X'}{\sqrt{\Delta_n}}, \ldots, \frac{\Delta_{j+d-1} X'}{\sqrt{\Delta_n}} \right).
\]

Then, recalling \( \alpha_n = \alpha_{\Delta_n}^{\omega} \), we set

\[
F_{u}(x_1, \ldots, x_d) = F(x_1, \ldots, x_d) \prod_{j=1}^{d} 1_{\{\|x_j\| \leq u\}} \quad \text{for} \ u > 0,
\]

\[
\phi_{j,k}^{n} = F_{\alpha_n/\sqrt{\Delta_n}}(\overline{X}_{j,k+1}^{n}) - F_{\alpha_n/\sqrt{\Delta_n}}(\overline{X}_{j,k}^{n}) \quad \text{for} \ j = 0, 1, \ldots, d - 1.
\]
Lemma A.4. Assume (SH) for some $r \in (0, 2]$ and suppose $F$ is a $d$-dimensional function satisfying

$$
\|F(x_1, \cdots, x_{j-1}, x_j + y, x_{j+1}, \cdots, x_d) - F(x_1, \cdots, x_d)\| \leq K(\|y\|^s + \|y\|^{s'}) \prod_{l=1}^{d}(1 + \|x_l\|^{p'}),
$$

for some $p' \geq 0$ and $s' \geq 1 \geq s > 0$. Let $m \geq 1$ and suppose that $d = 1$ or $\varpi \geq \frac{m(p'\sqrt{2})^{-2}}{2(m(p'\sqrt{2}) - r)}$. Then, with $\theta > 0$ arbitrarily fixed when $r > 1$ and $\theta = 0$ when $r \leq 1$, there is a sequence $\psi_n$ (depending on $m, s, s', \theta$) of positive numbers going to 0 as $n \to \infty$, such that

$$
\mathbb{E}[\|\sigma_{j,k}^n\|^{m} | \mathcal{F}_{t_n-1}] \leq \left(\frac{2r}{n}(1 - m\varpi) - \theta + \Delta_n(1 - r\varpi)(1^{-m\varpi}) - ms'\frac{1 - 2m\varpi}{\theta}\right)^{\psi_n}.
$$

A.3 Preliminary results

Lemma A.5. Let $u_n = n^{b\wedge(1-b)}$, at any time $t_i$ we have

$$
\sqrt{u_n}\left(\hat{\sigma}_{i+}^2 - \sigma_{i+}^2, \hat{\sigma}_{i-}^2 - \sigma_{i-}^2\right) \xrightarrow{\mathcal{F}_{t_i}} (V_{i+}^+, V_{i-}^-),
$$

where $(V_{i+}^+, V_{i-}^-)$ is a vector of normal random variables independent with $\mathcal{F}$. They have zero $\mathcal{F}$-conditional covariance and

$$
\mathbb{E}((V_{i\pm})^2 | \mathcal{F}) = 2\sigma_{i\pm}^4 T^b 1_{(b \in [0,0.5])} + \frac{1}{3}\left(\frac{d(\sigma_{i\pm}^2, \sigma_{i\pm}^2)}{dt}\right|_{t=t_{i\pm}})T^{-b} 1_{(b \in [0.5,1])}.
$$

Lemma A.6. For $i \neq j$ and $|i - j| \leq k_n$, we have

$$
\mathbb{E}_{i \wedge j - k_n}[(\Delta_{n}^u X)(\Delta_{n}^u X) R_i R_j] = O_p(\Delta_n^2),
$$

where $R_i$ is one of $\{\hat{\sigma}_{i+}^2, \hat{\sigma}_{i-}^2, \Delta_{n}^u \sigma^2\}$ and $R_j$ is one of $\{\hat{\sigma}_{j+}^2, \hat{\sigma}_{j-}^2, \Delta_{n}^u \sigma^2\}$.

Proof. First of all, it is easy to verify the result when $R_i = \Delta_{n}^u \sigma^2$ and $R_j = \Delta_{n}^u \sigma^2$.

Next, when $R_i \neq \Delta_{n}^u \sigma^2$, it amounts to prove that

$$
\mathbb{E}_{i \wedge j - k_n}[(\Delta_{n}^u X)(\Delta_{n}^u X)(\Delta_{n}^u X)^2(\Delta_{n}^u X)^2] = O_p(\Delta_n^2),
$$

where $u \in I_n^+(i)$ and $v \in I_n^+(j)$. The explicit expression of the right hand side depends on the relative order of $i, j, u, v$ and whether $u = v$. But its order with regard to $\Delta_n$ remains the same in all cases. To save space, we just show the calculation of one typical case. Let $i < j < u < v$, and denote $(i - k_n)\Delta_n$ as $\tau_0$, we have

$$
\mathbb{E}_{\tau_0}[(\Delta_{n}^u X')(\Delta_{n}^u X')(\Delta_{n}^u X')^2(\Delta_{n}^u X')^2]
$$
\[\mathbb{E}_{\tau_0}[(\Delta^n X')(\Delta^n X')(\Delta^n X')^2(\sigma^2_{\tau_{i+1}})] \Delta_n = \mathbb{E}_{t-k_n}[((\Delta_j^n X')(\Delta^n X')(\sigma^4_{\tau_{i+1}-})) \Delta^2_n\]
\[= \mathbb{E}_{\tau_0}((\Delta^n X')(\sigma^4_{\tau_{j-1}-})(a'_{\tau_{j-1}+} + 4\sigma^2_{\tau_{j-1}-})) \Delta^3_n\]
\[= (\sigma^4_{\tau_0-})\left(((a'_{\tau_0})^2 + \Gamma^{(X,\sigma)}_{\tau_0} + 4a'_{\tau_0}\tilde{\sigma}_{\tau_0}) + 4\left(a'_{\tau_0}\tilde{\sigma}_{\tau_0} + \Gamma^{(X,\tilde{\sigma})}_{\tau_0} + 4\tilde{\sigma}^2_{\tau_0}\right)\right) \Delta^4_n,\]

where \(\Gamma^{(X',Y)} := \mathbb{E}_0[d(X', Y)_t]/dt\).

Finally, the proof when exactly one of \(R_i\) and \(R_j\) equals to \(\Delta^n_i\sigma^2\) or \(\Delta^n_j\sigma^2\) is similar. Hence we omit the details.

## B Proof of Main Theorems

### B.1 Continuous leverage effect estimator

Using the localization technique, we can and will assume \((SH)\). Recall that in equation (4.1), we decompose the estimation error into three components: discretization error, volatility estimation error and truncation error. Here, we further classify discretization error into three parts. To introduce the notations, first of all, we decompose the volatility process into three parts

\[\sigma^2_t = \sigma^2_{t^2} + \sigma^2_{t^j} + \sigma^2_{t^d},\]

where \(\sigma^2_{t^2}\) is the continuous part, \(\sigma^2_{t^j}\) the joint jumps with price process, and \(\sigma^2_{t^d}\) the disjoint jumps.

Now we rewrite the estimation error profile as

\[\left(\langle X, \sigma^2 \rangle_T^n - \langle X, \sigma^2 \rangle_T^n\right) = T(\alpha_n)_t^n + V^n_t + D(1)^n_t + D(2)^n_t + D(3)^n_t,\]

where

\[T(\alpha_n)_t^n = \sum_{i=k_n+1}^{[t/\Delta_n]-k_n} \left(\Delta^n_i \alpha_n (\tilde{\sigma}^2_{i+} - \tilde{\sigma}^2_{i-}) - \Delta^n_i X'(\tilde{\sigma}^2_{i+} - \tilde{\sigma}^2_{i-})\right),\]

\[V^n_t = \sum_{i=k_n+1}^{[t/\Delta_n]-k_n} \left(\Delta^n_i X'(\tilde{\sigma}^2_{i+} - \tilde{\sigma}^2_{i-}) - \Delta^n_i X'\Delta^n_i \sigma^2\right),\]

\[D(1)^n_t = \sum_{i=k_n+1}^{[t/\Delta_n]-k_n} \Delta^n_i X'\Delta^n_i \sigma^2_d,\]

\[D(2)^n_t = -\left(\sum_{i=1}^{k_n} + \sum_{i=[t/\Delta_n]-k_n+1}^{[t/\Delta_n]}\right) \Delta^n_i X'\Delta^n_i \sigma^2,\]

\[= \mathbb{E}_{\tau_0}[(\Delta^n X')(\Delta^n X')(\Delta^n X')^2(\sigma^2_{\tau_{i+1}})] \Delta_n = \mathbb{E}_{t-k_n}[(\Delta^n X')(\Delta^n X')(\sigma^4_{\tau_{i+1}-})] \Delta^2_n\]

\[= \mathbb{E}_{\tau_0}((\Delta^n X')(\sigma^4_{\tau_{j-1}-})(a'_{\tau_{j-1}+} + 4\sigma^2_{\tau_{j-1}-})) \Delta^3_n\]

\[= (\sigma^4_{\tau_0-})\left(((a'_{\tau_0})^2 + \Gamma^{(X,\sigma)}_{\tau_0} + 4a'_{\tau_0}\tilde{\sigma}_{\tau_0}) + 4\left(a'_{\tau_0}\tilde{\sigma}_{\tau_0} + \Gamma^{(X,\tilde{\sigma})}_{\tau_0} + 4\tilde{\sigma}^2_{\tau_0}\right)\right) \Delta^4_n,\]

where the volatility process includes joint jumps with price process, and
\[ D(3)_t^n = \sum_{i=1}^{[t/\Delta_n]} \Delta_i^n X' \Delta_i^n \sigma^2 e - \int_0^T 2\sigma_i^2 \tilde{\sigma}_i dt. \]

B.1.1 Proof of Theorem 5.1

We prove the central limit theory first. In words, we are going to show that the properly scaled truncation error and discretization error converges in probability to zero, while the scaled volatility estimation error converges stably in law to the limiting process.

**Step 1.** To prove \( \sqrt{u_n} T(\alpha_n)_t^n \xrightarrow{u.c.p.} 0 \), it suffices to show

\[ \limsup_{n \to \infty} \sqrt{u_n} \mathbb{E}(|T(\alpha_n)_t^n|) = 0. \]

Since

\[
\sqrt{u_n} |T(\alpha_n)_t^n| = \sqrt{u_n} \left| \sum_{i=k_n+1}^{[t/\Delta_n]-k_n} \left( \Delta_i^n X_{\alpha_n} (\tilde{\sigma}_{i+}^2 - \tilde{\sigma}_{i-}^2) - \Delta_i^n X' (\tilde{\sigma}_{i+}'^2 - \tilde{\sigma}_{i-}'^2) \right) \right|
\]

\[ \leq \sum_{i=k_n+1}^{[t/\Delta_n]-k_n} \sum_{j \in I_{n}^{(i)}} \sqrt{u_n} \Delta_n \left| \Delta_i^n X_{\alpha_n} (\Delta_j^n X_{\alpha_n})^2 - \Delta_i^n X' (\Delta_j^n X')^2 \right|. \]

It is enough to prove that \( \mathbb{E} \left[ \left| \Delta_i^n X_{\alpha_n} (\Delta_j^n X_{\alpha_n})^2 - \Delta_i^n X' (\Delta_j^n X')^2 \right| \right] = \psi_n \Delta_n^2 / \sqrt{u_n}, \) where \( \psi_n \) converges to zero as \( n \) goes to infinity.

Consider the function \( F(x_i, x_j) = x_i(x_j)^2 \) and vectors

\[
\tilde{X}_{i,j}^n(1) = \frac{1}{\sqrt{\Delta_n}} (\Delta_i^n X, \Delta_j^n X),
\]

\[
\tilde{X}_{i,j}^n(2) = \frac{1}{\sqrt{\Delta_n}} (\Delta_i^n X, \Delta_j^n X'),
\]

\[
\tilde{X}_{i,j}^n(3) = \frac{1}{\sqrt{\Delta_n}} (\Delta_i^n X', \Delta_j^n X').
\]

Then denote \( \alpha_n / \sqrt{\Delta_n} \) by \( v_n \) and define

\[
\phi_{i,j}^n(1) = F_{v_n} (\tilde{X}_{i,j}^n(1)) - F_{v_n} (\tilde{X}_{i,j}^n(2)), \quad \phi_{i,j}^n(2) = F_{v_n} (\tilde{X}_{i,j}^n(2)) - F_{v_n} (\tilde{X}_{i,j}^n(3)),
\]

where \( F_{v}(x_1, x_2) = F(x_1, x_2) \prod_{i=1}^{2} 1_{\{||x_i|| \leq v\}}. \)

Now we have

\[
\Delta_i^n X_{\alpha_n} (\Delta_j^n X_{\alpha_n})^2 - \Delta_i^n X' (\Delta_j^n X')^2 = (\phi_{i,j}^n(1) + \phi_{i,j}^n(2)) \Delta_n^3/2.
\]
A look at the proof of Lemma 13.2.6 in in Jacod and Protter (2011) shows it does not make essential difference to use vectors of non-adjacent increments. Therefore, applying that lemma with $r < 1, m = s = 1, s' = p' = 2$ we can get for $l = 1, 2$

$$
\mathbb{E}_{r_{l-1}}[\|\phi^\alpha_{l,j}(l)\|] \leq \left( \frac{2-r}{2} + \Delta_n^{(2-r)\sigma} \right) \psi_n = \psi_n \Delta_n^{(2-r)\sigma} \left( 1 + \Delta_n^{(2-r)(1/2)\sigma} \right).
$$

Therefore, it is sufficient to have $(2 - r)\sigma \geq 3/4$, which amounts to $\sigma \geq \frac{3}{4(2-r)}$.

**Step 2.** We are going to show that $\sqrt{u_n} D(1)^n_t$ is asymptotically negligible.

$$
\sqrt{u_n} \mathbb{E}[|D(1)^n_t|] \leq \sqrt{u_n} \mathbb{E} \left[ \sum_{i=k_n+1}^{[t/\Delta_n]-k_n} |\Delta^n_i X' \Delta^d_i \sigma^2| \right].
$$

Conditioned on the filtration generated by volatility process, $\Delta^n_i X'$ behaves like a mixture of normal and we have

$$
\mathbb{E}[|\Delta^n_i X'||\mathcal{F}_\sigma] = O_p(\sqrt{\Delta_n}) \leq K \sqrt{\Delta_n},
$$

since the volatility process is bounded by Assumption (SH). Consequently, we obtain

$$
\sqrt{u_n} \mathbb{E} \left[ \sum_{i=k_n+1}^{[t/\Delta_n]-k_n} |\Delta^n_i X' \Delta^d_i \sigma^2| \right] \leq \sqrt{u_n} \Delta_n K \sum_{i=k_n+1}^{[t/\Delta_n]-k_n} \mathbb{E}[|\Delta^d_i \sigma^2|].
$$

Then as long as $\Delta^d_i \sigma^2$ is of finite variation, we have

$$
\limsup_{n \to \infty} \sqrt{u_n} \mathbb{E}[|D(1)^n_t|] \leq K \sqrt{u_n \Delta_n} \to 0,
$$

implying $D(1)^n_t$ is asymptotically negligible.

**Step 3.** In this step, we are going to prove the following result for $j = 2$ and 3

$$
\sqrt{u_n} D(j)^n_t \overset{u.c.p.}{\to} 0.
$$

By Cauchy-Schwarz inequality, we obtain

$$
\mathbb{E}[|\Delta^n_i X' \Delta^d_i \sigma^2|] \leq \sqrt{\mathbb{E}[(\Delta^d_i X')^2] \mathbb{E}[\Delta^d_i \sigma^2]^2} \leq K \Delta_n.
$$

Then $\sqrt{u_n} D(2)^n_t \overset{u.c.p.}{\to} 0$ readily follows from the following fact that

$$
\limsup_{n \to \infty} \sqrt{u_n} \mathbb{E}[|D(2)^n_t|] \leq \limsup_{n \to \infty} K \sqrt{u_n k_n \Delta_n} = 0.
$$
On the other hand, by Itô formula, we have
\[ (X^r_{t+s} - X^r_t)(\sigma^2_{t+s} - \sigma^2_t) = \int_0^s 2\sigma^2_{t+r} \bar{\sigma}_t dr + \int_0^s (X^r_{t+r} - X^r_t) d\sigma^2_{t+r} + \int_0^s (\sigma^2_{t+r} - \sigma^2_t) dX^r_t. \]
Consequently, we get
\[
\mathbb{E}_{t-1}[\Delta_n^n X' \Delta_n^n \sigma^2, - \int_0^{t_n} 2 \sigma^2_t \bar{\sigma}_t dt] = K \Delta_n^{3/2}, \\
\mathbb{E}_{t-1}[\Delta_n^n X' \Delta_n^n \sigma^2, - \int_0^{t_n} 2 \sigma^2_t \bar{\sigma}_t dt]^2 \leq K \Delta_n^2 .
\]
Denote \( \zeta_n = \sqrt{u_n}(\Delta_n^n X' \Delta_n^n \sigma^2, - \int_0^{t_n} 2 \sigma^2_t \bar{\sigma}_t dt) \). Note \( \zeta_n \) is \( \mathcal{F}_{t_n} \)-measurable. Then the above equations yield
\[
\sum_{i=1}^{[t/\Delta_n]} \mathbb{E}_{i-1}[\zeta_n^i] = K \sqrt{u_n \Delta_n} \rightarrow 0, \\
\sum_{i=1}^{[t/\Delta_n]} \mathbb{E}_{i-1}[\zeta_n^i]^2 \leq K \ u_n \Delta_n \rightarrow 0
\]
Hence, Lemma A.1 implies that \( \sqrt{u_n} D(3)_t \overset{u.c.}{\rightarrow} 0 \).

**Step 4.** In this step we analyze \( \sqrt{u_n} V^n_t \). Define
\[
\xi_n = \sqrt{u_n} \left( \Delta_n^n X' (\bar{\sigma}_i^2 - \bar{\sigma}_i^2) - \Delta_n^n X' \Delta_i^n \sigma^2 \right). \quad (B.2)
\]
Although the variable \( \xi_n \) has a vanishing \( \mathcal{F}_{(t-1)\Delta_n} \)-conditional expectation, it is not \( \mathcal{F}_{t\Delta_n} \)-measurable. In order to ensure some “conditional independence” of the successive summands, we split the sum over \( i \) into big blocks of size \( mk_n \) (\( m \) will eventually go to infinity), separated by small blocks of size \( 2k_n \). The condition on \( m \) is
\[
m \rightarrow \infty \text{ and } mk_n \Delta_n \rightarrow 0. \quad (B.3)
\]
More specifically, define \( I(m, n, l) = (l-1)(m+2)k_n + 1 \), then the \( l \)-th big block contains \( \xi_n \) for all \( i \) between \( I(m, n, l) + k_n + 1 \) and \( I(m, n, l) + (m+1)k_n \). Then the total number of such blocks is \( l_n(m, t) = \lfloor \frac{t}{(m+2)k_n} \rfloor \). Let
\[
\xi_n = \sum_{r=k_n+1}^{(m+1)k_n} \xi_{I(m, n, i) + r}, \quad Z_n = \sum_{i=1}^{l_n(m, t)} \xi_n, \quad \tilde{\xi}_n = \sum_{r=-k_n}^{k_n} \xi_{I(m, n, i) + r}, \quad \tilde{Z}_n = \sum_{i=2}^{l_n(m, t)} \tilde{\xi}_n.
\]
So \( \sqrt{u_n} V_t^n = Z(m)_t^n + \tilde{Z}(m)_t^n \).

We are going to prove that \( \tilde{Z}(m)_t^n \) is asymptotically negligible first. By successively conditioning, we get

\[
\mathbb{E}_{I(m,n,i)-k_n-1}[\xi_{I(m,n,i)+r}] = \mathbb{E}_{I(m,n,i)-k_n-1}\left[ \Delta^n_t X' \sqrt{u_n} \left( (\hat{\sigma}^2_{i+} - \sigma^2_{i}) - (\hat{\sigma}^2_{i-} - \sigma^2_{i-1}) \right) \right] \\
= \mathbb{E}_{I(m,n,i)-k_n-1}\left[ \Delta^n_t X' O_p(\phi_n) + O_p(\Delta_n) \sqrt{u_n}(\hat{\sigma}^2_{i-} - \sigma^2_{i-1}) \right] \\
= \Delta_n O_p(\psi_n),
\]

\[
\mathbb{E}_{I(m,n,i)-k_n-1}[\xi_{I(m,n,i)+r}^2] = \mathbb{E}_{I(m,n,i)-k_n-1}\left[ (\Delta^n_t X')^2 u_n \left( (\hat{\sigma}^2_{i+} - \sigma^2_{i}) - (\hat{\sigma}^2_{i-} - \sigma^2_{i-1}) \right)^2 \right] \\
= \mathbb{E}_{I(m,n,i)-k_n-1}\left[ (\Delta^n_t X')^2 \mathbb{E}((V^+_i)^2 + (V^-_i)^2 + 2V_i^+V_i^- | \mathcal{F}) \right] \\
= \rho_{I(m,n,i)-k_n-1}\Delta_n.
\]

Note there is no overlap among the sequence \( \tilde{\xi}(m)_i^n \) and it is easy to check that

\[
\mathbb{E}_{I(m,n,i)-k_n-1}(\tilde{\xi}(m)_i^n) = O_p(k_n \Delta_n \psi_n),
\]

and

\[
\mathbb{E}_{I(m,n,i)-k_n-1}(\tilde{\xi}(m)_i^n)^2 = \mathbb{E}_{I(m,n,i)-k_n-1}\left( \sum_{r=-k_n}^{k_n} (\xi_{I(m,n,i)+r})^2 \right) \\
+ \mathbb{E}_{I(m,n,i)-k_n-1}\left( \sum_{r,j=-k_n}^{k_n} 1_{(j \neq r)} \xi_{I(m,n,i)+r} \xi_{I(m,n,i)+j} \right) \\
\leq O_p(k_n \Delta_n) + \sum_{r,j=-k_n}^{k_n} \mathbb{E}_{I(m,n,i)-k_n-1}[\Delta^n_r X' \Delta^n_j X' O_p(1)] \\
= O_p(k_n \Delta_n) + O_p(k_n^2 \Delta^2_n).
\]

Then by condition B.3, Lemma A.1 yields that \( \tilde{Z}(m)_t^n \xrightarrow{a.s.} 0 \).

Now the variable \( \xi(m)_i^n \) has vanishing \( \mathcal{F}_{I(m,n,i)} \)-conditional expectation, and \( \mathcal{F}_{I(m,n,i+1)} \)-measurable, implying that it behaves like a martingale difference.

We are going to prove

\[
\begin{align*}
\sum_{i=1}^{l_{m,t}} \mathbb{E} \left[ (\xi(m)_i^n)^2 \middle| \mathcal{F}_{I(m,n,i)} \right] & \xrightarrow{p} \int_0^t \rho_s^2 ds, \\
\sum_{i=1}^{l_{m,t}} \mathbb{E} \left[ (\xi(m)_i^n)^4 \middle| \mathcal{F}_{I(m,n,i)} \right] & \xrightarrow{p} 0, \\
\sum_{i=1}^{l_{m,t}} \mathbb{E} \left[ \xi(m)_i^n \Delta_{i,m}^n M \middle| \mathcal{F}_{I(m,n,i)} \right] & \xrightarrow{p} 0,
\end{align*}
\]

(B.4)
where \( \Delta_{i,n}^n M = M_{(I(m,n,i)+(m+1)k_n)} - M_{(I(m,n,i)+k_n+1)} \).

The difficulty left now is that \( \xi_i^n \) may be overlapped within the big block. To deal with this, let \( I_n^i(i) = \{i - k_n - 1, \ldots, i - 1\} \) if \( i > k_n \) and \( I_n^i(i) = \{i + 1, \ldots, i + k_n + 1\} \) define two local windows in time of length \( k_n \Delta_n \) just before and after time \( i \Delta_n \), and \( I_n^i(i) \) be the union of them. Furthermore, let

\[
J(m, n, i, j) = \{I(n, m, i) + k_n + 1, \ldots, I(m, n, i) + (m + 1)k_n\} \setminus (I_n^i(j) \cup j).
\]

With these notations, we can decompose the conditional second moment of \( \xi(m)_i^n \) as follows:

\[
\mathbb{E}\left[ (\xi(m)_i^n)^2 \bigg| \mathcal{F}_I(m,n,i) \right] = \sum_{r=k_n+1}^{(m+1)k_n} \sum_{j=k_n+1}^{(m+1)k_n} \mathbb{E}\left[ \xi_I(m,n,i)+r \xi_I(m,n,i)+j \bigg| \mathcal{F}_I(m,n,i) \right] =: H(m,1)_i^n + H(m,2)_i^n + H(m,3)_i^n.
\]

From Lemma A.5 and A.6, we have

\[
\sum_{i=1}^{l_n(m,t)} H(m,1)_i^n = \sum_{i=1}^{l_n(m,t)} \sum_{r=k_n+1}^{(m+1)k_n} \mathbb{E}\left[ (\xi_I(m,n,i)+r)^2 \bigg| \mathcal{F}_I(m,n,i) \right] \to \int_0^t \rho_s^2 ds,
\]

\[
\sum_{i=1}^{l_n(m,t)} H(m,2)_i^n = \sum_{i=1}^{l_n(m,t)} \sum_{r=k_n+1}^{(m+1)k_n} \sum_{j \in I_n(r)} \mathbb{E}\left[ \xi_I(m,n,i)+r \xi_I(m,n,i)+j \bigg| \mathcal{F}_I(m,n,i) \right] \leq \sum_{i=1}^{l_n(m,t)} Km^2 k_n^2 \Delta_n^2 \to 0,
\]

\[
\sum_{i=1}^{l_n(m,t)} H(m,3)_i^n = \sum_{i=1}^{l_n(m,t)} \sum_{r=k_n+1}^{(m+1)k_n} \sum_{j \in J(m,n,i,r)} \mathbb{E}\left[ \xi_I(m,n,i)+r \xi_I(m,n,i)+j \bigg| \mathcal{F}_I(m,n,i) \right] \leq \sum_{i=1}^{l_n(m,t)} Km^2 k_n^2 \Delta_n^2 \psi_n \to 0.
\]

The calculation of the fourth moments is even more tedious, so we present partial results and omit the rest calculation.

\[
\sum_{i=1}^{l_n(m,t)} \sum_{r=k_n+1}^{(m+1)k_n} \mathbb{E}\left[ (\xi_I(m,n,i)+r)^4 \bigg| \mathcal{F}_I(m,n,i) \right]
\]
As for the last equation in (B.4), it holds when \( M \) is independent with \( X \). Besides, it is not hard to verify that it is also true when \( M = X \). In any other case, \( M \) could be decomposed as the sum of two components, one is a functional of \( X \) and the other is independent with \( X \). So the result readily follows.

**B.1.2 Proof of Theorem 4.1**

From the previous result, we have

\[
\begin{align*}
\{V_l^n t &= O_p(1/\sqrt{u_n}), \\
D(j)_t^n &= o_p(1/\sqrt{u_n}),
\end{align*}
\]

implying that they all converges to zero in probability as \( n \) goes to infinity.

In order to prove \( T(\alpha_n)_t^n \overset{P}{\longrightarrow} 0 \), it is enough to prove that

\[
\mathbb{E}\left[ |\Delta^n X(\Delta^n X)^2 - \Delta^n X'(\Delta^n X')^2| \right] = \psi_n \Delta^n_n^2,
\]

instead of \( \psi_n \Delta^n_n^2/\sqrt{u_n} \). Consequently, it is sufficient to have \( (2 - r) \varpi \geq 2 \), which amounts to \( \varpi \geq \frac{1}{2(2-r)} \).

**B.2 Discontinuous leverage effect estimator**

We refer to Jacod and Todorov (2010) for the proof when \( u_n \propto k_n \), or \( b < 1/2 \). The proof when \( b \geq 1/2 \) is essentially the same since the limiting process is determined by spot volatility estimation too.
B.3 Proof of Theorem 7.1

For this proof, we will refer to Corollary 16.5.2 in Jacod and Protter (2011). This proves that the estimation of spot volatility will need to be adjusted by a factor $\frac{3}{2}$. The rest of the proof can be easily carried out as in the proof of Theorem 4.1.

References


