Essays in Financial Econometrics

Dacheng Xiu

A Dissertation

Presented to the Faculty

of Princeton University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance

by the Program in

Applied and Computational Mathematics

Adviser: Yacine Aït-Sahalia

May 2011

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

The huge amount of tick-by-tick data provides rich and timely information regarding fluctuations of traded assets and their co-movements. Nevertheless, the existence of market microstructure noise interferes with estimation, especially when the sampling frequency increases to beyond every five minutes. The first chapter introduces the Quasi-Maximum Likelihood Estimator of volatility as a solution to this problem. In theory, the proposed approach is consistent, rate-efficient, and shares the model-free feature with non-parametric alternatives. In practice, it is also convenient and has better small sample performance without any tuning parameters.

When measuring covariance and correlation, the fact that the two assets may not trade or otherwise be observed at exactly the same time, known as observation asynchronycity, is another issue that may distort the estimates. The second chapter, jointly written with Yacine Aït-Sahalia and Jianqing Fan, extends the Quasi-Maximum Likelihood Estimator to explore asynchronous and noisy data, with the help of generalized synchronization scheme.

A fundamental problem in option pricing is to find explicit pricing formulae or efficient pricing algorithms. However, closed-form pricing formulae are very sparse, whereas numerical or simulation-based methods are computationally expensive and deliver no insight into the structure of the option price. The third chapter fills in the gap between closed-form solutions and numerical methods through expansions of option prices. This approach works with general dynamics without any requirement on affine dynamics or explicit characteristic functions, and quantitatively characterizes the relative importance of model parameters as the option approaches expiration. With closed-form expansions, we translate model features into option prices, such as stochastic interest rate, mean-reverting drift, and self-exciting or skewed jumps. Numerical examples illustrate the accuracy of this approach.

Acknowledgements

This dissertation would not have been possible without support from many people. First of all, I am extremely indebted to my adviser, Yacine Aït-Sahalia, for your generous help, inspiring comments, and encouraging advice. I can still remember how you accepted me when I asked you to be my adviser, how you inspired me when I showed you the rudimentary draft of the first chapter, and how you encouraged me when I told you my decision for graduation. It was you who helped build up my confidence and courage to work independently. I would like to show my deep gratitude to the rest of my committee members, Jianqing Fan, Wei Xiong, and René Carmona, in addition to my adviser, for meeting with me frequently to discuss how to come up with an idea, how to draft and revise a paper, how to make presentations, and how to be a referee. I feel so fortunate and proud to have had the opportunity to work with you all. I also benefited tremendously from conversations with Markus Brunnermeier, Patrick Cheridito, Gregory Chow, Erhan Çinlar, Bo E. Honoré, Jean Jacod, Jakub Jurek, Jia Li, Yingying Li, Ulrich Müller, Per Mykland, José Scheinkman, Ronnie Sircar, and Jialin Yu, who helped me improve these chapters. I am also grateful to Robert Calderbank, Ingrid Daubechies, and Weinan E for bringing me to this program, and Howard Bergman, Bobray Bordelon, Kathleen DeGennaro, Ellen DiPippo, Phyllis Fafalios, Laura Hedden, Todd Hines, Christina Lipsky, Valerie Marino, Karen Neukirchen, Jessica O'Leary, and Carol Smith, for your patience and help. I also enjoyed exchanging ideas with friends and students from Program in Applied and Computational Math, Department of Economics, and Department of Operations Research and Financial Engineering. Last but not least, my special thanks go to my beloved Yi Li, for your love, support, and faith in me. No one needs you more than I need you.

To my parents.

CONTENTS

CHAPTER 1

QUASI-MAXIMUM LIKELIHOOD ESTIMATION OF VOLATILITY WITH HIGH-FREQUENCY DATA

This chapter investigates the properties of the well-known maximum likelihood estimator in the presence of stochastic volatility and market microstructure noise, by extending the classic asymptotic results of quasi-maximum likelihood estimation. When trying to estimate the integrated volatility and the variance of noise, this parametric approach remains consistent, efficient and robust as a quasi-estimator under misspecified assumptions. Moreover, it shares the model-free feature with nonparametric alternatives, for instance realized kernels, while being advantageous over them in terms of finite sample performance. Comparisons with a variety of implementations of the Tukey-Hanning$_2$ kernel are provided using Monte Carlo simulations, and an empirical study with the Euro/US Dollar future illustrates its application in practice. This chapter is published in Xiu (2010), and presented at the workshop on Financial Econometrics at Fields Institute for Research in the Mathematical Sciences, the 2nd Princeton-Humboldt Finance workshop, the conference in Modeling High Frequency Data in Finance at Stevens Institute, the 2010 IMS Annual Meeting, and the 2010 Joint Statistical Meetings.

1.1 Introduction

The availability of high frequency data is a double-edged sword for the estimation of volatility. On the one hand, it facilitates our empirical studies of the asymptotic properties of a natural estimator, Realized Variance (RV), i.e. the sum of squared log-returns; while on the other hand, along with the data comes market microstructure noise, which disrupts all the desirable properties of the estimator. Without microstructure noise, this simple estimator is both consistent and efficient. In addition, other studies have shed light on its central limit theory, such as Jacod (1994) and Barndorff-Nielsen and Shephard (2002). However, the existence of noise interferes with the estimation especially when the sampling frequency approaches zero. A common practice is to sample sparsely, say every 5 minutes, discarding a large portion of the sample and the information therein.

The problem has received considerable attention recently. For instance, Aït-Sahalia et al. (2005) suggest sampling as frequently as possible at the cost of modeling the noise. This paper assumes constant volatility so that it can perform the maximum likelihood estimation (MLE). In the setting of stochastic volatility, Zhang et al. (2005) bring forward a nonparametric estimator, Two-Scale Realized Volatility (TSRV), which is the first consistent estimator in the presence of noise, despite a relatively low convergence rate $n^{\frac{1}{6}}$. Subsequently, Zhang (2006) advocates Multi-Scale Realized Volatility (MSRV), improving the convergence rate to $n^{\frac{1}{4}}$, which is the optimal rate a model can reach as shown by Gloter and Jacod (2000). Recently, Barndorff-Nielsen et al. (2008a) have designed various Realized Kernels (RKs) that can be used to deal with endogenous noise and endogenously spaced data and their convergence rates are the same as that of the MSRV. In simulation, these nonparametric estimators perform very well with optimally selected bandwidth or kernels; while in practice, estimators with bandwidth based on ad hoc choices, or based on small sample performance may behave better, as illustrated in Gatheral and Oomen (2009) and Bandi and Russell (2008). Another group of estimators dates back to Zhou (1996), who first applies the autocovariance-based-approach to constant volatility cases. Hansen and Lunde (2006) extend the Zhou estimator to the stochastic volatility models with serially dependent noise. However, the Zhou estimator and its extensions are inconsistent.

The motivation and inspiration of this article stem from Gatheral and Oomen (2009) and Aït-Sahalia and Yu (2009). Using artificially simulated "zero-intelligence" data, Gatheral and Oomen (2009) compare a comprehensive set of estimators including those listed above and their ad hoc modifications. According to their studies, the MLE, though possibly misspecified with time varying volatility, is among the best in terms of efficiency and robustness. In addition, Aït-Sahalia and Yu (2009) apply the MLE to analyze the liquidity of NYSE stocks. Their maximum likelihood estimates with the data simulated from stochastic volatility models indicate good properties, although this estimator is derived from a constant volatility assumption. Related works also include Hansen et al. (2008), where the authors suggest the consistency of the MLE by examining moving average filters. These studies motivate us to consider the MLE as a Quasi-Maximum Likelihood Estimator (QMLE)[1] under misspecified models, which dates back to as early as Amemiya (1973), White (1980) and White (1982). In these articles, the framework of misspecified estimation has been built and its close connection with Kullback-Leibler Information Criterion (KLIC) has been illustrated. Since these seminal works, Domowitz and White (1982) and Bates and White (1985) have extended the consistency results of the QMLE with i.i.d. models to various cases including dependent observations, Quasi-GMM-estimators, and Quasi-M-estimators.

Our work is thereby built on the fusion of high frequency data and misspecified likelihood estimation. The correct model specification features stochastic volatility. However, this model is intentionally misspecified to be one of constant volatility. Under this assumption, we perform (quasi) maximum likelihood estimation and analyze the estimator, which is essentially the same as the MLE in Aït-Sahalia et al. (2005). Remarkably, in the context of the correct model, the QMLE of volatility consistently estimates the integrated volatility at the most efficient rate $n^{\frac{1}{4}}$. Also, the QMLE of noise variance has the same asymptotic distribution as before. In other words, the maximum likelihood estimators are robust to stochastic volatility. In addition, they are still robust to random sampling intervals and non-Gaussian market microstructure noises.

[1]We give the MLE an alias QMLE sometimes in order to emphasize model misspecification and keep the notation in line with the classic results of misspecified models.

The chapter is organized as follows. Section 1.2 reviews the parametric likelihood estimator and provides the intuition and motivation for the QMLE. Section 1.3 outlines the classic asymptotic theory of the QMLE, and derives an extension to more general settings. Section 1.4 investigates the statistical properties of the QMLE, where the consistency, central limit theory and robustness of the QMLE are established. Section 1.5 compares it with nonparametric kernel estimators, and Section 1.6 uses Monte Carlo simulations to verify the conclusions obtained from the previous sections. Section 1.7 details an empirical study with the Euro/US Dollar future data. Section 1.8 concludes. The appendix provides all mathematical proofs.

1.2 Revisiting the MLE: the QMLE

In this section, we recapitulate the parametric approach and introduce our quasi-estimator. The traditional parametric methodology applies to cases where the true value of the parameter of interest is a special point in the parameter space. Therefore, Aït-Sahalia et al. (2005) have to make an assumption that the volatility is neither a stochastic process nor a deterministic function, but instead, a constant. That is, the latent efficient log price process satisfies

$$dX_t = \sigma dW_t \tag{1.1}$$

with the observed log transaction price $\tilde{X}_{\tau_i}$ contaminated by the microstructure noise U in a way such that $\tilde{X}_{\tau_i} = X_{\tau_i} + U_{\tau_i}$, where τ_i is the observation time. For simplicity, we assume that the data are regularly spaced, satisfying $\tau_i - \tau_{i-1} = \Delta$. The observations are made within $[0, T]$, where $T = n \cdot \Delta$ is fixed, so that the infill asymptotic behaviors are determined as n goes to ∞ and Δ goes to 0 simultaneously. The structure of the observed log return $Y_i s$ features MA(1), where

$$Y_i = \tilde{X}_{\tau_i} - \tilde{X}_{\tau_{i-1}} = \sigma(W_{\tau_i} - W_{\tau_{i-1}}) + U_{\tau_i} - U_{\tau_{i-1}}$$

If we postulate that the noise distribution is Gaussian, then our log likelihood function for $Y s$ is

$$l(\sigma^2, a^2) = -\frac{1}{2}\log\det(\Omega) - \frac{n}{2}\log(2\pi) - \frac{1}{2}Y'\Omega^{-1}Y \tag{1.2}$$

where

$$\Omega = \begin{pmatrix} \sigma^2\Delta + 2a^2 & -a^2 & 0 & \cdots & 0 \\ -a^2 & \sigma^2\Delta + 2a^2 & -a^2 & \ddots & \vdots \\ 0 & -a^2 & \sigma^2\Delta + 2a^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a^2 \\ 0 & \cdots & 0 & -a^2 & \sigma^2\Delta + 2a^2 \end{pmatrix} \tag{1.3}$$

The MLE $(\hat{\sigma}^2, \hat{a}^2)$ proves to be consistent at different rates for its volatility part and noise part even if the noise distribution turns out to be non-Gaussian:

$$\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}^2 - \sigma_0^2) \\ n^{\frac{1}{2}}(\hat{a}^2 - a_0^2) \end{pmatrix} \xrightarrow{\mathcal{L}} N\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 8a_0\sigma_0^3 T^{-\frac{1}{2}} & 0 \\ 0 & 2a_0^4 + \text{cum}_4[U] \end{pmatrix} \Big)$$

where $\text{cum}_4[U]$ is the fourth cumulant of the true noise U.

The optimal rate of convergence for volatility estimation turns to $n^{1/4}$ in the presence of noise. Given a typical 6.5 hours trading day, if we model the noise and sample as high as every 1 second, we have the normalization constant $n^{1/4}$ around 12.4, whereas sampling every 5 minutes and ignoring the noise lead to a normalization constant $n^{1/2}$ which approximately equals to 8.8. It appears that the gain in efficiency is marginal when sampling at highest frequency. However, if we further compare the asymptotic variances, $8a_0\sigma_0^3 T^{-\frac{1}{2}}$ is much smaller than $2\sigma_0^4$, since the standard deviation a_0 of the microstructure noise is extremely small in practice.

Implementing the MLE is more convenient than it appears. In fact, the likelihood function for the observed log-returns can be expressed in the following computationally efficient form, as a function of the transformed parameters (γ^2, η) by triangularizing the

matrix Ω (see, e.g. Aït-Sahalia et al. (2005)):

$$l(\eta, \gamma^2) = -\frac{1}{2}\sum_{i=1}^{N} \ln(2\pi d_i) - \frac{1}{2}\sum_{i=1}^{N} \frac{\tilde{Y}_i^2}{d_i}, \tag{1.4}$$

where

$$d_i = \gamma^2 \frac{1 + \eta^2 + ... + \eta^{2i}}{1 + \eta^2 + ... + \eta^{2(i-1)}}$$

and the $\tilde{Y}_i's$ are obtained recursively as $\tilde{Y}_1 = Y_1$ and for $i = 2, ..., N$:

$$\tilde{Y}_i = Y_i - \frac{\eta\left(1 + \eta^2 + ... + \eta^{2(i-2)}\right)}{1 + \eta^2 + ... + \eta^{2(i-1)}}\tilde{Y}_{i-1}.$$

This algorithm avoids the brute-force computation of Ω^{-1}, and hence significantly accelerates the optimization procedure.

Nevertheless, ample evidence of stochastic volatility calls this simplistic model into question. So, it is natural to ask: what is the impact of stochastic volatility on the MLE? Will stochastic volatility change the desired properties of the MLE, just as microstructure noise does to RV, or will the MLE still be consistent? Simulation studies by Aït-Sahalia and Yu (2009) and Hansen et al. (2008) seem to have suggested that the MLE may be a consistent estimator of the integrated volatility. Intuitively, this conjecture is plausible in that when volatility becomes stochastic, the integrated volatility, the parameter of interest, happens to be the average of the volatility process, which is expected to be a legitimate candidate for the estimate. If consistency is guaranteed, what would be the convergence rate and asymptotic variance? How would it compare with other alternative nonparametric estimators? Closer scrutiny of the estimator in the absence of microstructure noise may yield more insights.

Consider a stochastic volatility model with no noise:

$$dX_t = \sigma_t dW_t$$

The objective is to estimate the integrated volatility $\int_0^T \sigma_t^2 dt$. By design, we mistakenly

assume the spot volatility σ_t to be constant; therefore, the quasi-log likelihood function is

$$l(\omega, \sigma^2) = -\frac{n}{2}\log(\sigma^2\Delta) - \frac{n}{2}\log(2\pi) - \frac{1}{2\sigma^2\Delta}Y'Y$$

where $Y_i = X_{\tau_i} - X_{\tau_{i-1}} = \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t$ and $Y = (Y_1, Y_2, \ldots, Y_n)'$.

Apparently, the QMLE is

$$\hat{\sigma}_n^2 = \frac{1}{T}\sum_{i=1}^{n} Y_i^2 = \frac{1}{T}\sum_{i=1}^{n}(X_{\tau_i} - X_{\tau_{i-1}})^2$$

Here, the RV estimator recurs through a different argument, and it is of course the perfect estimator under the true stochastic volatility model.

However, the consistency of the QMLE is no longer straightforward in the presence of noise, because there may be no close form available for this estimator. Its asymptotic variance is far more complicated due to heteroskedasticity and autocorrelation, as mentioned by Hansen et al. (2008). The remaining paper rigorously investigates the asymptotic behavior of the QMLE in the setting of stochastic volatility and microstructure noise, which leads us to the classic asymptotic theory of quasi-estimators.

1.3 Asymptotic Theory of Quasi-M-Estimators

In this section, we at first briefly review the consistency of the QMLE under misspecified models, a theory initially developed in White (1980) and White (1982). The rationale behind the theory is related to Kullback - Leibler Information Criterion (KLIC). More precisely, suppose that we have an i.i.d. random sample, and let $g(X)$ be the true unknown data generating density, and $f(X, \theta)$ our possibly misspecified density indexed by a parameter $\theta \in \Theta$. White (1982) claims that under certain regular conditions, the QMLE is consistent to θ^* which minimizes KLIC:

$$I(g : f, \theta) = E(\log[g(X)/f(X, \theta)])$$

where the expectation is taken under the true model. If the model is correctly specified, that is, there exists $\theta_0 \in \Theta$, such that $g(X) = f(X, \theta_0)$, then the KLIC attains its minimum at $\theta^* = \theta_0$, hence this result is in agreement with the consistency of regular maximum likelihood estimators. Otherwise, the model is misspecified, and intuitively, θ^* minimizes our ignorance of the true structure.

In Domowitz and White (1982), the authors generalize these results to include dependent observations, and show that the QMLE $\hat{\theta}_n$, which maximizes $Q_n(\omega, \theta)$, still converges in probability to θ_n^*[2], a maximizer of $\bar{Q}_n(\omega, \theta)$, the expectation of $Q_n(\omega, \theta)$ under the true model. In addition, Bates and White (1985) extend the theory to general quasi-M-estimators, by introducing discrepancy functions.

Now, we add more randomness to the misspecification theory, which enables our applications to stochastic volatility models, where the parameter of interest itself is random. The reasoning of the proof is similar to the regular one given in White (1980) and Newey and McFadden (1994).

Theorem 1. *Let $Q_n(\omega, \theta)$ and $\bar{Q}_n(\omega, \theta)$ be two random functions such that for each θ in Θ, a compact subset of R^k, they are measurable functions on Ω and, for each $\omega \in \Omega$, continuous functions on Θ. In addition, $\bar{Q}_n(\omega, \theta)$ is almost surely maximized at $\theta_n^*(\omega)$. Further, the following two conditions are satisfied as $n \to \infty$:*

1. *Uniform Convergence:*

$$\sup_{\theta \in \Theta} \|Q_n(\omega, \theta) - \bar{Q}_n(\omega, \theta)\| \xrightarrow{\mathrm{P}} 0. \tag{1.5}$$

2. *Identifiability: for every $\epsilon > 0$, there exists a constant $\delta_0 > 0$, such that*

$$P(\bar{Q}_n(\omega, \theta_n^*) - \max_{\theta \in \Theta : \|\theta - \theta_n^*\| \geq \epsilon} \bar{Q}_n(\omega, \theta) > \delta_0) \to 1. \tag{1.6}$$

Then any sequence of estimators $\hat{\theta}_n$ such that $Q_n(\omega, \hat{\theta}_n) \geq \sup_{\theta \in \Theta} Q_n(\omega, \theta) + o_p(1)$ converges in probability to θ_n^*, i.e., $\hat{\theta}_n - \theta_n^* \xrightarrow{\mathrm{P}} 0$.

Note that the identifiability condition here is slightly different from that in a common setting, for example Newey and McFadden (1994). It not only requires some uniqueness

[2] θ^* might depend on n when model is misspecified.

like property of the maximizer, but also a proper normalization such that the maximizer can be distinguished asymptotically.

Now we modify the assumptions to accommodate to the M-Estimators setting, in case we need different normalizations for different parameters.

Theorem 2. *Let $\Psi_n(\omega,\theta)$and $\bar{\Psi}_n(\omega,\theta)$ be random vector-valued functions. For each θ in Θ, a compact subset of $\mathbb{R}^k$, they are measurable function on Ω, and for each ω in Ω, continuous functions on Θ. In addition, there exists a sequence of θ_n^*, satisfying $\bar{\Psi}_n(\omega,\theta_n^*)=0$ almost surely, such that as $n\to\infty$,*

1. *Uniform Convergence:*

$$\sup_{\theta\in\Theta}\|\Psi_n(\omega,\theta)-\bar{\Psi}_n(\omega,\theta)\| \xrightarrow{\mathrm{P}} 0. \tag{1.7}$$

2. *Identifiability: For every $\epsilon>0$, there exists a constant $\delta_0>0$, such that,*

$$P(\min_{\theta\in\Theta:\|\theta-\theta_n^*\|\geq\epsilon}\|\bar{\Psi}_n(\omega,\theta)\|>\delta_0)\to 1. \tag{1.8}$$

Then any sequence of estimators $\hat{\theta}_n$ such that $\Psi_n(\omega,\hat{\theta}_n)=o_p(1)$ converges in probability to θ_n^*, i.e., $\hat{\theta}_n-\theta_n^*\xrightarrow{\mathrm{P}}0$.

In the following discussions, we will choose Ψ_n as the score function of a misspecified model, up to an appropriate normalization, and $\bar{\Psi}_n$ is carefully specified corresponding to Ψ_n.

The central limit result is given by the next theorem, which is an extension of Theorem 2.4 in Domowitz and White (1982).

Theorem 3. *Suppose that the conditions of Theorem 2 are satisfied. In addition, $\Psi_n(\omega,\theta)$ and $\bar{\Psi}_n(\omega,\theta)$ are continuously differentiable of order 1 on Θ. Also, there exists a sequence of positive definite matrices $\{V_n(\omega)\}$ such that*

$$-V_n(\omega)\Psi_n(\omega,\theta_n^*)\xrightarrow{\mathcal{L}}N(0,I_k) \tag{1.9}$$

If $\nabla\bar{\Psi}_n(\omega,\theta)$ is stochastic equicontinuous, and $|\nabla\Psi_n(\omega,\theta)-\nabla\bar{\Psi}_n(\omega,\theta)|\xrightarrow{\mathrm{P}}0$, uniformly for all

$\theta \in \Theta$, *then*

$$V_n(\omega)\nabla\bar{\Psi}_n(\omega,\theta_n^*)(\hat{\theta}_n-\theta_n^*) \xrightarrow{\mathcal{L}} N(0,I_k) \tag{1.10}$$

The extensions of the classic asymptotic theory pave the way for a thorough inquiry of asymptotic properties of the QMLE.

1.4 Statistical Properties of the QMLE

This section shares the setup with most volatility estimators available in the literature. More specifically, we make the following assumptions.

Assumption 4. *The underlying latent log price process satisfies*

$$dX_t = \sigma_t dW_t$$

with the volatility process a positive and locally bounded Itô semimartingale.[3]

Assumption 5. *The noise U_t is independently and identically distributed, and independent of price and volatility processes, with mean 0, variance a_0^2 and finite fourth moment.*

The assumptions of i.i.d. noise and its independence with prices are not always consistent with the empirical evidence, as pointed out in Hansen and Lunde (2006). Aït-Sahalia et al. (2005) have discussed the way to modify the log likelihood function with more parameters, according to the assumed parametric structure of the noise. See also Gatheral and Oomen (2009) for an MA(2) implementation. As to the TSRV, Aït-Sahalia et al. (2009) have extended it to the case with serially dependent noise. Kalnina and Linton (2008) have proposed a modification of the TSRV with endogenous noise and heteroskedastic measurement error. Independently, Barndorff-Nielsen et al. (2008a) have shown the robustness of their realized kernels with respect to endogenous noise. However, serially dependence plus endogenous assumption are still unsatisfactory, since in reality the transaction prices

[3]This assumption accommodates virtually all continuous time financial models. See Jacod (2008) Hypothesis (L-s) for more details.

are recorded with round-off errors. Li and Mykland (2007) have discussed the robustness of the TSRV with respect to rounding errors while Jacod et al. (2009a) have recently proposed a pre-averaging approach that works well for a general class of errors including certain combination of rounding and additive errors. In this paper, being parsimonious and for simplicity, we consider the i.i.d. white noise and provide in addition a heuristic argument for time-dependent noise to validate the applications of the QMLE in practice.

1.4.1 Consistency of the QMLE

Consider first what would happen in the absence of noise. Based on the discussion in Section 1.2, the score function (up to a proper normalization) under misspecified model is,

$$\Psi_n(\omega,\sigma^2) = -\frac{1}{n}\frac{dl(\omega,\sigma^2)}{d\sigma^2} = \frac{1}{2}\{\frac{1}{\sigma^2} - \frac{1}{n\sigma^4\Delta}Y'Y\}$$

and its root is

$$\hat{\sigma}_n^2 = \frac{1}{T}\sum_{i=1}^{n} Y_i^2 = \frac{1}{T}\sum_{i=1}^{n}(X_i - X_{i-1})^2$$

Then we choose

$$\bar{\Psi}_n(\omega,\sigma^2) = \frac{1}{2}\{\frac{1}{\sigma^2} - \frac{1}{n\sigma^4\Delta}\int_0^T \sigma_t^2 dt\} \tag{1.11}$$

Therefore, it has a root $\sigma_n^{2*} = \frac{1}{T}\int_0^T \sigma_t^2 dt$.

Because σ^2 is in a compact set Θ, and if we require the parameter space to be bounded away from zero, i.e., there exist $\tilde{\sigma}^2 \in \Theta$ such that $\sigma^2 \geq \tilde{\sigma}^2 > 0$, then

$$\sup_{\sigma^2\in\Theta} |\Psi_n(\omega,\sigma^2) - \bar{\Psi}_n(\omega,\sigma^2)| = \frac{1}{2\tilde{\sigma}^4 T}|\sum_{i=1}^{n} Y_i^2 - \int_0^T \sigma_t^2 dt| \xrightarrow{\text{P}} 0 \tag{1.12}$$

Proof of the last step is well-known, see e.g. Karatzas and Shreve (1991, pp32 Theorem 5.8), hence uniform convergence in probability in Theorem 2 is shown. Identifiability condition trivially holds, so the consistency follows from Theorem 2. The rate of convergence

depends on the rate of (1.12), which is $n^{\frac{1}{2}}$, as given by Jacod (1994) for instance. Therefore,

$$\hat{\sigma}_n^2 - \sigma_n^{2*} = \frac{1}{T}\sum_{i=1}^{n} Y_i^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt = O_p(n^{-\frac{1}{2}})$$

Apparently, we do not need Theorem 2 at all here. In general, however, this is certainly not the case. One inspiration from this simple example is regarding the selection of $\bar{\Psi}_n$ in (1.11). If we add no leverage assumption, that is to say, the volatility process is conditionally deterministic, then $\bar{\Psi}_n$ is nothing but the conditional expectation of Ψ_n under the true model.

When the observed data are noisy, we have $Y_i = \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t + U_{\tau_i} - U_{\tau_{i-1}}$. Besides the constant volatility assumption, we mistakenly assume that the noises are normally distributed with variance a^2. Therefore, the quasi-log likelihood function is exactly (1.2), hence the same likelihood estimator recurs in the form of the QMLE under misspecified model. However, closed-form expressions are no longer available, so we have to turn to Theorem 2 for consistency.

Denote $\theta = (\sigma^2, a^2)$ and we also assume that the parameters stay in a compact set, which is bounded away from zero. As in the no noise case, we choose the following score functions up to some proper normalizations.

$$\begin{aligned}
\Psi_n &= (\Psi_n^1(\omega,\theta), \Psi_n^2(\omega,\theta))' \\
&= (-\frac{1}{\sqrt{n}}\frac{\partial l(a^2,\sigma^2)}{\partial \sigma^2}, -\frac{1}{n}\frac{\partial l(a^2,\sigma^2)}{\partial a^2})' \\
&= \Big(\frac{1}{2\sqrt{n}}\{\frac{\partial \log(\det\Omega)}{\partial \sigma^2} + Y'\frac{\partial \Omega^{-1}}{\partial \sigma^2}Y\}, \frac{1}{2n}\{\frac{\partial \log(\det\Omega)}{\partial a^2} + Y'\frac{\partial \Omega^{-1}}{\partial a^2}Y\}\Big)'
\end{aligned}$$

and correspondingly,

$$\begin{aligned}
\bar{\Psi}_n &= (\bar{\Psi}_n^1(\omega,\theta), \bar{\Psi}_n^2(\omega,\theta))' \\
&= \Big(\frac{1}{2\sqrt{n}}\{\frac{\partial \log(\det\Omega)}{\partial \sigma^2} + tr(\frac{\partial \Omega^{-1}}{\partial \sigma^2}\Sigma_0)\}, \frac{1}{2n}\{\frac{\partial \log(\det\Omega)}{\partial a^2} + tr(\frac{\partial \Omega^{-1}}{\partial a^2}\Sigma_0)\}\Big)'
\end{aligned}$$

where Σ_0 is given by

$$\Sigma_0 = \begin{pmatrix} \int_0^{\tau_1} \sigma_t^2 dt + 2a_0^2 & -a_0^2 & 0 & \cdots & 0 \\ -a_0^2 & \int_{\tau_1}^{\tau_2} \sigma_t^2 dt + 2a_0^2 & -a_0^2 & \ddots & \vdots \\ 0 & -a_0^2 & \int_{\tau_2}^{\tau_3} \sigma_t^2 dt + 2a_0^2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -a_0^2 \\ 0 & \cdots & 0 & -a_0^2 & \int_{\tau_{n-1}}^{T} \sigma_t^2 dt + 2a_0^2 \end{pmatrix}$$

Being aware that the convergence rates may be different for the two parameters, as shown in Section 1.2, we choose different normalizations accordingly. These normalizations are essential to ensure the identifiability condition.

Denote $\Omega^{-1} = (\omega^{ij})$, and $\epsilon_j = U_{\tau_j} - U_{\tau_{j-1}}$. The difference between Ψ_n^1 and $\bar{\Psi}_n^1$, for instance is

$$\begin{aligned} & 2\sqrt{n}(\Psi_n^1 - \bar{\Psi}_n^1) \\ =& Y' \frac{\partial \Omega^{-1}}{\partial \sigma^2} Y - tr(\frac{\partial \Omega^{-1}}{\partial \sigma^2} \Sigma_0) \\ =& \sum_{i=1}^n \frac{\partial \omega^{ii}}{\partial \sigma^2} \underbrace{\{(\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\}}_{\text{martingale difference}} + \sum_{i=1}^n \sum_{j \neq i}^n \frac{\partial \omega^{ij}}{\partial \sigma^2} \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t \int_{\tau_{j-1}}^{\tau_j} \sigma_t dW_t \\ & + 2 \sum_{i=1}^n \sum_{j=1}^n \frac{\partial \omega^{ij}}{\partial \sigma^2} \epsilon_j \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t + \underbrace{\sum_{i=1}^n \sum_{j \neq i, i-1, i+1}^n \frac{\partial \omega^{ij}}{\partial \sigma^2} \epsilon_i \epsilon_j}_{\text{due to the noise}} \\ & \underbrace{+ \sum_{i=1}^n \frac{\partial \omega^{ii}}{\partial \sigma^2} (\epsilon_i^2 - 2a_0^2) + \sum_{i=1}^n \frac{\partial \omega^{i,i-1}}{\partial \sigma^2} (\epsilon_i \epsilon_{i-1} + a_0^2) + \sum_{i=1}^n \frac{\partial \omega^{i,i+1}}{\partial \sigma^2} (\epsilon_i \epsilon_{i+1} + a_0^2)}_{\text{due to the noise}} \end{aligned} \tag{1.13}$$

The first term is a linear combination of martingale differences, while the second term is the sum of cross products over different pieces of integrals. The third one is a mixture of the martingale part and the noise part. The rest are purely related to the noise. It is clear that the first two terms as a whole, the third one and the rest are pairwise uncorrelated given the proposed assumptions.

Now we proceed with variance calculations. The lesson learned from the failure of RV

indicates that the variance of the noise may dominate the others. So for the noise part, we accurately compute the variance of the derivatives with respect to different parameters.

Lemma 1. *Given Assumptions 4 and 5, we have*

$$\sum_{i=1}^{n}\omega^{ii}\{(\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t)^2-\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2 dt\}=O_p(1) \tag{1.14}$$

$$\sum_{i=1}^{n}\sum_{j\neq i}^{n}\omega^{ij}\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t\int_{\tau_{j-1}}^{\tau_j}\sigma_t dW_t=O_p(n^{\frac{1}{4}}) \tag{1.15}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\omega^{ij}\epsilon_j\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t=O_p(n^{\frac{1}{4}}) \tag{1.16}$$

Lemma 2. *As to the noise part, we have*

$$var(\epsilon'\frac{\partial\Omega}{\partial\sigma^2}\epsilon)=\frac{\sqrt{T}a_0^4 n^{\frac{1}{2}}}{16a^5\sigma^3}+o(n^{\frac{1}{2}}) \tag{1.17}$$

$$var(\epsilon'\frac{\partial\Omega}{\partial a^2}\epsilon)=\frac{n(2a_0^4+cum_4[U])}{a^8}+o(n) \tag{1.18}$$

By verifying the identifiability condition, solving the equations $\bar{\Psi}_n=0$ and applying the above lemmas, we can prove

Theorem 6. *Ψ_n and $\bar{\Psi}_n$ are as given, $\hat{\theta}_n=(\hat{\sigma}_n^2,\hat{a}_n^2)$ is as defined in Theorem 2. Given Assumptions 1 and 2, the QMLE$(\hat{\sigma}^2,\hat{a}^2)$ satisfies: $\hat{a}_n^2-a_0^2\xrightarrow{\mathrm{P}}0$ and $\hat{\sigma}_n^2-\frac{1}{T}\int_0^T\sigma_t^2dt\xrightarrow{\mathrm{P}}0$.*

As expected, the consistency of the volatility estimator still holds, even though the volatility process becomes stochastic.

1.4.2 Central Limit Theorem of the QMLE

Although we have shown the consistency of the QMLE, we have not explored whether the QMLE has the optimal convergence rates, neither have we shown anything about the magnitude of the asymptotic variance or possible asymptotic bias. To answer these questions, the following lemma and theorem provide the central limit theory.

Lemma 3. *Given Assumptions 4 and 5, we have*[4]

$$\begin{pmatrix} n^{\frac{1}{4}}(\Psi_n^1 - \bar{\Psi}_n^1) \\ n^{\frac{1}{2}}(\Psi_n^2 - \bar{\Psi}_n^2) \end{pmatrix} \xrightarrow{\mathcal{L}_X}$$
$$MN\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{4}(\frac{5\int_0^T \sigma_t^4 dt}{16a\sigma^7\sqrt{T}} + \frac{a_0^2\int_0^T \sigma_t^2 dt}{8\sigma^5 a^3\sqrt{T}} + \frac{a_0^4\sqrt{T}}{16a^5\sigma^3}) & 0 \\ 0 & \frac{2a_0^4 + cum_4[U]}{4a^8} \end{pmatrix} \Big)$$

Theorem 7. *Ψ_n and $\bar{\Psi}_n$ are as given, $\hat{\theta}_n = (\hat{\sigma}_n^2, \hat{a}_n^2)$ is as defined in Theorem 2. Given Assumptions 4 and 5, the QMLE $(\hat{\sigma}^2, \hat{a}^2)$ satisfies:*

$$\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt) \\ n^{\frac{1}{2}}(\hat{a}^2 - a_0^2) \end{pmatrix} \xrightarrow{\mathcal{L}_X}$$
$$MN\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{5a_0\int_0^T \sigma_t^4 dt}{T(\int_0^T \sigma_t^2 dt)^{\frac{1}{2}}} + \frac{3(\int_0^T \sigma_t^2 dt)^{\frac{3}{2}} a_0}{T^2} & 0 \\ 0 & 2a_0^4 + cum_4[U] \end{pmatrix} \Big)$$

In the constant volatility case, QMLE attains the optimal efficiency, as can be expected since the QMLE is constructed in the same way as the MLE. In general, QMLE achieves the optimal convergence rates. However, this QMLE method might not provide a straightforward estimator for the integrated quarticity, that is, $\int_0^T \sigma_t^4 dt$. Thus, as to the construction of confidence intervals, one can instead use the method given by Jacod et al. (2009a).

1.4.3 Robustness of the QMLE

Drift

What happens to our conclusions if the underlying X process has a drift? More precisely, suppose X_t to be of the Itô type,

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

[4]MN is a notation of mixed normal used in Barndorff-Nielsen et al. (2008a), and $\mathcal{L}_X$ means $\sigma(X)$-stable convergence in law.

where μ_t is a locally bounded and progressively measurable process. Instead of parameterizing the drift term and modifying our likelihood estimator accordingly, we completely ignore the presence of drift, or in other words, misspecify the model with drift 0. In this case, the estimator is unchanged, and we only need to replace $\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t$ with $\int_{\tau_{i-1}}^{\tau_i} \mu_t dt + \int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t$ in the proof. Loosely speaking, this will not alter our conclusions since in the high frequency setting, the drift is of order dt, which is mathematically negligible with respect to the diffusive component of order $(dt)^{\frac{1}{2}}$. An alternative argument given by Mykland and Zhang (2009) is to zero out the drift by changing probability measures.

Random Sampling Intervals

What if the observations are made randomly? If the sampling intervals between two consecutive observations are i.i.d., and independent of the price process, we may add one more misspecification that the data are regularly spaced; hence we obtain the same estimator as before. Consistency can be established directly by conditioning on the observation time. In fact, this estimator can be viewed as the Pretend Fixed Maximum Likelihood (PFML) estimator discussed in Aït-Sahalia and Mykland (2003). In light of this, we may perform the Full Information Maximum Likelihood (FIML) or Integrated Out Maximum Likelihood (IOML) estimation, in order to utilize the information regarding the distribution of the random intervals, if available.

Also, it is possible to extend the i.i.d. sampling scheme to endogenous and stochastically spaced observations, by considering the time-changed processes as in Barndorff-Nielsen et al. (2008a). The extension is effortless in light of the connection between QMLE and RKs spelled out in Section 1.5.

Non-Gaussian and Serial Dependent Microstructure Noise

What about the robustness with respect to the noise? From Lemma 2 and Theorem 6, we find that the distribution of the i.i.d. noise does not affect the consistency. In other words, whatever the distribution of the noise is, it is misspecified to be Gaussian, and the QMLE obtained by maximizing (1.2) gives the same estimator of a_0^2 and the same order of

convergence rate, though the asymptotic variance may be different.

On the other hand, if the noises are in fact time-dependent, we can combine the QMLE with the subsampling method. For instance, if the noise itself features an MA(k) structure, we can divide the whole sample into $k+1$ disjoint parts such that the noises associated with the adjacent points in each subsample are uncorrelated. Then, we can simply apply the QMLE to each subsample and aggregate the estimates by taking the average.

Jumps

If the price process has jumps, then $\hat{\sigma}^2$ will converge to $\frac{1}{T}(\int_0^T \sigma_t^2 dt + \sum_t (\Delta X_t)^2)$ instead, which coincides with the TSRV estimator. The problem of separating jumps from volatility in this setting is more tedious, and may contribute little, since in most cases large jumps do not happen very often within a day. If they do occur, we may use the wavelet method in Fan and Wang (2007) to remove jumps before estimation, or separate the estimation periods by jumps and add up every piece of integrated volatility together, if the positions of jumps can be located accurately.

1.5 Comparisons with Realized Kernels

1.5.1 Estimation Methods

Realized Kernels (RKs) include a series of nonparametric estimators designed for volatility estimation in the presence of noise. Flat-top RKs take on the following form

$$K(\tilde{X}_\tau) = \gamma_0(\tilde{X}_\tau) + \sum_{h=1}^{n-1} k(\frac{h-1}{H})(\gamma_h(\tilde{X}_\tau) + \gamma_{-h}(\tilde{X}_\tau)) \tag{1.19}$$

where

$$\gamma_h(\tilde{X}_\tau) = \sum_{j=1}^{n} (\tilde{X}_{\tau_j} - \tilde{X}_{\tau_{j-1}})(\tilde{X}_{\tau_{j-h}} - \tilde{X}_{\tau_{j-h-1}}) \tag{1.20}$$

is the h^{th} sample autocovariance function and the kernel $k(\cdot)$ is a weight function. The most common finite-lag flat-top kernels are of the modified Tukey-Hanning type, defined

by:

$$k_{TH_p}(x) = \sin^2(\frac{\pi}{2}(1-x)^p) \cdot 1_{\{0 \leq x \leq 1\}}$$

In addition, there are infinite-lag realized kernels which may assign nonzero weight to all autocovariances, such as the optimal kernel:

$$k_{opt}(x) = (1+x)e^{-x}$$

As expected, the statistical properties of RKs vary as different kernels and bandwidths are selected. Therefore, it enables us to make different choices for specific purposes. However, the choice of the bandwidth may also become a burden in practice, in that the optimal bandwidth given by the theory cannot be estimated very accurately, and that the rule-of-thumb approximation of the bandwidth may not perform as well as the one selected in ad hoc ways. This problem may become even worse if the estimates were sensitive to the choice of bandwidth. By contrast, the QMLE is designed in a parametric way, which is free of bandwidth and kernel selection, while sharing the desired model-free feature with nonparametric estimators. It is by nature a different estimator and may not be regarded as one of the RKs directly.

1.5.2 Asymptotic Behavior and Finite Sample Performance

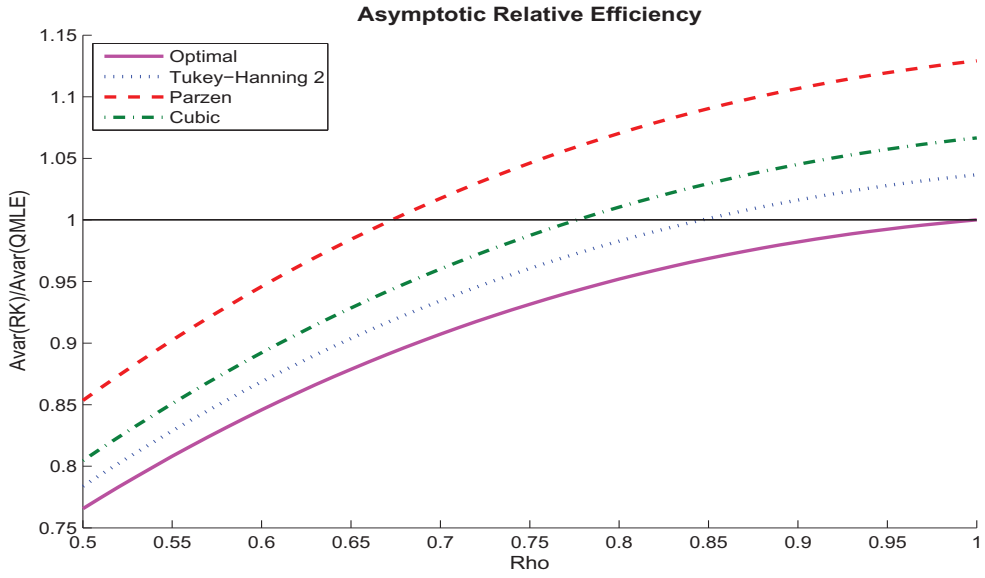
In regard to asymptotic efficiency, the kernel and bandwidth can be chosen for RKs to achieve the same optimal convergence rate as the QMLE. Nevertheless, when volatility is constant, the asymptotic variance of finite-lag kernels can only approximate the parametric variance bound, which, by contrast, can be obtained by the QMLE and the optimal kernel with infinite lags. When volatility is stochastic, the relative efficiency of the QMLE and

RKs depends on the extent of heteroskedasticity. More precisely, we have[5]

$$\frac{Avar(\text{RK})}{Avar(\text{QMLE})} = \frac{16\sqrt{\rho k_0 k_1}}{3} \frac{\left(1+\sqrt{1+3k_0k_2/\rho k_1^2}\right)^{-\frac{1}{2}} + \left(1+\sqrt{1+3k_0k_2/\rho k_1^2}\right)^{\frac{1}{2}}}{5\rho^{-\frac{1}{2}} + 3\rho^{\frac{3}{2}}}$$

where $\rho = \int_0^T \sigma_u^2 du / \sqrt{T\int_0^T \sigma_u^4 du}$ measures the variability of the volatility process, and k_0, k_1 and k_2 are constants determined by the selected kernel. Figure 1.1 plots the relative efficiency of four typical RKs against the QMLE. Apparently, the QMLE tends to be more favorable than finite-lag kernels as ρ becomes larger, while RKs are better when ρ is small. The optimal kernel with theoretically optimal bandwidth, however, fully dominates the QMLE asymptotically except for the case $\rho = 1$, when volatility is constant. Intuitively, the smaller ρ is, the further the misspecified model deviates from the truth.

Figure 1.1: Asymptotic Relative Efficiency of the QMLE and RKs



A major drawback of RKs is that they require a number of out-of-period intraday returns because of the construction of the autocovariance estimator $\gamma_h(\tilde{X}_\tau)$. For this reason, infinite-lag kernels, in particular, are not implementable empirically. In practice, the feasi-

[5]The asymptotic variance of the RK obtained here requires a refinement of the endpoints as well as the optimal bandwidth.

ble autocovariance estimator is implemented by

$$\tilde{\gamma}_{\pm h}(\tilde{X}_\tau) = \sum_{j=H+1}^{n-H} (\tilde{X}_{\tau_j} - \tilde{X}_{\tau_{j-1}})(\tilde{X}_{\tau_{j-h}} - \tilde{X}_{\tau_{j-h-1}}) \tag{1.21}$$

With finite-lag flat-top kernels, the cut-off near the boundary is not an issue asymptotically, however it may yield a large bias in finite sample. As the sample size n decreases, namely, H/n increases, this bias becomes more evident. The next section applies Monte Carlo simulations to demonstrate the edge effect.

1.5.3 Quadratic Representation and Weighting Matrices

An intuitive way to understand the similarities and differences between the QMLE and RKs is to regard them as quadratic estimators. More specifically, we have the following quadratic iterative representation of the QMLE.

Theorem 8. *The QMLE $(\hat{\sigma}^2, \hat{a}^2)$ satisfies the following equations:*

$$\hat{\sigma}^2 T = Y' W_1(\hat{\sigma}^2, \hat{a}^2) Y \tag{1.22}$$

$$\hat{a}^2 = Y' W_2(\hat{\sigma}^2, \hat{a}^2) Y \tag{1.23}$$

The weighting matrices satisfy:

$$W_1(\sigma^2, a^2) = \frac{n \cdot tr(\Omega^{-2}\Lambda) \cdot \Omega^{-1}\Lambda\Omega^{-1} - n \cdot tr(\Omega^{-2}\Lambda^2) \cdot \Omega^{-2}}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2}) \cdot tr(\Omega^{-2}\Lambda^2)} \tag{1.24}$$

$$W_2(\sigma^2, a^2) = \frac{tr(\Omega^{-2}\Lambda) \cdot \Omega^{-2} - tr(\Omega^{-2}) \cdot \Omega^{-1}\Lambda\Omega^{-1}}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2}) \cdot tr(\Omega^{-2}\Lambda^2)} \tag{1.25}$$

where Ω is given by (1.3), and

$$\Lambda = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & -1 & 2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}$$

Also, $W_1(\sigma^2, a^2)$ and $W_2(\sigma^2, a^2)$ depend on σ^2 and a^2 only through $\lambda^2 = a^2/\sigma^2 T$.

Similarly, a feasible realized kernel estimator can be expressed as

$$K(\tilde{X}_\tau) = Y'WY \tag{1.26}$$

where W is determined by the kernel $k(\cdot)$ and bandwidth H.

$$\begin{aligned} W_{i,i} =& 1_{\{1+H\leq i\leq n-H\}} \\ W_{i,j} =& k(\frac{|i-j|-1}{H}) \cdot 1_{\{1\leq |i-j|\leq H\}} \cdot 1_{\{1+H\leq j\leq n-H\}} \end{aligned}$$

Figure 1.2 plots the weighting matrices against the row and column indices for the QMLE and Turkey-Hanning$_2$ kernel. The plot displays similar features such as the weights on the diagonals are either 1 or very close to 1 within several lags, and decay to zero beyond that. The following theorem illustrates the implicit connection between the two estimators in terms of the asymptotic behavior.

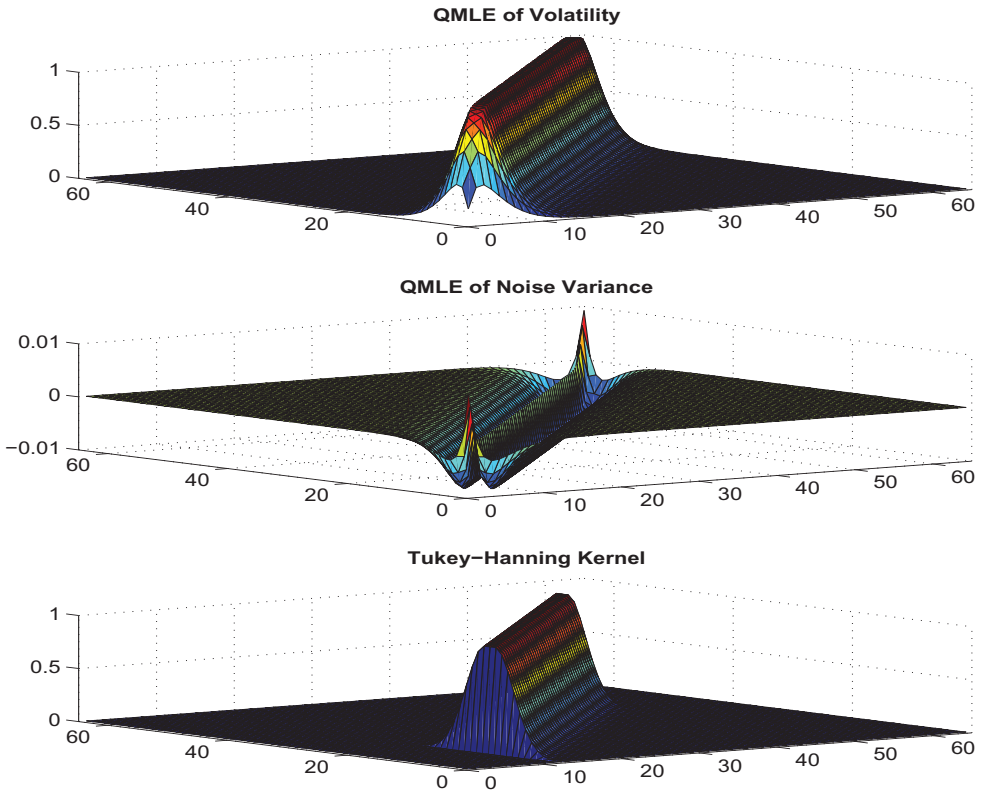
Theorem 9. *The QMLE is asymptotically equivalent to the optimal kernel with implicit bandwidth $H = \hat{\lambda} \cdot n^{\frac{1}{2}} = a_0/(\int_0^T \sigma_t^2 dt)^{\frac{1}{2}} n^{\frac{1}{2}}$. In other words, for any $K = n^{\frac{1}{2}+\delta}$, $0 < \delta < \frac{1}{2}$, and any $K \leq i, j \leq n-K$, we have*

$$W_{1,i,j}(\sigma^2, a^2) \approx k_{opt}(\frac{|i-j|}{\lambda \cdot n^{\frac{1}{2}}})$$

Theorem 9 also points out that the weighting matrix $W_1(\sigma^2, a^2)$ is approximately a symmetric Toeplitz matrix with equal weight along the diagonal, barring the boundary effect. The implicit bandwidth depends on the parameters of interest and is suboptimal. It therefore explains why the optimal kernel with optimal bandwidth asymptotically dominates the QMLE in Figure 1.1. Nevertheless, the ending points of the diagonals have different patterns, indicating different treatments of the edge effect. Apparently, the QMLE controls the weights on the boundary in a more natural way, and hence has better finite sample performance.

The quadratic representation also sheds light on the differences in the procedures of

Figure 1.2: Weighting Matrices of the QMLE and the RK



Note: We plot the two weighting matrices against their column and row indices. The parameters are $\sigma^2 = 0.1$, $a = 0.005$, $n = 65$ and $T = 1/252$. The bandwidth $H = 5.74a\sqrt{n/\sigma^2 T}$.

estimation. The bandwidth H of the RK is either estimated as first step or selected in an ad hoc way, while the bandwidth $\hat{\lambda}\cdot n^{\frac{1}{2}}$ of the QMLE is automatically updated by the optimization algorithm, or more intuitively, by iterating (1.22) and (1.23). The weighting matrix of the QMLE is therefore more adaptive. In view of this, it is natural to construct a one-step alternative for the QMLE, which, instead of running nonlinear optimization, employs a consistent plug-in of $\hat{\lambda}$ for (1.22) and (1.23). This one-step volatility estimator coincides with one of the unbiased quadratic estimators given by Sun (2006), where its asymptotic properties are discussed under the constant volatility assumption. The quadratic forms of other nonparametric estimators are also included in Andersen et al. (2009).

1.6 Simulation Studies with High Frequency Data

1.6.1 Asymptotic Behavior of the QMLE

We at first conduct Monte Carlo simulations to verify the asymptotic results given in Theorem 7. We fix T as 1 day. Within $[0, T]$, the data are simulated using Euler scheme based on stochastic volatility models, for instance the Heston Model with jumps in volatility process.

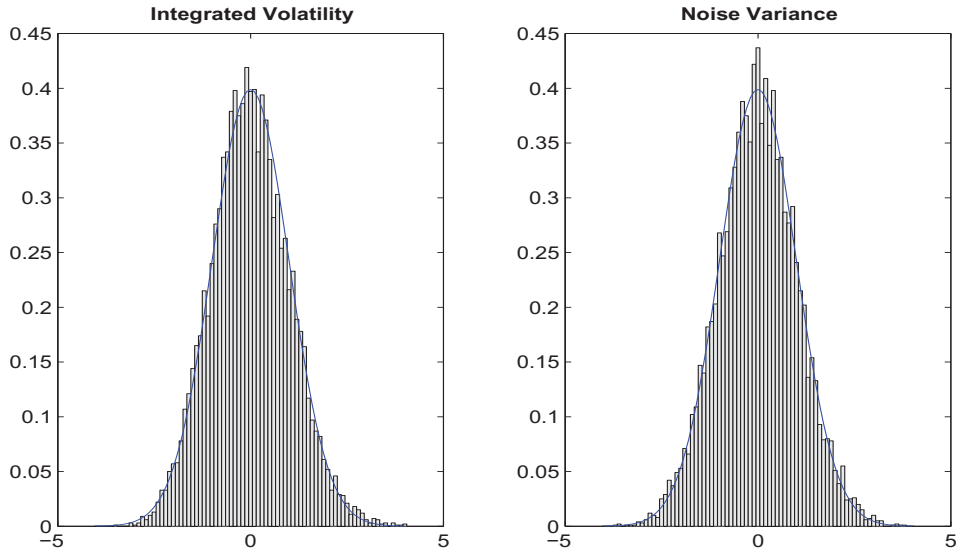
$$\begin{aligned} dX_t =& \mu dt + \sigma_{t-} dW_t \\ d\sigma_t^2 =& \kappa(\bar{\sigma}^2 - \sigma_t^2)dt + \delta\sigma_{t-} dB_t + \sigma_{t-} J_t^V dN_{2t} \end{aligned}$$

where $E(dW_t \cdot dB_t) = \rho dt$. The arrival of transactions follows a Poisson process and the mean interarrival time is 1 second. The true values of parameters are consistent with those chosen by Aït-Sahalia and Yu (2009). Specifically, the drift μ is 0.03, and to add the leverage effect, ρ is selected to be -0.75. $\kappa = 5$ and the volatility of volatility is $\delta = 0.4$. σ_0^2 is sampled from the stationary distribution of the CIR process, that is Gamma$(2\kappa\bar{\sigma}^2/\delta^2, \delta^2/2\kappa)$, so that the unconditional mean of volatility process is exactly $\bar{\sigma}^2$, chosen as 0.1. The noise satisfies normal distribution with standard deviation $a_0 = 0.5\%$. The jumps follow a Poisson process N_{2t} independent of the price and volatility processes with intensity $\lambda = 12$. The jump size in volatility is $J_t^V = \exp(z)$, where $z \sim N(-5, 1)$. The number of Monte

Carlo sample paths is 10000.

Figure 1.3 shows the histograms of the standardized estimates compared with their asymptotic distributions. Moreover, Table 1.1 compares the sample quartile statistics with their theoretic benchmarks given by the standard Gaussian distribution. All these simulation results reconfirm our asymptotic theory.

Figure 1.3: Histograms of the Standardized Estimates



Note: We plot the histograms of the standardized estimates. The true noise variance is 0.005^2. The simulations include random intervals and volatility jumps. The density of the asymptotic distribution is also plotted.

1.6.2 Comparisons with Realized Kernels

For comparison purpose, we implement the Tukey-Hanning$_2$ kernel which is one of the flat-top kernels that converge at the best rate. It is considerably efficient compared with other kernels in Barndorff-Nielsen et al. (2008a), and it does not require too many out-of-period data.

As a benchmark, we first implement the Tukey-Hanning$_2$ kernel with the infeasible bandwidth, $H = c^* \xi n^{\frac{1}{2}}$ with c^* given by

$$c^* = \sqrt{\rho \frac{k_1}{k_0}(1 + \sqrt{1 + \frac{3k_0 k_2}{\rho k_1^2}})}$$

Table 1.1: Finite Sample Performance of the QMLE

No. obs	Mean	Stdv.	0.5%	2.5%	5%	95%	97.5%	99.5%
σ^2								
130	0.0099	1.1867	0.0001	0.0112	0.0422	0.9073	0.9339	0.9672
390	-0.0164	1.0807	0.0010	0.0170	0.0455	0.9273	0.9541	0.9830
780	0.0045	1.0861	0.0015	0.0175	0.0469	0.9248	0.9539	0.9818
1170	0.0044	1.0502	0.0022	0.0175	0.0454	0.9311	0.9579	0.9877
2340	-0.0017	1.0328	0.0025	0.0183	0.0440	0.9353	0.9639	0.9890
4680	-0.0032	1.0112	0.0025	0.0167	0.0432	0.9403	0.9660	0.9898
23400	-0.0057	1.0131	0.0031	0.0236	0.0492	0.9464	0.9706	0.9912
a^2								
130	-0.0084	1.2284	0.0135	0.0464	0.0818	0.9079	0.9400	0.9779
390	-0.0069	1.1146	0.0079	0.0355	0.0641	0.9236	0.9578	0.9859
780	0.0015	1.0909	0.0073	0.0334	0.0631	0.9315	0.9609	0.9878
1170	0.0037	1.0656	0.0057	0.0297	0.0558	0.9356	0.9627	0.9893
2340	-0.0186	1.0350	0.0066	0.0294	0.0554	0.9439	0.9707	0.9934
4680	-0.0063	1.0319	0.0063	0.0290	0.0551	0.9440	0.9700	0.9939
23400	-0.0026	1.0163	0.0047	0.0279	0.0533	0.9465	0.9712	0.9932

Note: In this table, we report the finite sample quartiles of the infeasible standardized QMLE, which employs the theoretic asymptotic variance. The benchmark quartiles are those for the limit distribution $N(0,1)$.

where $\xi^2 = a_0^2/\sqrt{T\int_0^T \sigma_u^4 du}$, $\rho = \int_0^T \sigma_u^2 du/\sqrt{T\int_0^T \sigma_u^4 du}$, $k_0 = 0.219$, $k_1 = 1.71$ and $k_2 = 41.7$. Also, the edge effect is eliminated in that the calculation, using formula (1.20), includes the out-of-period data. Hence, this estimator would provide us with the best Tukey-Hanning$_2$ kernel estimate (RK$_1$) in theory.

Next, we make the edge effect stand out by approximating the autocovariances using (1.21), while keeping the theoretical optimal bandwidth as before. We denote the second estimator as RK$_2$.

Finally, to mimic the empirical applications of these estimators, we redo the experiments with the simple rule-of-thumb choice of the bandwidth:

$$\hat{H} = 5.74\hat{a}\sqrt{n/RV_{10}(\tilde{X})}$$

where $\hat{a} = \sqrt{RV(\tilde{X})/2n}$ and $RV_{10}(\tilde{X})$ is the RV estimator based on 10 min returns. So we obtain the corresponding RK$_3$ and RK$_4$.

The simulations are based on the same stochastic volatility model and parameters as before with a longer time window $[0, 3T]$. We regard $[T, 2T]$ as in-sample period, and the rest time interval is only used for infeasible kernels. The results of estimation are reported in Table 1.2. It is evident that the QMLE dominates the four implementations of the Tukey-Hanning$_2$ kernel in terms of the RMSE. As expected, RK$_1$ is very close to the QMLE, and better than the other three kernels, in that it employs more data and is equipped with theoretically optimal bandwidth. Comparing the first two kernels, we can also see that the edge effect results in a large bias when the sample size is relatively small, and become dampened as the sample size goes up. In addition, the differences between the first two kernels and the last two are the losses due to the suboptimal choice of the bandwidth. In practice, only RK$_4$ can be applied, which may suffer from larger bias and variance because of the two sources of losses.

Table 1.2: Comparison between the QMLE and Tukey-Hanning$_2$ Kernel

		1 sec	5 sec	10 sec	20 sec	30 sec	1 min	3 min
QMLE	Bias	0.0189	0.0480	-0.0040	-0.0065	-0.0341	-0.0452	-0.0838
	Stdv	1.1860	1.8172	2.1611	2.6040	2.9270	3.5828	4.9203
	RMSE	1.1861	1.8178	2.1611	2.6040	2.9272	3.5831	4.9210
RK$_1$	Bias	0.0178	0.0465	0.0074	0.0092	-0.0079	-0.0192	0.0003
	Stdv	1.2328	1.8844	2.2424	2.6835	3.0087	3.6851	5.0185
	RMSE	1.2329	1.8850	2.2424	2.6836	3.0087	3.6851	5.0185
RK$_2$	Bias	-0.1620	-0.3659	-0.5794	-0.8407	-1.0288	-1.4913	-2.5874
	Stdv	1.2197	1.8468	2.1847	2.5873	2.8704	3.4239	4.4245
	RMSE	1.2305	1.8827	2.2602	2.7205	3.0492	3.7345	5.1255
RK$_3$	Bias	0.0186	0.0247	-0.0048	0.0699	-0.0452	0.0009	0.3615
	Stdv	1.8603	2.7555	3.2438	3.7953	4.1695	4.8708	6.2255
	RMSE	1.8604	2.7556	3.2438	3.7960	4.1697	4.8708	6.2360
RK$_4$	Bias	-0.0641	-0.1568	-0.2591	-0.3008	-0.4934	-0.6316	-0.7731
	Stdv	1.8546	2.7293	3.2091	3.7537	4.0768	4.7444	5.9674
	RMSE	1.8557	2.7337	3.2196	3.7658	4.1066	4.7863	6.0173

Note: This table reports the estimates for $100 \cdot (\hat{\sigma}^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt)$, where $\hat{\sigma}^2$ is given by the QMLE and various implementations of the Tukey-Hanning$_2$ kernel respectively. Among them, RK$_1$ and RK$_2$ have the theoretical bandwidth, while RK$_1$ and RK$_3$ employ out-of-period data.

1.7 Empirical Work with the EUR/USD Futures

In this section, we are interested in estimating the realized variance of the Euro/US dollar futures carried out on the Chicago Mercantile Exchange (CME) in the year 2008. The contracts are actively traded on the 24-hour clock, and quoted in terms of the unit value of the Euro as measured in US dollars. The high frequency data are available from Tick Data Inc.

The foreign exchange markets are less active during the weekends and holidays; therefore, we eliminate the transactions on Saturdays and Sundays as well as US federal holidays. In addition, we also exclude January 2, the day after Thanksgiving, December 24 to 26, and December 31. Last, we make everyday begin at 5pm Chicago Time when the electronic trading starts so as to eliminate the potential price jumps between the one hour trading gap from 4pm to 5pm. The summary statistics of the data are provided in Table 1.3. Evidently, the microstructure noise displays an MA(1) structure.

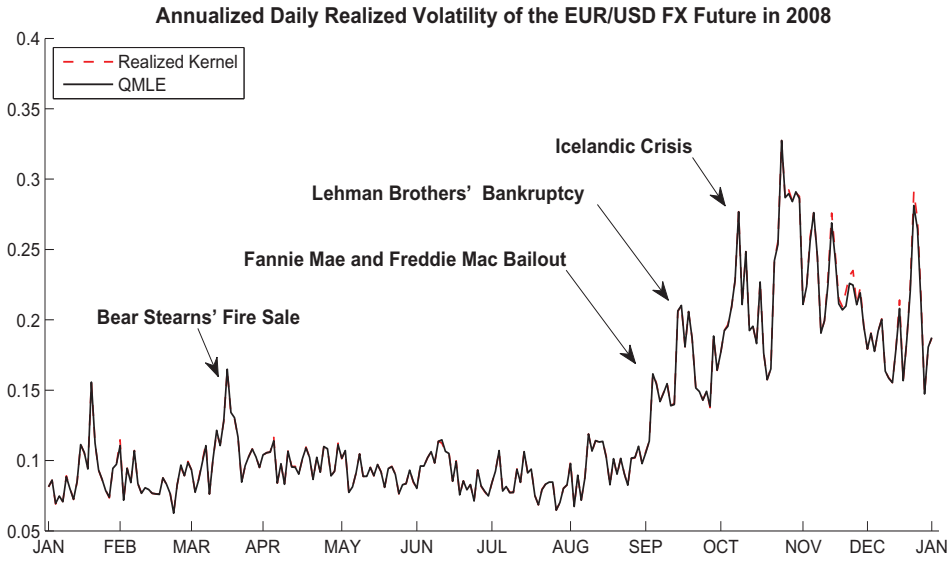
Table 1.3: Summary Statistics of the Log Returns of the Euro/US Dollar Future

Avg No of Obs	Avg Freq	Mean	Std Err	1st Lag	2nd Lag
19657	4.44s	6.48e-09	6.12e-05	-0.075	0.0092

We apply both the QMLE and the Tukey-Hanning$_2$ kernel with the rule-of-thumb bandwidth and plot the annualized daily realized volatility in Figure 1.4. It is apparent from the plot that the two methods give almost identical estimates and the differences are statistically insignificant. This is in agreement with our Monte Carlo simulation result, where the two estimates are indistinguishable when the trading frequency is as high as several seconds.

Moreover, Figure 1.4 provides ample evidence of structure breaks or jumps in the volatility process starting from September 2008, which echoes the fact that, since then, the global financial crisis has entered its most critical stage. Further, these jumps or large movements are usually associated with financial news: for instance, the government seizure of Fannie Mae and Freddie Mac, Lehman Brothers' bankruptcy in early September, and the collapse of three major banks of Iceland in early October, etc. These findings are harder to obtain by modeling lower frequency data.

Figure 1.4: The EUR/USD Future: A Case Study



1.8 Conclusions

This chapter contributes to the estimation of integrated volatility by showing that the popular MLE, as a new quasi-estimator, is consistent, efficient and robust with respect to stochastic volatility. Moreover, this parametric estimator only involves an optimization procedure, which is free from bandwidth selection, and hence very convenient in practice. More interestingly, this seemingly inappropriate estimator turns out to be asymptotically equivalent to the optimal realized kernel with an implicitly specified bandwidth, and dominant over alternative realized kernels in terms of the finite sample accuracy.

This chapter makes another contribution to the analysis of model specification by extending the classic asymptotic theory of the QMLE to a stochastic parameter setting, and by showing an example of misspecifying models on purpose, which gives rise to facility and feasibility in estimation. This study may be applicable to more cases where stochastic volatility plagues the estimation, such as covariance estimation, or more generally, where the object of interest is stochastic.

1.9 Appendix

1.9.1 Proof of Theorem 1

For each $\delta > 0$, with probability approaching 1 (w.p.a.1), we have[6]

$$\left.\begin{aligned} Q_n(\omega, \hat{\theta}_n) &> Q_n(\omega, \theta_n^*) - \tfrac{\delta}{3} \\ \bar{Q}_n(\omega, \hat{\theta}_n) &> Q_n(\omega, \hat{\theta}_n) - \tfrac{\delta}{3} \\ Q_n(\omega, \theta_n^*) &> \bar{Q}_n(\omega, \theta_n^*) - \tfrac{\delta}{3} \end{aligned}\right\} \text{Due to Uniform Convergence}$$

Thus, $\forall\ \delta > 0$, w.p.a.1, $\bar{Q}_n(\omega, \hat{\theta}_n) > \bar{Q}_n(\omega, \theta_n^*) - \delta$.

Further, for any $\epsilon > 0$, let $N(\omega) := \{\theta : \|\theta - \theta_n^*\| < \epsilon\}$. Then for each ω, $N^c \cap \Theta$ is compact, and $\max_{\theta \in N^c \cap \Theta} \bar{Q}_n(\omega, \theta) = \bar{Q}_n(\omega, \tilde{\theta}) \leq \bar{Q}_n(\omega, \theta_n^*)$. Let

$$\delta_n(\omega) := \bar{Q}_n(\omega, \theta_n^*) - \max_{\theta \in N^c \cap \Theta} \bar{Q}_n(\omega, \theta)$$

We have, for any $\delta < \delta_0$,

$$\begin{aligned} &P(\bar{Q}_n(\omega, \hat{\theta_n}) > \max_{\theta \in N^c \cap \Theta} \bar{Q}_n(\omega, \theta)) \\ =&P(\bar{Q}_n(\omega, \hat{\theta}_n) > \bar{Q}_n(\omega, \theta_n^*) - \delta_n) \\ \geq&P(\bar{Q}_n(\omega, \hat{\theta}_n) > \bar{Q}_n(\omega, \theta_n^*) - \delta_n, \delta_n > \delta) \\ \geq&P(\bar{Q}_n(\omega, \hat{\theta}_n) > \bar{Q}_n(\omega, \theta_n^*) - \delta, \delta_n > \delta) \to 1 \end{aligned}$$

Therefore, $P(\|\hat{\theta}_n - \theta_n^*\| < \epsilon) \to 1$.

1.9.2 Proof of Theorem 2

It follows from Theorem 1, on applying it to $Q_n(\omega, \theta) = -\|\Psi_n(\omega, \theta)\|$ and $\bar{Q}_n(\omega, \theta) = -\|\bar{\Psi}_n(\omega, \theta)\|$.

[6]The statements in this proof are well posed if any desired measurability is guaranteed. However, this measurability issue can be ignored by redefining all concepts in terms of outer measure, see Newey and McFadden (1994, pp 2121).

1.9.3 Proof of Theorem 3

For simplicity, it is sufficient to prove this result for $k = 1$. Because

$$\Psi_n(\omega, \theta) - \Psi_n(\omega, \theta_n^*) = \nabla\Psi_n(\omega, \tilde{\theta}_n)(\theta - \theta_n^*)$$

Plug in $\hat{\theta}_n$ and multiply both sides by V_n, then we have

$$V_n(\omega)\nabla\Psi_n(\omega, \tilde{\theta}_n)(\hat{\theta}_n - \theta_n^*) = -V_n(\omega)\Psi_n(\omega, \theta_n^*)$$

Since $\hat{\theta}_n - \theta_n^* \xrightarrow{\text{P}} 0$, $\nabla\bar{\Psi}_n$ is stochastic equicontinuous, and $|\nabla\Psi_n(\omega, \theta) - \nabla\bar{\Psi}_n(\omega, \theta)| \xrightarrow{\text{P}} 0$, uniformly for all $\theta \in \Theta$, it follows from an analogous reasoning as in Theorem 2.3 in Domowitz and White (1982) that $\nabla\Psi_n(\omega, \tilde{\theta}_n) - \nabla\bar{\Psi}_n(\omega, \theta_n^*) \xrightarrow{\text{P}} 0$, which concludes the proof.

1.9.4 Proof of Lemma 1

Let $\Omega^{-1} = (\omega^{ij})$ and $X_{\tau_i} = \int_0^{\tau_i} \sigma_t dW_t$. We also define $M_{\tau_i} = X_{\tau_i}^2 - \langle X_{\tau_i}, X_{\tau_i}\rangle = (\int_0^{\tau_i} \sigma_t dW_t)^2 - \int_0^{\tau_i} \sigma_t^2 dt$. Then $\{X_{\tau_i}\}_{1\le i\le n}$ and $\{M_{\tau_i}\}_{1\le i\le n}$ are martingales. Also, we can add one more condition that $B^{-1} \le \sigma_t^2 \le B, \forall\ t \in [0, T]$, in that a regular localization procedure always applies. For convenience, we change the variables:

$$\gamma^2(1 + \eta^2) = \sigma^2\Delta + 2a^2$$
$$\gamma^2\eta = -a^2$$

Then the inverse change of variables is given by,

$$\eta = \frac{1}{2a^2}\{-2a^2 - \sigma^2\Delta + \sqrt{\sigma^2\Delta(4a^2 + \sigma^2\Delta)}\} \tag{1.27}$$
$$\gamma^2 = \frac{1}{2}\{2a^2 + \sigma^2\Delta + \sqrt{\sigma^2\Delta(4a^2 + \sigma^2\Delta)}\} \tag{1.28}$$

and we have

$$\omega^{ij} = \frac{(-\eta)^{|i-j|} - (-\eta)^{i+j} - (-\eta)^{2n-i-j+2} + (-\eta)^{2n-|i-j|+2}}{\gamma^2(1-\eta^2)(1-\eta^{2n+2})} \tag{1.29}$$

One can check $\omega^{ij} \geq 0$. In addition, ω^{ii} is a concave function of i, and clearly attains its maximum at $i = \frac{n+1}{2}$ and minimum at the boundary. In addition, by (1.27) and (1.28)

$$\begin{aligned}
\eta =& -1 + \sqrt{\frac{\sigma^2 T}{a^2}} n^{-\frac{1}{2}} - \frac{\sigma^2 T}{2a^2} n^{-1} + O(n^{-\frac{3}{2}}) \\
\eta^2 =& 1 - 2\sqrt{\frac{\sigma^2 T}{a^2}} n^{-\frac{1}{2}} + \frac{2\sigma^2 T}{a^2} n^{-1} + O(n^{-\frac{3}{2}}) \\
\log \eta^2 =& -2\sqrt{\frac{\sigma^2 T}{a^2}} n^{-\frac{1}{2}} + O(n^{-\frac{3}{2}}) \\
\gamma^2 =& a^2 + \sqrt{a^2 \sigma^2 T} n^{-\frac{1}{2}} + \frac{\sigma^2 T}{2} n^{-1} + O(n^{-\frac{3}{2}})
\end{aligned}$$

Then, it can be easily shown that $\omega^M := \omega^{\frac{n+1}{2}, \frac{n+1}{2}} = O(n^{\frac{1}{2}})$. By the martingale property, we have

$$\begin{aligned}
& E(\sum_{i=1}^{n} \omega^{ii} \{(\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\})^2 \\
=& \sum_{i=1}^{n} (\omega^{ii})^2 E\{(\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\}^2 \\
\leq& (\omega^M)^2 \sum_{i=1}^{n} E\{(\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\}^2
\end{aligned}$$

As in the no noise case, we have

$$\sum_{i=1}^{n} \{(\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\} = O_p(n^{-\frac{1}{2}}) \tag{1.30}$$

Combining the two, we obtain

$$E(\sum_{i=1}^{n} \omega^{ii} \{(\int_{\tau_{i-1}}^{\tau_i} \sigma_t dW_t)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\})^2 = (O(n^{\frac{1}{2}}))^2 O(n^{-1}) = O(1)$$

Thus, we prove (1.14) by Chebyshev's inequality. To prove (1.15), note that $\omega^{ij} = \omega^{ji}$, so it follows from the martingale property that

$$
\begin{aligned}
& E(\sum_{i=1}^{n}\sum_{j\neq i}^{n}\omega^{ij}\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t\int_{\tau_{j-1}}^{\tau_j}\sigma_t dW_t)^2 \\
=&4E(\sum_{i}\sum_{j>i}\sum_{k}\sum_{l>k}\omega^{ij}\omega^{kl}\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t\int_{\tau_{j-1}}^{\tau_j}\sigma_t dW_t\int_{\tau_{k-1}}^{\tau_k}\sigma_t dW_t\int_{\tau_{l-1}}^{\tau_l}\sigma_t dW_t) \\
=&4\sum_{j=1}^{n}E\{(\int_{\tau_{j-1}}^{\tau_j}\sigma_t dW_t)^2(\sum_{k=1}^{j-1}\omega^{kj}\int_{\tau_{k-1}}^{\tau_k}\sigma_t dW_t)^2\} \\
\leq&4B\Delta\sum_{j=1}^{n}E(\sum_{k=1}^{j-1}\omega^{kj}\int_{\tau_{k-1}}^{\tau_k}\sigma_t dW_t)^2 \\
=&4B\Delta\sum_{j=1}^{n}\sum_{k=1}^{j-1}(\omega^{kj})^2E(\int_{\tau_{k-1}}^{\tau_k}\sigma_t dW_t)^2 \\
\leq&4B^2\Delta^2\sum_{j=1}^{n}\sum_{k<j}(\omega^{kj})^2 \\
=&O(n^{\frac{1}{2}})
\end{aligned}
$$

Finally, we prove (1.16). Because for any $j \neq i$, $2\omega^{ji} - \omega^{j-1,i} - \omega^{j+1,i} \leq 0$, and $\omega^{ij} > 0$.

$$
\begin{aligned}
& E(\sum_{i=1}^{n}\sum_{j=1}^{n}\omega^{ij}\epsilon_j\int_{\tau_{i-1}}^{\tau_i}\sigma_t dW_t)^2 \\
=&a_0^2\sum_{i=1}^{n}\sum_{j=1}^{n}\{\omega^{ij}(2\omega^{ji} - \omega^{j-1,i} - \omega^{j+1,i})E\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2 dt\} \\
\leq&a_0^2 B\Delta\sum_{i=1}^{n}\omega^{ii}(2\omega^{ii} - \omega^{i-1,i} - \omega^{i+1,i}) \\
=&O(n^{\frac{1}{2}})
\end{aligned}
$$

So all the conclusions claimed in the lemma hold.

1.9.5 Proof of Lemma 2

Let $\epsilon = (\epsilon_{\tau_1}, \epsilon_{\tau_2}, \ldots, \epsilon_{\tau_n})'$. Denote K^i, $K^{i,j}$, $K^{i,j,k}$, and $K^{i,j,k,l}$ the corresponding cumulants of $\{\epsilon_{\tau_i}\}$. Then $K^i = 0$, since the mean of ϵ_{τ_i} is zero. According to McCullagh (1987, Section

3.3) , we have

$$
\begin{aligned}
& var(\epsilon'\Omega\epsilon) \\
=& \sum_{i,j,k,l=1}^{n} \omega^{ij}\omega^{kl}\Big(K^{i,j,k,l} + 4K^{i}K^{j,k,l} + 2K^{i,k}K^{j,l} + 4K^{i}K^{k}K^{j,l}\Big) \\
=& \sum_{i,j,k,l=1}^{n} \omega^{ij}\omega^{kl}\Big(K^{i,j,k,l} + 2K^{i,k}K^{j,l}\Big) \\
=& \sum_{i,j,k,l=1}^{n} \omega^{ij}\omega^{kl}\Big(\text{cum}(\epsilon_{\tau_i},\epsilon_{\tau_j},\epsilon_{\tau_k},\epsilon_{\tau_l}) + 2\text{cov}(\epsilon_{\tau_i},\epsilon_{\tau_k})\text{cov}(\epsilon_{\tau_j},\epsilon_{\tau_l})\Big) \\
:=& V_1(\omega,\omega) + V_2(\omega,\omega)
\end{aligned}
$$

where

$$
\begin{aligned}
V_1(v,\omega) =& \sum_{i,j,k,l=1}^{n} v^{ij}\omega^{kl}\text{cum}(\epsilon_{\tau_i},\epsilon_{\tau_j},\epsilon_{\tau_k},\epsilon_{\tau_l}) \\
V_2(v,\omega) =& \sum_{i,j,k,l=1}^{n} v^{ij}\omega^{kl}2\text{cov}(\epsilon_{\tau_i},\epsilon_{\tau_k})\text{cov}(\epsilon_{\tau_j},\epsilon_{\tau_l}) \\
=& 2a_0^4\sum_{i=1}^{N}\sum_{j=1}^{N}\{v^{ij}(\omega^{j-1,i-1} + \omega^{j-1,i+1} - 2\omega^{j-1,i} + \omega^{j+1,i-1} + \omega^{j+1,i+1} \\
& - 2\omega^{j+1,i} - 2(\omega^{j,i-1} + \omega^{j,i+1} - 2\omega^{j,i}))\}
\end{aligned}
$$

Recall that in Lemma 1 of Aït-Sahalia et al. (2005),

$$
\begin{aligned}
& \text{cum}(\epsilon_{\tau_i},\epsilon_{\tau_j},\epsilon_{\tau_k},\epsilon_{\tau_l}) \\
=& \begin{cases} 2\text{cum}_4[U], & \text{if } i=j=k=l; \\ (-1)^{s(i,j,k,l)}\text{cum}_4[U], & \text{if } \max(i,j,k,l) = \min(i,j,k,l)+1; \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

where $s(i,j,k,l)$ denotes the number of indices among (i,j,k,l) that are equal to $\min(i,j,k,l)$.

So using this lemma, we can obtain

$$V_1(v,\omega)$$

$$
\begin{aligned}
=&\text{cum}_4[U]\Big(2\sum_{i=1}^{n} v^{ii}\omega^{ii} + \sum_{i=1}^{n-1}(-4v^{i,i+1}\omega^{i+1,i+1} - 4v^{i+1,i+1}\omega^{i,i+1} + 2v^{i,i}\omega^{i+1,i+1} \\
&+ 4v^{i,i+1}\omega^{i,i+1})\Big)
\end{aligned}
$$

To facilitate calculation, we rewrite

$$
\begin{aligned}
&V_2(\frac{\partial\omega}{\partial\sigma^2}, \frac{\partial\omega}{\partial\sigma^2}) \\
=&(\frac{\partial\eta}{\partial\sigma^2})^2 V_2(\frac{\partial\omega}{\partial\eta}, \frac{\partial\omega}{\partial\eta}) + \frac{1}{\gamma^4}(\frac{\partial\gamma^2}{\partial\sigma^2})^2 V_2(\omega,\omega) - \frac{1}{\gamma^2}\frac{\partial\eta}{\partial\sigma^2}\frac{\partial\gamma^2}{\partial\sigma^2}\frac{\partial V_2(\omega,\omega)}{\partial\eta}
\end{aligned}
$$

Then direct computation yields,

$$
\begin{aligned}
V_2(\frac{\partial\omega}{\partial\eta}, \frac{\partial\omega}{\partial\eta}) =& \frac{4a_0^4\{3(9-7n)\eta^2 + 3(19+3n)\eta^3 + 15(1+n)\eta^4 - 3(1+n)\eta^5\}}{3\gamma^4(-1-\eta)\eta^2(1-\eta)^6} + o(1) \\
=& \frac{a_0^4 n^{\frac{3}{2}}}{4a^3\sigma\sqrt{T}} + o(n^{\frac{3}{2}}) \qquad (1.31) \\
V_2(\omega,\omega) =& \frac{4a_0^4}{\gamma^4(1-\eta)^4}\{-1-4\eta+\eta^2+n(1-\eta)(3-\eta)\} + o(1) \\
=& \frac{2na_0^4}{a^4} + o(n) \qquad (1.32) \\
\frac{\partial V_2(\omega,\omega)}{\partial\eta} =& \frac{-8a_0^4\{4(-1+n)\eta - 5(1+n)\eta^2 + (1+n)\eta^3\}}{-\gamma^4\eta(1-\eta)^5} + o(1) \\
=& \frac{5na_0^4}{2a^4} + o(n) \qquad (1.33)
\end{aligned}
$$

Note that

$$
\begin{aligned}
\frac{\partial\eta}{\partial\sigma^2} =& \frac{1}{2}\sqrt{\frac{T}{a^2\sigma^2}}n^{-\frac{1}{2}} - \frac{T}{2a^2}n^{-1} + O(n^{-\frac{3}{2}}) \\
\frac{\partial\gamma^2}{\partial\sigma^2} =& \frac{1}{2}\sqrt{\frac{a^2T}{\sigma^2}}n^{-\frac{1}{2}} + \frac{T}{2}n^{-1} + O(n^{-\frac{3}{2}})
\end{aligned}
$$

Combining (1.31), (1.32) and (1.33), we obtain,

$$
V_2(\frac{\partial\omega}{\partial\sigma^2}, \frac{\partial\omega}{\partial\sigma^2}) = \frac{\sqrt{T}a_0^4 n^{\frac{1}{2}}}{16a^5\sigma^3} + o(n^{\frac{1}{2}}) \qquad (1.34)
$$

Similarly,

$$\begin{aligned}\frac{\partial \eta}{\partial a^2} =& -\frac{\sqrt{a^2\sigma^2 T}}{2a^4}n^{-\frac{1}{2}} + \frac{\sigma^2 T}{2a^4}n^{-1} + O(n^{-\frac{3}{2}}) \\ \frac{\partial \gamma^2}{\partial a^2} =& 1 + \frac{\sqrt{a^2\sigma^2 T}}{2a^2}n^{-\frac{1}{2}} + O(n^{-\frac{3}{2}})\end{aligned}$$

Therefore,

$$\begin{aligned}& V_2(\frac{\partial \omega}{\partial a^2}, \frac{\partial \omega}{\partial a^2}) \\ =& (\frac{\partial \eta}{\partial a^2})^2 V_2(\frac{\partial \omega}{\partial \eta}, \frac{\partial \omega}{\partial \eta}) + \frac{1}{\gamma^4}(\frac{\partial \gamma^2}{\partial a^2})^2 V_2(\omega, \omega) - \frac{1}{\gamma^2}\frac{\partial \eta}{\partial a^2}\frac{\partial \gamma^2}{\partial a^2}\frac{\partial V_2(\omega, \omega)}{\partial \eta} \\ =& \frac{2na_0^4}{a^8} + o(n)\end{aligned}$$

It follows from the similar calculation as above (see also Aït-Sahalia et al., 2005, pp 396) that

$$\begin{aligned}V_1(\frac{\partial \omega}{\partial \eta}, \frac{\partial \omega}{\partial \eta}) =& \frac{4n}{\gamma^4(1-\eta)^4}\text{cum}_4[U] + o(n) = \frac{n}{4a^4}\text{cum}_4[U] + o(n) \\ V_1(\omega, \omega) =& \frac{4n}{\gamma^4(1-\eta)^2}\text{cum}_4[U] + o(n) = \frac{n}{a^4}\text{cum}_4[U] + o(n) \\ \frac{\partial V_1(\omega, \omega)}{\partial \eta} =& \frac{8n}{\gamma^4(1-\eta)^3}\text{cum}_4[U] + o(n) = \frac{n}{a^4}\text{cum}_4[U] + o(n)\end{aligned}$$

Therefore,

$$\begin{aligned}V_1(\frac{\partial \omega}{\partial \sigma^2}, \frac{\partial \omega}{\partial \sigma^2}) &= O(1) \\ V_1(\frac{\partial \omega}{\partial a^2}, \frac{\partial \omega}{\partial a^2}) &= \frac{n\text{cum}_4[U]}{a^8} + o(n)\end{aligned}$$

Hence, combining the above equalities, we obtain

$$\begin{aligned}var(\epsilon' \frac{\partial \Omega}{\partial \sigma^2}\epsilon) &= V_1(\frac{\partial \omega}{\partial \sigma^2}, \frac{\partial \omega}{\partial \sigma^2}) + V_2(\frac{\partial \omega}{\partial \sigma^2}, \frac{\partial \omega}{\partial \sigma^2}) = \frac{\sqrt{T}a_0^4 n^{\frac{1}{2}}}{16a^5\sigma^3} + o(n^{\frac{1}{2}}) \\ var(\epsilon' \frac{\partial \Omega}{\partial a^2}\epsilon) &= V_1(\frac{\partial \omega}{\partial a^2}, \frac{\partial \omega}{\partial a^2}) + V_2(\frac{\partial \omega}{\partial a^2}, \frac{\partial \omega}{\partial a^2}) = \frac{n(2a_0^4 + \text{cum}_4[U])}{a^8} + o(n)\end{aligned}$$

This concludes the proof.

1.9.6 Proof of Theorem 6

We want to prove it by verifying the conditions of Theorem 2. It follows from Lemmas 1 and 2 that

$$\Psi_n^1 - \bar{\Psi}_n^1 = \frac{1}{2\sqrt{n}}\{O_p(1) + O_p(n^{\frac{1}{4}}) + O_p(n^{\frac{1}{4}}) + O_p(n^{\frac{1}{4}})\} = O_p(n^{-\frac{1}{4}})$$
$$\Psi_n^2 - \bar{\Psi}_n^2 = \frac{1}{2n}\{O_p(1) + O_p(n^{\frac{1}{4}}) + O_p(n^{\frac{1}{4}}) + O_p(n^{\frac{1}{2}})\} = O_p(n^{-\frac{1}{2}})$$

So far, we have shown that $|\Psi_n^i - \bar{\Psi}_n^i| \xrightarrow{\mathrm{P}} 0$, for any $\theta \in \Theta$, $i = 1, 2$. To prove uniform convergence in probability, we need stochastic equicontinuity of $\Psi_n^i - \bar{\Psi}_n^i$. As argued in Newey and McFadden (1994, pp 2133-2134) (see also Rockafellar (1970, Theorem 10.8)), the point-wise convergence of concave functions implies uniform convergence. Note that $\Psi_n^i - \bar{\Psi}_n^i$ can be regarded as the difference of two concave functions from (1.13), then a slightly modified proof of Theorem 10.8 in Rockafellar (1970), using triangle inequalities, still gives rise to uniform convergence. Therefore, as in Andersen and Gill (1982), we arrive at

$$\sup_{\theta\in\Theta} \|\Psi_n(\theta) - \bar{\Psi}_n(\theta)\| \xrightarrow{\mathrm{P}} 0.$$

Next, we show the identifiability condition holds.

$$\begin{aligned}\bar{\Psi}_n^2 =& \frac{1}{2n}\{tr(\Omega^{-1}\frac{\partial\Omega}{\partial a^2}) + tr(\frac{\partial\Omega^{-1}}{\partial a^2}\Sigma_0)\} \\ =& \frac{1}{n}\{(tr\Omega^{-1} - tr\Omega^{-1}J) + a_0^2(tr\frac{\partial\Omega^{-1}}{\partial a^2} - tr\frac{\partial\Omega^{-1}}{\partial a^2}J) + \frac{1}{2}\sum_{i=1}^{n}\frac{\partial\omega^{ii}}{\partial a^2}\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2 dt\}\end{aligned}$$

where $J = (J_{ij})$ where $J_{i-1,i} = 1$, and the other components of J is 0.

$$\begin{aligned}tr\Omega^{-1} - tr(\Omega^{-1}J) =& \frac{n(1+\eta)(1+\eta^{2n+1}) - \eta(1+\eta^{2n}) - \frac{2\eta^2(1-\eta^{2n-1})}{1-\eta}}{\gamma^2(1-\eta^2)(1-\eta^{2n+2})} \\ =& \frac{n}{2a^2} - \frac{\sqrt{Ta^2\sigma^2}}{4a^4}n^{\frac{1}{2}} + O(1) \qquad (1.35)\end{aligned}$$

So

$$tr\frac{\partial\Omega^{-1}}{\partial a^2} - tr(\frac{\partial\Omega^{-1}}{\partial a^2}J) = -\frac{n}{2a^4} + \frac{3\sqrt{a^2\sigma^2T}}{8a^6}n^{\frac{1}{2}} + O(1) \tag{1.36}$$

On the other hand, choose $K = n^{\frac{2}{3}}$, and let [7] $\omega^m = \omega^{K,K}$ and $\omega^M = \omega^{\frac{n+1}{2},\frac{n+1}{2}}$. By (1.29), we have

$$\begin{aligned}
1 \leq \frac{\omega^M}{\omega^m} =& \frac{(1-\eta^{n+1})(1-\eta^{n+1})}{(1-\eta^{2(n^{\frac{2}{3}})})(1-\eta^{2n-2(n^{\frac{2}{3}})+2})} \\
=& \frac{(1-e^{-\sqrt{\frac{\sigma^2T}{a^2}}n^{\frac{1}{2}}+O(n^{-\frac{1}{2}})})(1-e^{-\sqrt{\frac{\sigma^2T}{a^2}}n^{\frac{1}{2}}+O(n^{-\frac{1}{2}})})}{(1-e^{-\sqrt{\frac{\sigma^2T}{a^2}}n^{\frac{1}{6}}+O(n^{-\frac{5}{6}})})(1-e^{-2\sqrt{\frac{\sigma^2T}{a^2}}n^{\frac{1}{2}}+O(n^{-\frac{1}{2}})})} \\
\to& 1
\end{aligned}$$

Therefore, for any $K \leq i \leq n-K$, $\omega^{ii} = \omega^m(1+o(1))$; for $i < K$ or $i > n-K$, ω^{ii} is dominated by ω^m, and the integration is over an interval which shrinks at the rate $K/n = n^{-1/3}$, so

$$\sum_{i=1}^{n}\omega^{ii}\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2dt = \omega^m\int_0^T\sigma_t^2dt(1+o_p(1)) = \frac{n^{\frac{1}{2}}(a^2)^{-\frac{1}{2}}}{2\sqrt{\sigma^2T}}\int_0^T\sigma_t^2dt(1+o_p(1))$$

Similarly, for $K \leq i \leq n-K$,

$$\begin{aligned}
\frac{\partial\omega^{ii}}{\partial a^2} =& \frac{\partial\omega^{ii}}{\partial\eta}\frac{\partial\eta}{\partial a^2} - \frac{\omega^{ii}}{\gamma^2}\frac{\partial\gamma^2}{\partial a^2} = (\frac{2\eta}{1-\eta^2}\frac{\partial\eta}{\partial a^2} - \frac{1}{\gamma^2}\frac{\partial\gamma^2}{\partial a^2})\omega^{ii}(1+o(1)) \\
=& -\frac{1}{2a^2}\omega^{ii}(1+o(1))
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\omega^{ii}}{\partial\sigma^2} =& \frac{\partial\omega^{ii}}{\partial\eta}\frac{\partial\eta}{\partial\sigma^2} - \frac{\omega^{ii}}{\gamma^2}\frac{\partial\gamma^2}{\partial\sigma^2} = (\frac{2\eta}{1-\eta^2}\frac{\partial\eta}{\partial\sigma^2} - \frac{1}{\gamma^2}\frac{\partial\gamma^2}{\partial\sigma^2})\omega^{ii}(1+o(1)) \\
=& -\frac{1}{2\sigma^2}\omega^{ii}(1+o(1))
\end{aligned}$$

while for $i < K$ or $i > n-K$, $\frac{\partial\omega^{ii}}{\partial a^2}$ and $\frac{\partial\omega^{ii}}{\partial\sigma^2}$ are dominated by ω^{ii}. Therefore,

$$\sum_{i=1}^{n}\frac{\partial\omega^{ii}}{\partial a^2}\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2dt = -\frac{n^{\frac{1}{2}}(a^2)^{-\frac{3}{2}}}{4\sqrt{\sigma^2T}}\int_0^T\sigma_t^2dt(1+o_p(1)) \tag{1.37}$$

[7] Here and subsequently, $n^{\frac{2}{3}}$ can be replaced by $n^{\frac{1}{2}+\delta}$ with any $0 < \delta < \frac{1}{2}$.

and

$$\sum_{i=1}^{n}\frac{\partial\omega^{ii}}{\partial\sigma^2}\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2dt=-\frac{n^{\frac{1}{2}}(\sigma^2)^{-\frac{3}{2}}}{4\sqrt{a^2T}}\int_0^T\sigma_t^2dt(1+o_p(1)) \tag{1.38}$$

By calculation, we can also obtain

$$\bar{\Psi}_n^2=(\frac{1}{2a^2}-\frac{a_0^2}{2a^4})+(\frac{3a_0^2\sqrt{\sigma^2T}}{8a^5}-\frac{\sqrt{\sigma^2T}}{4a^3}-\frac{\int_0^T\sigma_t^2dt}{8a^3\sqrt{\sigma^2T}})n^{-\frac{1}{2}}+o_p(n^{-\frac{1}{2}}) \tag{1.39}$$

Hence, if we solve $\bar{\Psi}_n^2=0$, we get

$$a_n^{2*}=a_0^2+(\frac{3a_0^2\sqrt{\sigma_n^{2*}T}}{4a_n^*}-\frac{\sqrt{\sigma_n^{2*}T}a_n^*}{2}-\frac{a_n^*\int_0^T\sigma_t^2dt}{4\sqrt{\sigma_n^{2*}T}})n^{-\frac{1}{2}}+o_p(n^{-\frac{1}{2}}) \tag{1.40}$$

We assume all these likelihood estimates are bounded almost surely, since the parameter space is itself bounded. Also,

$$\begin{aligned}\frac{\partial\bar{\Psi}_n^2}{\partial\sigma^2}=&(\frac{\int_0^T\sigma_t^2dt}{16a^3\sqrt{T\sigma^6}}+\frac{3a_0^2}{16a^5\sqrt{\sigma^2T}}-\frac{T}{8a^3\sqrt{\sigma^2T}})n^{-\frac{1}{2}}+o_p(n^{-\frac{1}{2}})\\ \frac{\partial\bar{\Psi}_n^2}{\partial a^2}=&\frac{a_0^2}{a^6}-\frac{1}{2a^4}+(\frac{3\int_0^T\sigma_t^2dt}{16a^5\sqrt{\sigma^2T}}-\frac{15a_0^2\sqrt{\sigma^2T}}{16a^7}+\frac{3\sqrt{\sigma^2T}}{8a^5})n^{-\frac{1}{2}}+o_p(n^{-\frac{1}{2}})\end{aligned}$$

On the other hand, let $\Gamma:=\Sigma-\Omega$ be the diagram matrix with the i^{th} element of the diagonal $\Gamma_i:=\int_{\tau_{i-1}}^{\tau_i}\sigma_t^2dt-\sigma^2\Delta$, then

$$\begin{aligned}&\bar{\Psi}_n^1(\sigma^2,a^2)\\ =&\frac{1}{2\sqrt{n}}\{tr(\Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2})+\frac{\partial tr(\Omega^{-1}\Sigma_0)}{\partial\sigma^2}\}\\ =&\frac{1}{2\sqrt{n}}\{tr(\Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2})+\frac{\partial tr(\Omega^{-1}(\Omega+(2I-J-J')(a^2-a_0^2)+\Gamma))}{\partial\sigma^2}\}\\ =&\frac{1}{2\sqrt{n}}\{tr(\Omega^{-1}\frac{\partial\Omega}{\partial\sigma^2})+\sum_{i=1}^{n}\frac{\partial(\omega^{ii}\Gamma_i)}{\partial\sigma^2}\}-\frac{(a^2-a_0^2)T}{2a^2n^{\frac{3}{2}}}\sum_{i=1}^{n}(\omega^{ii}+\sigma^2\frac{\partial\omega^{ii}}{\partial\sigma^2})\\ =&\frac{1}{2\sqrt{n}}\{\Delta\sum_{i=1}^{n}\omega^{ii}+\sum_{i=1}^{n}\frac{\partial\omega^{ii}}{\partial\sigma^2}\Gamma_i-\Delta\sum_{i=1}^{n}\omega^{ii}\}-\frac{\sqrt{T}}{8a^3\sigma}(a^2-a_0^2)+O_p(n^{-\frac{1}{2}})\\ =&\frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\frac{\partial\omega^{ii}}{\partial\sigma^2}\Gamma_i-\frac{\sqrt{T}}{8a^3\sigma}(a^2-a_0^2)+O_p(n^{-\frac{1}{2}})\end{aligned}$$

$$= -\frac{1}{8a\sigma^3\sqrt{T}}(\int_0^T \sigma_t^2 dt - \sigma^2 T)(1+o(1)) - \frac{\sqrt{T}}{8a^3\sigma}(a^2 - a_0^2) + O_p(n^{-\frac{1}{2}}) \tag{1.41}$$

where the last equality is given by (1.38).

Set $\bar{\Psi}_n^1(\sigma^2, a^2) = 0$, that is,

$$\int_0^T \sigma_t^2 dt - \sigma_n^{2*} T = -\frac{\sigma_n^{2*} T}{a_n^{2*}}(a_n^{2*} - a_0^2) + O_p(n^{-\frac{1}{2}}) = O_p(n^{-\frac{1}{2}}) \tag{1.42}$$

Again, it follows from direct calculation that

$$\begin{aligned}
\frac{\partial \bar{\Psi}_n^1}{\partial \sigma^2} =& \frac{(-a_0^2 + a^2)\sqrt{T}}{16a^3\sigma^3} + \frac{\sqrt{T}}{8a\sigma^3} + \frac{3(\int_0^T \sigma_t^2 dt - T\sigma^2)}{16a\sigma^5\sqrt{T}} + o_p(1) \\
\frac{\partial \bar{\Psi}_n^1}{\partial a^2} =& \frac{3(-a_0^2 + a^2)\sqrt{T}}{16a^5\sigma} - \frac{\sqrt{T}}{8a^3\sigma} + \frac{\int_0^T \sigma_t^2 dt - T\sigma^2}{16a^3\sigma^3\sqrt{T}} + o_p(1)
\end{aligned}$$

As $n \to \infty$, $\theta_n^* \xrightarrow{\text{P}} \theta_0$, so with probability approaching 1,

$$\min_{\theta \in \Theta : \|\theta - \theta_n^*\| \geq \epsilon} \|\bar{\Psi}_n(\theta)\| \geq \min_{\theta \in \Theta : \|\theta - \theta_0\| \geq \frac{\epsilon}{2}} \|\bar{\Psi}_n(\theta)\|$$

Observe that,

$$\begin{aligned}
\|\bar{\Psi}_n(\theta)\| =& |\bar{\Psi}_n^1|^2 + |\bar{\Psi}_n^2|^2 \\
=& \frac{1}{4a^4}\{1 - \frac{a_0^2}{a^2} + (\frac{3a_0^2\sqrt{\sigma^2 T}}{8a^5} - \frac{\sqrt{\sigma^2 T}}{4a^3} - \frac{\int_0^T \sigma_t^2 dt}{8a^3\sqrt{\sigma^2 T}})n^{-\frac{1}{2}} + o(n^{-\frac{1}{2}})\}^2 \\
&+ \{\frac{1}{8a\sigma^3\sqrt{T}}(\int_0^T \sigma_t^2 dt - \sigma^2 T)(1+o(1)) + \frac{\sqrt{T}}{8a^3\sigma}(a^2 - a_0^2) + O(n^{-\frac{1}{2}})\}^2 \\
=& \frac{1}{4a^4}(1 - \frac{a_0^2}{a^2})^2 + \frac{T}{64a^2\sigma^2}\{\frac{1}{\sigma^2 T}(\int_0^T \sigma_t^2 dt - \sigma^2 T) + (1 - \frac{a_0^2}{a^2})\}^2 + o_p(1) \\
>& \epsilon_0 + o_p(1)
\end{aligned}$$

when $\|\theta - \theta_0\| = \sqrt{(\sigma^2 - \frac{1}{T}\int_0^T \sigma_t^2 dt)^2 + (a^2 - a_0^2)^2} > \frac{\epsilon}{2}$ and n goes to ∞. Hence,

$$\begin{aligned}
P(\min_{\theta \in \Theta : \|\theta - \theta_n^*\| \geq \epsilon} \|\bar{\Psi}_n(\theta)\| > \frac{\epsilon_0}{2}) >& P(\min_{\theta \in \Theta : \|\theta - \theta_0\| \geq \frac{\epsilon}{2}} \|\bar{\Psi}_n(\theta)\| > \frac{\epsilon_0}{2}) \\
>& P(-o_p(1) < \frac{\epsilon_0}{2})
\end{aligned}$$

$\to 1$

Therefore, it follows that the identifiability condition holds, hence by Theorem 2, $\hat{\sigma}^2 - \sigma_n^{2*} \xrightarrow{\text{P}} 0$, and $\hat{a}^2 - a_n^{2*} \xrightarrow{\text{P}} 0$. This concludes the proof by (1.40) and (1.42).

1.9.7 Proof of Lemma 3

The argument is very similar to the proof of Theorem 1 given in the appendix of Barndorff-Nielsen et al. (2008a), which involves the concept of stable convergence in law. The details about it have been discussed in Jacod and Shiryaev (2003) and Jacod (2007).

Assume for convenience that $\omega^{ij} = 0$, for $i, j < 1$ or $i, j > n$. Also, denote

$$\begin{aligned}
M_1^{(\beta)} &= \sum_{i=1}^{n} \frac{\partial \omega^{ii}}{\partial \beta} \{(\Delta_i^n X)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_t^2 dt\} \\
M_2^{(\beta)} &= \sum_{i=1}^{n} \sum_{j<i} \frac{\partial \omega^{ij}}{\partial \beta} \Delta_i^n X \Delta_j^n X \\
M_3^{(\beta)} &= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \beta} \epsilon_j \Delta_i^n X := \sum_{i=1}^{n} \iota_i^{(\beta)} \Delta_i^n X \\
M_4^{(\beta)} &= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \beta} (\epsilon_i \epsilon_j - E\epsilon_i \epsilon_j)
\end{aligned}$$

where $\Delta_i^n X = X_{\tau_i} - X_{\tau_{i-1}}$ and $\iota_i^{(\beta)} = \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \beta} \epsilon_j$.

We begin with $M_2^{(\sigma^2)}$. Pick $K = n^{\frac{2}{3}}$, and consider $K \le i \le n - K$. When $|i - j| = O(n^{\frac{1}{2}+\delta})$, $\omega^{ij} \to 0$ exponentially, for any $0 < \delta < \frac{1}{2}$. Rewrite it as:

$$n^{-\frac{3}{4}} M_2^{(\sigma^2)} \approx \Delta \sum_i f(\frac{\Delta_i^n X}{\sqrt{\Delta}}, \frac{\Delta_{i-1}^n X}{\sqrt{\Delta}}, ..., \frac{\Delta_{i-K}^n X}{\sqrt{\Delta}})$$

where

$$f(x_i, x_{i-1}, ..., x_{i-K}) = n^{-\frac{3}{4}} \sum_{h=1}^{K} \frac{\partial \omega^{i,i-h}}{\partial \sigma^2} x_i x_{i-h}$$

It follows from Theorem 7.1 in Jacod (2007) that, $\Delta^{-\frac{1}{2}} \cdot n^{-\frac{3}{4}} M_2^{(\sigma^2)}$ converges stably in law, and its asymptotic variance can be calculated using formula (7.2) and (7.3) in Jacod

(2007):

$$Avar = \lim_{n\to\infty} \sum_i \sigma_{\tau_i}^4 R_{i,n}(f)\Delta$$

where

$$\begin{aligned} R_{i,n}(f) =& n^{-\frac{3}{2}}\{E(\sum_{h=1}^{K} \frac{\partial \omega^{i,i-h}}{\partial \sigma^2} U_i U_{i-h})^2 - (E\sum_{h=1}^{K} \frac{\partial \omega^{i,i-h}}{\partial \sigma^2} U_i U_{i-h})^2\} \\ =& n^{-\frac{3}{2}} \sum_{i-K\le j<i} (\frac{\partial \omega^{ij}}{\partial \sigma^2})^2 \end{aligned}$$

and U_is are i.i.d. standard Gaussian random variables.

For any $K \le i \le n-K$, direct calculation yields

$$n^{-\frac{3}{2}} \sum_{i-K\le j<i} (\frac{\partial \omega^{ij}}{\partial \sigma^2})^2 \longrightarrow \frac{5}{64T^{\frac{3}{2}}\sigma^7 a}$$

By the same argument used in the proof of Theorem 6, we have

$$n^{-\frac{1}{4}} M_2^{(\sigma^2)} \xrightarrow{\mathcal{L}_X} MN(0, \frac{5}{64\sqrt{T}\sigma^7 a}\int_0^T \sigma_t^4 dt) \tag{1.43}$$

Since we have shown in Lemma 1 that $n^{-\frac{1}{4}} M_1^{(\sigma^2)} \xrightarrow{\mathrm{P}} 0$, it follows from Lemma 2 in Barndorff-Nielsen et al. (2008a) that

$$n^{-\frac{1}{4}}(M_1^{(\sigma^2)} + 2M_2^{(\sigma^2)}) \xrightarrow{\mathcal{L}_X} MN(0, \frac{5}{16\sqrt{T}\sigma^7 a}\int_0^T \sigma_t^4 dt) \tag{1.44}$$

As to $M_3^{(\sigma^2)}$, we at first notice that

$$E((\iota_i^{(\sigma)})^2 | \sigma(X)) = a_0^2 \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \sigma^2}(2\frac{\partial \omega^{ij}}{\partial \sigma^2} - \frac{\partial \omega^{i,j-1}}{\partial \sigma^2} - \frac{\partial \omega^{i,j+1}}{\partial \sigma^2})$$

and that for $K \leq i \leq n-K$,

$$n^{-\frac{1}{2}} \sum_{j=1}^{n} \frac{\partial \omega^{ij}}{\partial \sigma^2} (2\frac{\partial \omega^{ij}}{\partial \sigma^2} - \frac{\partial \omega^{i,j-1}}{\partial \sigma^2} - \frac{\partial \omega^{i,j+1}}{\partial \sigma^2}) \longrightarrow \frac{1}{8\sqrt{T}\sigma^5 a^3}$$

and that $\{(\sum_{i=1}^{n}(\Delta_i^n X)^2)^{-\frac{1}{2}}\Delta_i^n X\}$ become uniformly asymptotically negligible. Therefore, by the standard central limit theorem, conditional on the filtration $\sigma(X)$, we have

$$(\sum_{i=1}^{n}(\Delta_i^n X)^2)^{-\frac{1}{2}} n^{-\frac{1}{4}} M_3^{(\sigma^2)} \xrightarrow{\mathcal{L}} N(0, \frac{a_0^2}{8\sqrt{T}\sigma^5 a^3})$$

Since $\sum_{i=1}^{n}(\Delta_i^n X)^2 \xrightarrow{\mathrm{P}} \int_0^T \sigma_t^2 dt$, and by Lemma 1 and Proposition 5 in Barndorff-Nielsen et al. (2008a), we can conclude that

$$n^{-\frac{1}{4}} M_3^{(\sigma^2)} \xrightarrow{\mathcal{L}_X} N(0, \frac{a_0^2}{8\sqrt{T}\sigma^5 a^3} \int_0^T \sigma_t^2 dt) \tag{1.45}$$

and $n^{-\frac{1}{4}}(M_1^{(\sigma^2)} + 2M_2^{(\sigma^2)})$ and $n^{-\frac{1}{4}} M_3^{(\sigma^2)}$ jointly converge $\sigma(X)$-stably in law.

Also, as to the noise part, conditional on the filtration $\sigma(X)$,

$$n^{-\frac{1}{4}} M_4^{(\sigma^2)} \xrightarrow{\mathcal{L}} N(0, \frac{\sqrt{T}a_0^4}{16a^5\sigma^3}) \tag{1.46}$$

The same reasoning as above yields

$$\begin{aligned} & n^{\frac{1}{4}}(\Psi_n^1 - \bar{\Psi}_n^1) \\ =& n^{-\frac{1}{4}}(M_1^{(\sigma^2)} + 2M_2^{(\sigma^2)} + M_3^{(\sigma^2)} + M_4^{(\sigma^2)}) \\ \xrightarrow{\mathcal{L}_X}& MN(0, \frac{5}{16\sqrt{T}\sigma^7 a} \int_0^T \sigma_t^4 dt + \frac{a_0^2}{8\sqrt{T}\sigma^5 a^3} \int_0^T \sigma_t^2 dt + \frac{\sqrt{T}a_0^4}{16a^5\sigma^3}) \end{aligned} \tag{1.47}$$

As to $\Psi_n^2 - \bar{\Psi}_n^2$, since the noise part dominates, it implies that

$$n^{\frac{1}{2}}(\Psi_n^2 - \bar{\Psi}_n^2) \xrightarrow{\mathcal{L}} MN(0, \frac{2a_0^4 + \mathrm{cum}_4[U]}{4a^8}) \tag{1.48}$$

Finally, the covariance of the $\Psi_n^1 - \bar{\Psi}_n^1$ and $\Psi_n^2 - \bar{\Psi}_n^2$ is of the order $O(n^{-1})$, thus the joint central limit theorem follows from Proposition 5 in Barndorff-Nielsen et al. (2008a).

1.9.8 Proof of Theorem 7

Now we derive the central limit theorem of the estimators. Because

$$\Psi_n^i(\omega,\hat{\theta}) - \Psi_n^i(\omega,\theta_n^*) = \nabla\Psi_n^i(\omega,\tilde{\theta}_i)(\hat{\theta}-\theta_n^*)$$

where $\tilde{\theta}_i$ is between $\hat{\theta}$ and θ_n^*, and by Lemma 1 in Barndorff-Nielsen et al. (2008a), it is clear that we only need to verify the assumptions of Theorem 3. In fact, because of (1.13), we will consider $Y'\Omega Y$ and $tr(\Omega\Sigma_0)$ and their derivatives multiplied by proper normalizations. It is straightforward to check that these functions are either convex or concave, which guarantees stochastic equicontinuity (see the proof of Theorem 6).

Combining Theorem 6, we have

$$\begin{aligned}
&\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}^2-\sigma_n^{2*}) \\ n^{\frac{1}{2}}(\hat{a}^2-a_n^{2*}) \end{pmatrix} \\
&= \begin{pmatrix} n^{\frac{1}{4}} & 0 \\ 0 & n^{\frac{1}{2}} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{T}}{8a_n^*\sigma_n^{3*}} & -\frac{\sqrt{T}}{8a^{3*}\sigma^*} \\ 0 & \frac{a_0^2}{a^{6*}} - \frac{1}{2a^{4*}} \end{pmatrix}^{-1} \begin{pmatrix} \Psi_n^1(\omega,\theta_n^*) - \bar{\Psi}_n^1(\omega,\theta_n^*) \\ \Psi_n^2(\omega,\theta_n^*) - \bar{\Psi}_n^2(\omega,\theta_n^*) \end{pmatrix} + \begin{pmatrix} o_p(1) \\ o_p(1) \end{pmatrix} \\
&\xrightarrow{\mathcal{L}_X} MN\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{5a_0\int_0^T \sigma_t^4 dt}{T(\int_0^T \sigma_t^2 dt)^{\frac{1}{2}}} + \frac{3a_0(\int_0^T \sigma_t^2 dt)^{\frac{3}{2}}}{T^2} & 0 \\ 0 & 2a_0^4 + \text{cum}_4[U] \end{pmatrix} \Big)
\end{aligned}$$

Note from (1.40) and (1.42) that because of the consistency results above, $a_n^{2*} - a_0^2 = o_p(n^{-\frac{1}{2}})$, $\sigma_n^{2*} - \frac{1}{T}\int_0^T \sigma_t^2 dt = O_p(n^{-\frac{1}{2}})$, therefore, the potential asymptotic biases are eliminated, which concludes the proof.

1.9.9 Proof of Theorem 8

From the likelihood function (1.2), we can obtain the score functions:

$$\begin{aligned}
tr(\Omega^{-1}\Lambda) - tr(Y'\Omega^{-1}\Lambda\Omega^{-1}Y) = 0 \\
tr(\Omega^{-1}) - tr(Y'\Omega^{-2}Y) = 0
\end{aligned}$$

where $a^2\Lambda = \Omega - \sigma^2\Delta I$.

On the other hand, the following equalities hold:

$$
\begin{aligned}
a^2 tr(\Omega^{-1}\Lambda) &= tr(\Omega^{-1}(\Omega - \sigma^2\Delta I)) = n - \sigma^2\Delta tr(\Omega^{-1}) \\
a^2 tr(\Omega^{-2}\Lambda) &= tr(\Omega^{-2}(\Omega - \sigma^2\Delta I)) = tr(\Omega^{-1}) - \sigma^2\Delta tr(\Omega^{-2}) \\
a^4 tr(\Omega^{-2}\Lambda^2) &= n - 2\sigma^2\Delta tr(\Omega^{-1}) + \sigma^4\Delta^2 tr(\Omega^{-2})
\end{aligned}
$$

As a result,

$$
\begin{aligned}
\sigma^2\Delta &= \frac{tr(\Omega^{-2}\Lambda)tr(\Lambda\Omega^{-1}) - tr(\Omega^{-2}\Lambda^2)tr(\Omega^{-1})}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2})\cdot tr(\Omega^{-2}\Lambda^2)} \\
a^2 &= \frac{tr(\Omega^{-2}\Lambda)tr(\Omega^{-1}) - tr(\Omega^{-2})tr(\Omega^{-1}\Lambda)}{(tr(\Omega^{-2}\Lambda))^2 - tr(\Omega^{-2})\cdot tr(\Omega^{-2}\Lambda^2)}
\end{aligned}
$$

Plugging in the score functions gives the representation. Clearly, the two quadratic forms are not co-linear. The second claim is obvious. In fact, we can write $\Omega = \sigma^2 T\tilde{\Omega}$, and $\tilde{\Omega}$ only depends on λ. Plugging it into the representation is amount to replacing Ω with $\tilde{\Omega}$ directly.

1.9.10 Proof of Theorem 9

According to (1.29), we have that for $K \le i, j \le n - K$,

$$
(\Omega^{-1})_{i,j} = \omega^{ij} \approx \frac{(-\eta)^{|i-j|}}{\gamma^2(1-\eta^2)}
$$

with the difference exponentially small. Therefore, we can further deduce that

$$
\begin{aligned}
(\Omega^{-2})_{i,j} =& \sum_l \omega^{il}\omega^{lj} \\
\approx& \frac{(-\eta)^{|i-j|}}{\gamma^4(1-\eta^2)^2}\Big(\frac{2\eta^2}{1-\eta^2} + |i-j|\Big) \\
\approx& \frac{n(-\eta)^{|i-j|}}{4a^4}\frac{a^2}{\sigma^2 T}\Big(\frac{a}{\sigma T^{\frac{1}{2}}}n^{\frac{1}{2}} + |i-j|\Big)
\end{aligned}
$$

$$
\begin{aligned}
(\Omega^{-1}\Lambda\Omega^{-1})_{i,j} =& \frac{1}{a^2}\Big((\Omega^{-1})_{i,j} - \sigma^2\Delta(\Omega^{-2})_{i,j}\Big) \\
\approx& \frac{(-\eta)^{|i-j|}}{4a^4}\Big(\frac{a}{\sigma T^{\frac{1}{2}}}n^{\frac{1}{2}} - |i-j|\Big)
\end{aligned}
$$

On the other hand, direct calculations deduce that

$$\begin{aligned} tr(\Omega^{-2}) =& \frac{n^{\frac{5}{2}}}{4a\sigma^3 T^{\frac{3}{2}}} + O(n^2) \\ tr(\Omega^{-2}\Lambda) =& \frac{n^{\frac{3}{2}}}{4a^3\sigma\sqrt{T}} + O(n) \\ tr(\Omega^{-2}\Lambda^2) =& \frac{n}{a^4} + O(n^{\frac{1}{2}}) \end{aligned}$$

Therefore, it follows from the representation that

$$W_{1,i,j} \approx (1 + \frac{|i-j|}{\lambda \cdot n^{\frac{1}{2}}})(-\eta)^{|i-j|} \approx (1 + \frac{|i-j|}{\lambda \cdot n^{\frac{1}{2}}})e^{-\lambda^{-1} \cdot n^{-\frac{1}{2}}|i-j|}$$

CHAPTER 2

HIGH-FREQUENCY COVARIANCE ESTIMATES WITH NOISY AND ASYNCHRONOUS DATA

This chapter proposes a consistent and efficient estimator of the high frequency covariance of two arbitrary assets, observed asynchronously with market microstructure noise. This estimator is built upon the marriage of the quasi-maximum likelihood estimator of the quadratic variation and the proposed Generalized Synchronization scheme. It is therefore not influenced by the Epps effect. Moreover, the estimation procedure is free of tuning parameters or bandwidths and readily implementable. Monte Carlo simulations show the advantage of this estimator by comparing it with a variety of estimators with specific synchronization methods. The empirical studies of six foreign exchange future contracts illustrate the time-varying correlations of the currencies during the global financial crisis in 2008, discovering the similarities and differences in their roles as key currencies in the global market. Another empirical application with stock and oil futures is provided to show some short-term abnormal correlation pattern that may not be captured with lower frequency observations. This chapter is published in Aït-Sahalia et al. (2010b), and presented at the 2010 Econometrics Society World Congress and the 2009 Joint Statistical Meetings.

2.1 Introduction

The covariation between asset returns plays a crucial role in modern finance. For instance, the covariance matrix and its inverse are the key statistics in portfolio optimization and risk management. Many recent financial innovations involve complex derivatives, like exotic options written on the minimum, maximum or difference of two assets, or some structured financial products, such as CDOs. All of these innovations are built upon, or in order to exploit, the correlation structure of two or more assets. As technological developments make high frequency data commonly available, much effort has been put into developing statistical inference methodologies for continuous time models with intra-day data, enabling us to capture the daily variation of some interesting statistics that were otherwise unobservable from daily or weekly data.

Realized variance estimation is an example of such statistics. Unfortunately, unlike those low frequency time series that are homogeneously spaced, tick-by-tick transactions of different assets usually occur randomly and asynchronously; in addition, with high frequency data comes market microstructure noise. These factors make it difficult to employ a Realized Covariance (RC) estimator directly. Popular estimators in the univariate variance case include Two Scales Realized Volatility (TSRV) of Zhang et al. (2005), the first consistent estimator for integrated volatility in the presence of noise, Multi-Scale Realized Volatility (MSRV), a modification of TSRV which achieves the best possible rate of convergence proposed by Zhang (2006), Realized Kernels (RK) by Barndorff-Nielsen et al. (2008a) and the Pre-Averaging (PA) approach by Jacod et al. (2009b), both of which contain sets of nonparametric estimators that can also achieve the best convergence rate. In contrast with these nonparametric estimators, Xiu (2010) has extended the Maximum-Likelihood Estimator of Aït-Sahalia et al. (2005) designed for parametric volatility to a Quasi-Maximum Likelihood Estimator (QMLE), as a misspecified maximum likelihood estimator in the setting of stochastic volatility. This parametric estimator has also proved to be consistent and efficient without any tuning parameters. Related work includes Bandi and Russell (2003), Delattre and Jacod (1997), Fan and Wang (2007), Gatheral and Oomen (2009), Hansen and Lunde (2006), Kalnina and Linton (2008), Li and Mykland (2007), Aït-Sahalia et al. (2009),

Zhang et al. (2009) and Li et al. (2010).

These advances in variance estimation pave the way for the efficient estimation of covariance with noisy data. Studies of correlation estimated from asynchronous high frequency stock price returns date back to at least Epps (1979), who documented the Epps effect, i.e. the fact that the sample correlation tends to have a strong bias towards zero as the sampling interval progressively shrinks. The same effect has been documented for exchange rates, see, e.g., Guillaume et al. (1997) and Muthuswamy et al. (2001). Since then, researchers have been trying to resolve this puzzle. There are at least two possible approaches. First, dealing with asynchronous data can be achieved with subsampling, using previous tick or other interpolation methods. This procedure may induce a potential bias. Hayashi and Yoshida (2005) have proposed a modification of the Realized Covariance (RC) estimator (HY), which is consistent and immune to this bias. Also, it is reasonable to conjecture that microstructure noise might be at least partly responsible for the documented bias as well, based on the fact that the magnitude of the noise relative to that of the price signal will increase the realized variance estimator, which serves as the denominator in the correlation calculation: see Large (2007) and Griffin and Oomen (2008), for example. Voev and Lunde (2007) provided a bias correction procedure for the HY estimator to correct for the effect of the noise, but the estimator does not achieve consistency. Zhang (2009) has demonstrated theoretically that there may be a bias associated with the RC estimator with the previous tick method due to asynchronocity of the data. Zhang (2009) has put forward a consistent Two Scales Realized Covariance estimator (TSCV) using the previous tick method, which is capable of dealing with asynchronous and noisy data. Recently, Barndorff-Nielsen et al. (2008b) suggested to synchronize the high frequency prices using a Refresh Time scheme, and implemented Multivariate Realized Kernels (MRK) to provide a consistent and semi-definite estimator of the covariance matrix, while Kinnebrock and Podolskij (2008) proposed a multivariate PA estimator. These nonparametric estimators involve the selection of bandwidths or other tuning parameters, and implementation can therefore raise subtle issues in practice.

This chapter proposes another consistent and rate-efficient estimator based on the QMLE and generalized time synchronization method. Unlike some of the alternatives, it involves

no tuning parameters to be set and is very easy to implement. The chapter is organized as follows. Section 2.2 details the proposed covariance estimator and the synchronization method. Section 2.4 includes Monte Carlo simulations to compare different covariance estimators. Section 2.5 provides a detailed empirical study. Section 2.7 concludes. The appendix contains the mathematical proofs.

2.2 Covariance and Correlation Estimation via the QMLE

2.2.1 Model Setup

We now extend the previous results to incorporate a two-dimensional log-price process $\boldsymbol{X}_t = (X_{1t}, X_{2t})$, discretely observed over the interval $[0, T]$. Suppose that the observations are recorded at times $0 = t_{i,0} \leq t_{i,1} \leq t_{i,2} \leq \ldots \leq t_{i,n_i} = T$ respectively, where $i = 1, 2$. As in the univariate case, one can only observe $\tilde{X}_{i,t}$, contaminated by an additive error $U_{i,t}$, associated at each observation point. Further, we make the following assumption:

Assumption 10. *The latent log-price process satisfies*

$$dX_{it} = \mu_{it}dt + \sigma_{it}dW_{it}$$

with $E(dW_{1t} \cdot dW_{2t}) = \rho_t dt$, and the volatility processes positive and locally bounded Itô semi-martingales, and the drifts locally bounded and progressively measurable processes. The noise $\boldsymbol{U}_t$ is an i.i.d. 2-dimensional vector with mean 0, diagonal covariance matrix $\boldsymbol{\Theta}$ and has a finite fourth moment.

2.2.2 Covariance and Correlation Estimation for Synchronized Data

The estimator is based on the following identity:

$$Cov(X_1, X_2) = \frac{1}{4}\Big(Var(X_1 + X_2) - Var(X_1 - X_2)\Big)$$

Therefore, we propose

$$\widehat{Cov}(\tilde{X}_1, \tilde{X}_2) = \frac{1}{4}\Big(\widehat{Var}(\tilde{X}_1 + \tilde{X}_2) - \widehat{Var}(\tilde{X}_1 - \tilde{X}_2)\Big) \tag{2.1}$$

where $\widehat{Cov}(\cdot,\cdot)$ is our estimator, $\widehat{Var}(\cdot,\cdot)$ denotes the QMLE of the quadratic variation, and $\cdot$ indicates the data we are actually using. However, in order to compute the prices $\tilde{X}_1+\tilde{X}_2$ and $\tilde{X}_1 - \tilde{X}_2$, we need the two assets be synchronically traded. We will discuss in details the next subsection how to deal with nonsynchronized trading.

Straightforwardly, the correlation estimator is given by:

$$\widehat{Corr}(\tilde{X}_1, \tilde{X}_2) = \frac{\widehat{Cov}(\tilde{X}_1, \tilde{X}_2)}{\sqrt{\widehat{Var}(\tilde{X}_1)}\sqrt{\widehat{Var}(\tilde{X}_2)}}. \tag{2.2}$$

Remark. *It may be more efficient to consider linear combinations*

$$\widehat{Cov}(\tilde{X}_1, \tilde{X}_2) = \frac{1}{4\gamma(1-\gamma)}\Big(\widehat{Var}(\gamma\tilde{X}_1 + (1-\gamma)\tilde{X}_2) - \widehat{Var}(\gamma\tilde{X}_1 - (1-\gamma)\tilde{X}_2)\Big) \tag{2.3}$$

where γ can be selected to minimize the asymptotic variance. If $\gamma = 1/2$, (2.3) recurs to (2.1). One alternative γ may be $\widehat{Var}(\tilde{X}_2)/\big(\widehat{Var}(\tilde{X}_1) + \widehat{Var}(\tilde{X}_2)\big)$.

The constructed covariance matrix estimates may not be positive semi-definite. This property is nevertheless essential to many applications in practice. To enforce it, one possibility consists in projecting the resulting symmetric matrix onto the space of positive semi-definite matrices: see for example an application to portfolio allocation in Fan et al. (2010).

In Section 1.3, we have shown consistency and a central limit theorem for the QMLE of the variance. Based on Theorem 7, we can show that this estimator is $n^{1/4}$-consistent for the covariance and correlation. This rate is the optimal one. The following theorem provides the central limit results for the covariance estimator under an idealized data observation scheme where the two assets are observed at synchronized times:

Theorem 11. *Given Assumption 10, and that the data are synchronized and equally spaced, that is, $n := n_1 = n_2$ and $\tau_j := t_{1,j} = t_{2,j}$, and that $\Delta = \tau_j - \tau_{j-1}$, for $1 \le j \le n$, then the following*

Central Limit Theorem holds:

$$n^{\frac{1}{4}}\Big(\widehat{Cov}(\tilde{X}_1,\tilde{X}_2)-\frac{1}{T}\int_0^T\rho_t\sigma_{1t}\sigma_{2t}dt\Big)\xrightarrow{\mathcal{L}_{\mathrm{X}}}MN(0,V) \tag{2.4}$$

$$n^{\frac{1}{4}}\Big(\widehat{Corr}(\tilde{X}_1,\tilde{X}_2)-\frac{\int_0^T\rho_t\sigma_{1t}\sigma_{2t}dt}{\sqrt{\int_0^T\sigma_{1t}^2dt}\sqrt{\int_0^T\sigma_{2t}^2dt}}\Big)\xrightarrow{\mathcal{L}_{\mathrm{X}}}MN(0,\tilde{V}) \tag{2.5}$$

where V and $\tilde{V}$ are given in the appendix.

2.3 Synchronization Scheme

2.3.1 Generalized Synchronization Method

We would stop here if the data were synchronized, meaning that the prices of the two assets were observed at the same times. However, this is not the case in practice, at least for high frequency financial data. In most cases, high frequency transactions for two assets occur at times that are not synchrone. This practical issue may induce a large bias for the estimation, and may be (at least partly) responsible for the Epps effect. We note further that both terms on the right side of (2.1) need synchronized data. The remaining question is what kind of data synchronization procedure one should use. Clearly, if we apply the QMLE to estimate the diagonal elements in the covariance matrix, it would be better to cross out a small number of data points rather than adding more through an interpolation method, because the former strategy may suffer from efficiency loss, while the latter one may result in inconsistency due to the change in the autocorrelation structure. We define the following concept, which we then use to propose a general synchronization scheme.

Definition 1. *A sequence of time points $\{\tau_0,\tau_1,\tau_2,...,\tau_n\}$ is said to be the Generalized Sampling Time for a collection of M assets, if*

1. $0=\tau_0<\tau_1<...<\tau_{n-1}<\tau_n=T$.
2. *There exists at least one observation for each asset between consecutive τ_is.*
3. *The time intervals, $\{\Delta_j=\tau_j-\tau_{j-1},1\le j\le n\}$, satisfies $\sup_i\Delta_i\xrightarrow{\mathrm{P}}0$.*

The Generalized Synchronization method is built upon the Generalized Sampling Time by selecting an arbitrary observation $\tilde{X}_{i,\check{t}_j}$ for the ith asset between the time interval $(\tau_{j-1}, \tau_j]$. The synchronized data sets are, therefore, $\{\tilde{X}^{\tau}_{i,\tau_j}, 1 \leq i \leq M, 1 \leq j \leq n\}$ such that $\tilde{X}^{\tau}_{i,\tau_j} = \tilde{X}_{i,\check{t}_j}$.

The concept of Generalized Synchronization method is more general than that of the Previous Tick approach discussed in Zhang (2009), and the Refresh Time scheme proposed by Barndorff-Nielsen et al. (2008b), namely, the Replace All scheme in deB. Harris et al. (1995).

More precisely, if we require $\{\tau_j\}$ to be equally spaced between $[0, T]$, the previous tick for each asset before τ_j to be selected, we are back to the Previous Tick approach. Or, if we choose τ_j recursively as

$$\tau_{j+1} = \max_{1 \leq i \leq M} \{t_{i,N_i(\tau_j)+1}\}$$

where $\tau_1 = \max\{t_{1,1}, t_{2,1}, ..., t_{M,1}\}$ and $N_i(t)$ measures the number of observations for asset i before time t, and if we select those ticks that occur right before or at τ_js, we return to the Refresh Time scheme. In both cases, the previous ticks of the assets, if needed, are regarded as if they were observed at the sampling time τ_js. By contrast, we advocate choosing an arbitrary tick for each asset within each interval. In practice, it may happen that the order of consecutive ticks is not recorded correctly. Because our synchronization method has no requirement on tick selection, the estimator is robust to data misplacement error, as long as these misplaced data points are within the same sampling intervals.

It is apparent that the Refresh Time scheme is highly dependent on the relatively illiquid asset. On the one hand, the number of the synchronized pairs are smaller than the number of the observations of this asset, inducing an inevitable loss of data for the other asset. More importantly, it is very likely that the Refresh Time points are determined by the occurrence of the relatively more illiquid asset, rendering the selected observations of the other asset always ahead of the corresponding illiquid asset. This hidden effect may induce some additional bias in the estimation.

Alternatively, we can design the synchronization scheme requiring each asset to lead in turn. Take two assets, for example. If we require the first asset to lead, we choose

$\tau_1 = t_{2,N_2(t_{1,1})+1}$. Recursively,

$$\tau_i = t_{2,N_2(t_{1,N_1(\tau_{i-1})+1})+1}$$

Literally, it means that right after τ_{i-1}, we find the first observation of $\tilde{X}_{1t}$, which should happen at $t_{1,N_1(\tau_{i-1})+1}$, and then the next Generalized Sampling Time is defined to be the point when the first $\tilde{X}_{2t}$ is observed right after $t_{1,N_1(\tau_{i-1})+1}$. In this case, at all sampling time points, the second asset would always have records. The previous tick of the first asset, if needed, is regarded as if it were observed a bit later at the sampling time. Hence, in the synchronized pairs, the first asset always leads the second one.

Remark. *Refresh Time includes the largest amount of data among all Generalized Sampling Time.*

Combining the synchronization scheme with our estimator, we make the following assumption:

Assumption 12. *The Generalized Sampling Time $\{\tau_j\}$ is independent of the price process, the volatility process and the noise. The time intervals, $\{\Delta_j = \tau_j - \tau_{j-1}, 1 \leq j \leq n\}$, are i.i.d. with mean $\bar{\Delta}$. The number of observations n is therefore random, of order $O_P(1/\bar{\Delta})$.*

Replacing the idealized data with the products of the Generalized Sampling Time, we obtain:

Theorem 13. *Given Assumptions 10 and 12, the QMLE of the quadratic variation for $\tilde{X}_1^\tau + \tilde{X}_2^\tau$ is consistent, that is,*

$$\widehat{Var}(\tilde{X}_1^\tau + \tilde{X}_2^\tau) - \frac{1}{T}\int_0^T \sigma_{1t}^2 + \sigma_{2t}^2 + 2\rho_t\sigma_{1t}\sigma_{2t}dt = O_P(\bar{\Delta}^{\frac{1}{4}})$$

Therefore,

$$\widehat{Cov}(\tilde{X}_1^\tau, \tilde{X}_2^\tau) - \frac{1}{T}\int_0^T \rho_t\sigma_{1t}\sigma_{2t}dt = O_P(\bar{\Delta}^{\frac{1}{4}})$$
$$\widehat{Corr}(\tilde{X}_1^\tau, \tilde{X}_2^\tau) - \frac{\int_0^T \rho_t\sigma_{1t}\sigma_{2t}dt}{\sqrt{\int_0^T \sigma_{1t}^2 dt}\sqrt{\int_0^T \sigma_{2t}^2 dt}} = O_P(\bar{\Delta}^{\frac{1}{4}})$$

In other words, the rate of convergence of the estimators are the same as those given in Theorem 11.

2.3.2 Synchronization Comparison with the HY Estimator

It is interesting to compare the synchronization method embedded in the HY estimator and the synchronization scheme proposed here. Recall that Hayashi and Yoshida (2005) proposed

$$\langle X_1, X_2 \rangle_{HY} = \sum_{i,j} (X_{1,t_{1,j}} - X_{1,t_{1,j-1}})(Y_{1,t_{2,i}} - Y_{1,t_{2,i-1}}) 1_{\{(t_{1,j-1},t_{1,j}] \cap (t_{2,i-1},t_{2,i}] \neq \emptyset\}} \tag{2.6}$$

where X_1 and X_2 are the observations, in an estimator that assumes no noise and is therefore infeasible in our setting. Nevertheless, the synchronization method proposed therein may be applicable. A priori, the HY method has advantages over the Refresh Time scheme in that the former utilizes all possible data. However, this ignores the fact that the HY method effectively deletes data through some cancelation in the calculation as well. If we fix index j and sum over i at first, we can rewrite the formula (2.6) in the following way:

$$\langle X_1, X_2 \rangle_{HY} = \sum_{j} (X_{1,t_{1,j}} - X_{1,t_{1,j-1}})(Y_{1,t_{2,j+}} - Y_{1,t_{2,(j-1)-}}) \tag{2.7}$$

where $t_{2,j+} = \inf\{t_{2,k} : t_{2,k} \geq t_{1,j}\}$, and $t_{2,j-} = \sup\{t_{2,k} : t_{2,k} \leq t_{1,j}\}$.

It follows from (2.7) and Figure 2.1 that at least any records for the second asset that occur in $(t_{2,(j-1)+}, t_{2,j-})$ will not play a role in the calculation due to cancelation. Similarly, if three consecutive observations of the first asset form two intervals which share the same corresponding interval of the second asset, then the middle observation of the first asset will not be used either. In the simulation and empirical studies, we will compare the effective sample size of the HY method and the Refresh Time scheme.

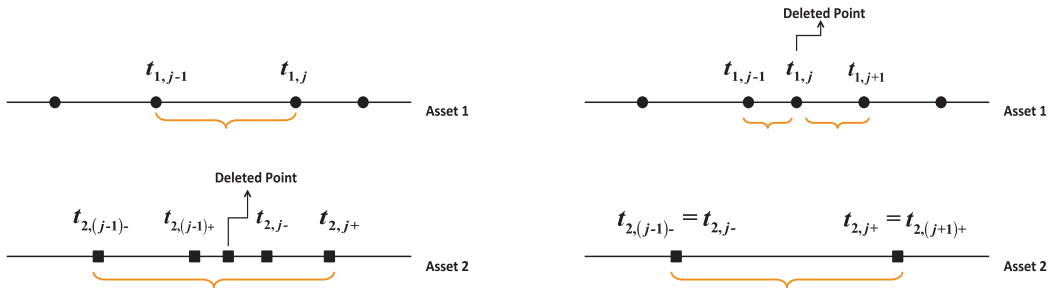


Figure 2.1: These two graphs illustrate how the HY synchronization method delete data points.

2.4 Simulation Results and Comparisons

We now conduct Monte Carlo simulations to verify the applicability of the above theoretical results in realistic settings. The data is generated using stochastic volatility models, and for this purpose we consider mainly the Heston Model. For $i = 1, 2$, the data generating process is

$$dX_{it} = \sigma_{it} dW_{it}$$
$$d\sigma_{it}^2 = \kappa_i(\bar{\sigma_i}^2 - \sigma_{it}^2) + s_i \sigma_{it} dB_{it} + \sigma_{it-} J_{it}^V dN_{it}$$

where $E(dW_{it} \cdot dB_{jt}) = \delta_{ij}\rho_i dt$ and $E(dW_{1t} \cdot dW_{2t}) = \rho dt$.

We generate sample paths using the Euler method, where the first observation for volatility process σ_{i0}^2 is sampled from a Gamma distribution $\Gamma(2\kappa_i\bar{\sigma_i}^2/s_i^2,\ s_i^2/2\kappa_i)$. The jump size in volatility is $J_t^V = \exp(z)$, where $z \sim N(\theta_i, \mu_i)$, and N_{it} is a Poisson Processes independent of Brownian motions with intensity λ_i. The noise of each asset is i.i.d. $N(0, a_i^2)$. The parameter values are reported in Table 2.1.

First, we generate N random intervals from the exponential distribution for both processes. The data are equally spaced and therefore, there are no issues with synchronization and we can easily estimate the true covariance using the RC method. Then we ran-

Table 2.1: Summary of the Parameter Values in the Monte Carlo Simulation

Asset	X_{i0}	κ_i	s_i	$\bar{\sigma}_i^2$	ρ_i	λ_i	θ_i	μ_i	a_i	ρ
$i=1$	$\log(100)$	6	0.5	0.16	-0.6	12	-5	0.8	0.005	0.5
$i=2$	$\log(40)$	4	0.3	0.09	-0.75	36	-6	1.2	0.001	$\cdot$

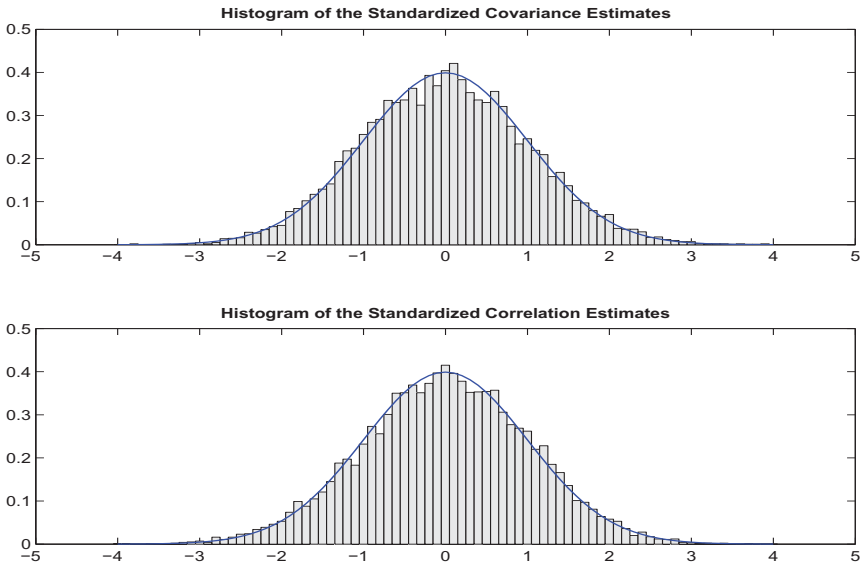


Figure 2.2: In this figure, we plot the histograms of the covariance and correlation estimator. Here, the data are synchronized, and the average length of the sampling intervals is $\overline{\Delta} = 1$s.

domly select about $N/2$ observations for X_1 and $2N/3$ for X_2 from the whole sample using Bernoulli trials, and contaminate the data with noise. Next we synchronize the data using the Refresh Time procedure and apply the QMLE estimator to $\tilde{X}_1^\tau \pm \tilde{X}_2^\tau$, $\tilde{X}_1^\tau$ and $\tilde{X}_2^\tau$ respectively, and construct the estimator for the covariance. We also include a modified version QMLE* as defined in (2.3).

Alternatively, we also implement the TSCV estimator by Zhang (2009) for comparison. Suppose the number of subsamples are K and J respectively. Define

$$[\tilde{X}_1^\tau, \tilde{X}_2^\tau]_T^S = \frac{1}{S}\sum_{i=S}^{n}(\tilde{X}_{1,\tau_i}^\tau - \tilde{X}_{1,\tau_{i-S}}^\tau)(\tilde{X}_{2,\tau_i}^\tau - \tilde{X}_{2,\tau_{i-S}}^\tau)$$

and then the TSCV estimator is given by

$$\langle \tilde{X}_1^\tau, \tilde{X}_2^\tau \rangle_{TSCV} = ([\tilde{X}_1^\tau, \tilde{X}_2^\tau]_T^K - \frac{\bar{n}_K}{\bar{n}_J}[\tilde{X}_1^\tau, \tilde{X}_2^\tau]_T^J)$$

where $1 \leq J << K = O(n^{\frac{2}{3}})$, and $\bar{n}_S = (n - S + 1)/S$, $S = K, J$.

We also implement the MRK proposed in Barndorff-Nielsen et al. (2008b) for comparison. Suppose we have n number of observations for each asset after synchronization. First we redefine the initial and final time points to control the edge effect associated with realized kernels. Fix $m \in \mathbb{N}$, with $N = n + 1 - 2m$. Define $\tilde{\boldsymbol{X}}_{\tau_j} = (\tilde{X}_{1,\tau_j}^\tau, \tilde{X}_{2,\tau_j}^\tau)'$.

$$\check{\boldsymbol{X}}_0 = \frac{1}{m}\sum_{j=1}^{m}\tilde{\boldsymbol{X}}_{\tau_j}$$
$$\check{\boldsymbol{X}}_N = \frac{1}{m}\sum_{j=1}^{m}\tilde{\boldsymbol{X}}_{\tau_{n-m+j}}$$

And $\check{\boldsymbol{X}}_j = \tilde{\boldsymbol{X}}_{\tau_{j+m}}$, where $j = 1, 2, ..., N-1$. Let $\boldsymbol{x}_j = \check{\boldsymbol{X}}_j - \check{\boldsymbol{X}}_{j-1}$, $j = 1, 2, ..., N$.

The multivariate realized kernel estimator is

$$\boldsymbol{K}(X) = \sum_{h=-H}^{H} k(\frac{|h|}{H+1})\boldsymbol{\Gamma}_h$$

where

$$\mathbf{\Gamma}_h = \begin{cases} \sum_{j=|h|+1}^{N} \boldsymbol{x}_j \boldsymbol{x}'_{j-h}, & h \geq 0; \\ \sum_{j=|h|+1}^{N} \boldsymbol{x}_{j+h} \boldsymbol{x}'_j, & h < 0. \end{cases}$$

and the kernel function is chosen to be the Parzen kernel:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3, & 0 \leq x \leq 1/2; \\ 2(1-x)^3, & 1/2 \leq x \leq 1; \\ 0, & x > 1 \text{ or } x < 0. \end{cases}$$

Also, we make the choice of bandwidth $H = \max(H_1, H_2)$, with $H_i = c^* \xi_i^{4/5} n^{3/5}$, where $\xi_i^2 = a_i^2 / \sqrt{T \int_0^T \sigma_{it}^4 dt}$ and $c^* = 3.5134$. Here, in order to optimize the behavior of the MRK, we try the infeasible case, i.e., H_i is calculated using the true values.

The benchmark is the RC estimator using the original complete synchronized data set without being contaminated by the market microstructure noise. This ideal estimator is only available in simulations, and is unfeasible in practice. This benchmark estimator is only computed here so that other estimators can be compared to it. We also implement the HY estimator with asynchronous data but no noise. The comparison results with various sampling frequencies are shown in Table 2.2.

It is clear from the RMSE reported here that the QMLE, the TSCV and the MRK behave quite well from the simulation studies and that neither of them suffers from a synchronization problem. But in some case, the bias cannot be ignored. One common drawback for the two nonparametric estimators is the choice of bandwidth, in that the selected bandwidth is suboptimal when trying to estimate the diagonal elements of the covariance matrix. By contrast, the QMLE has no tuning parameters, and is asymptotically unbiased at least for synchronous data, and is designed to have a smaller asymptotic variance since the rate is higher than that of the $n^{1/5}$-consistent MRK and the $n^{1/6}$-consistent TSCV. The comparison of the TSCV and the MRK indicates that the asymptotic bias associated with the MRK estimator overcomes its edge from its higher convergence rate in finite sample.

As to the synchronization comparison, Table 2.2 provides an example where the Refresh Time scheme may actually outperform the HY method in terms of data inclusion.

Empirically, there is not a wide gap between the amount of data retained by each method, as shown in Table 2.4. The HY method does not account for the noise and consequently may interfere with the autocorrelation structure introduced by the microstructure noise; hence it is not surprising that is may be dominated in our setting where the data are noisy.

Figure 2.2 verifies in simulations the central limit distribution given in Theorem 11. We estimate the covariance and correlation using synchronized assets and plot the histogram of the standardized estimates. The histogram matches well with the standard Gaussian distribution predicted by the theorem.

2.5 Empirical Study I: Foreign Exchange Futures

2.5.1 Data Description

Foreign exchange future contracts are traded on the Chicago Mercantile Exchange (CME) on a 24-hour clock. These marked-to-market futures contracts are very liquid. We are interested in estimating the covariations among pairs of currencies; we focus on the Australian dollar, Canadian dollar, Euro, British pound, Japanese yen and Swiss franc futures contracts. The contracts are quoted in terms of the unit value of the foreign currency measured in US dollars. We use currency futures instead of currency spot prices because the former contracts are traded in an exchange setting. By contrast, the currency spot markets, while extremely liquid, are over-the-counter, and may therefore suffer from additional microstructure issues, e.g. the financial stability of the counterparties and the quality of the execution they provide. Another advantage with futures data is that the interest rate differentials are embedded in them, which facilitates the calculation of the carry trade returns, so that we can avoid calculating them in the same way as in Brunnermeier et al. (2009). The cleaned data are available from Tick Data, Inc., among other data vendors.

Timing is a critical issue in the FX future market. Our sample period runs from January 1, 2007 to June 30, 2009, covering the most critical stage of the recent financial crisis. The future contracts are automatically rolled over to provide continuous price records. The tick-by-tick transaction prices are recorded in exchange time (Chicago Time). Therefore, in a given day, trading activity starts from Asia and Europe, followed by North America

and then Australia. There is no overlap between the opening hours of the Asian market and the US market. The advantage of using the Exchange Time is that we do not need to deal with different daylight saving rules for different continents in that the traders on the CME are likely to adjust their trading behavior according to the US time (where most of the activity takes place) rather than their local time. The electronic trading starts at 5pm on the previous day and ends at 4pm on current day. In between, there is a one-hour gap. Therefore, we redefine the day in accordance with the electronic trading system, so that we can avoid including potential jumps from the market close price to its open price.

Our data preprocessing eliminates the transactions on Saturdays and Sundays, US federal holidays, the day after Thanksgiving, December 24 to 26, and December 31 to January 2, because of the relatively smaller volume of activity. We are left with 618 days in the sample. For each particular FX contract, we take the average price for any multiple transactions that happen exactly at the same time stamp. The summary statistics for each FX futures are listed in Table 2.3. It appears that there exists an MA(1) structure for each futures contract, indicating that our model assumptions based on market microstructure noise are plausible.

Figure 2.3 plots the average hourly trading volume for each contract over the whole sample period, indicating that traders on CME tend to trade these currency futures simultaneously, and that there is no clear evidence of home bias in our sample, that is, e.g., traders tend to trade JPY instead of EUR, during their active trading hours. Therefore, it would be reasonable to apply the Refresh Time scheme to synchronize the data sets.

2.5.2 Empirical Findings

The estimation results are reported in Table 2.5. We find that the correlations are all positive, which is consistent with the findings in the literature based on lower frequency data, see e.g. Campbell et al. (2009). As shown in Figure 2.4, the relationship between the Swiss franc and the Euro exhibits a large correlation, followed by the British Pound, consistently with the high degree of integration among the European economies as well as the exchange rate policies followed by the Swiss Central Bank. The correlation between the Swiss franc

and the Euro slightly decreased in the middle of the crisis, but remained at a high level. As one of the world's major commodity currency, the Australian dollar has lower correlations with the other currencies, while the Canadian dollar is even further detached from them, owing perhaps to its close dependence on the American economy rather than the European or Asian economies. The large impact of the carry trade demise at various points during the financial crisis on the Australian dollar (especially in relation to the Japanese yen) is partly responsible for large swings in that exchange rate: high interest rates in Australia and low interest rates in Japan combine to make this pair of currencies an attractive target for carry traders.

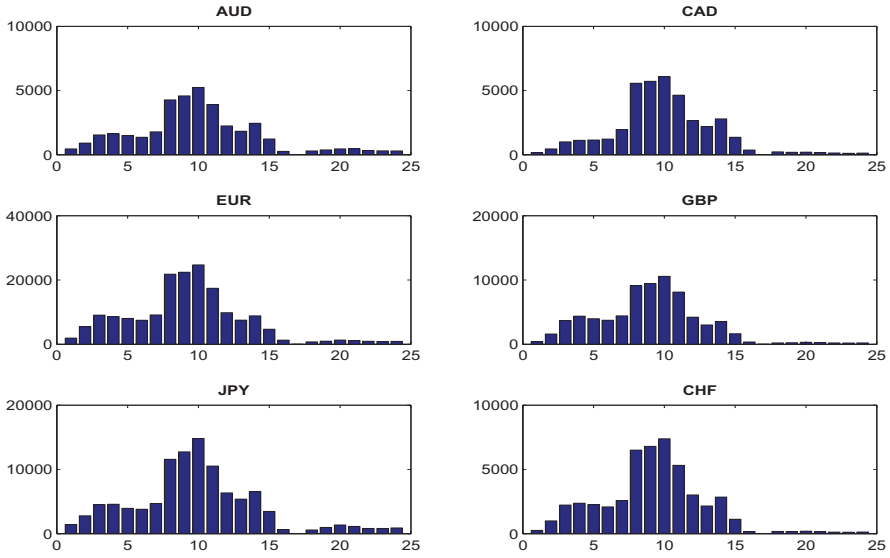


Figure 2.3: In this figure, we plot the average hourly trading volume of the foreign exchange future contracts from Jan 2007 to Jun 2009. The x-axis is measured in Chicago time.

Figure 2.4 also describes very similar patterns for those pairs that include the Japanese yen: correlations decreased from their respective normal levels during the period of highest uncertainty. This is probably due to the characteristics of the Japanese economy and its role in the financial crisis. The Japanese economy is strongly dependent on exports, and the yen played the role of a reserve currency during the crisis since the Japanese banking system emerged relatively unscathed (as did the Canadian one). An additional factor was likely the return of yen deposits previously borrowed by hedge funds and similar actors

in the pursuit of carry trades that were the victims of heavy deleveraging at the height of the crisis.

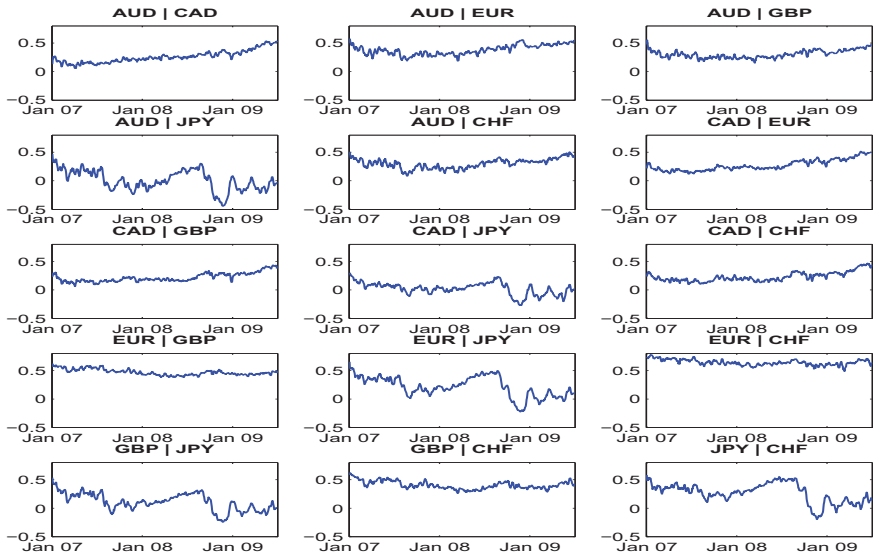


Figure 2.4: In this figure, we report the time series of correlation estimates of all six FX futures. Each curve is a 5-day moving average of the daily correlations.

2.5.3 Robustness Checks

Unlike stock prices that mainly respond to individual company events, the FX markets is driven primarily by macroeconomic events such as interest rate policy changes, for instance. The correlation estimates may therefore be sensitive to specific news announcement. The most active trading hours, as can be seen from Figure 2.3, are from 7 AM to 11 AM Chicago Time when both the Europe and US trading desks are open. In addition, most of the US macro news announcements are released during this period, see, e.g., Andersen et al. (2003). To check for the effect of these announcements, we divide the intraday observations into three trading sessions: 5 PM - 7 AM, 7 AM - 11 PM and 11 PM - 4PM. In the first session, the Asian and Australian markets are open, with European markets joining later. The second session is the most active period within the day. Only the North American markets are open in the last session. The estimates are reported in Table 2.6. The correlation estimates in the second session are mostly higher than the other two, but the

differences between the first session and last session are not significant.

Refresh Time tends to be selected when less liquid assets trade, hence the price of the liquid asset in the synchronized pairs are more likely to be stale. The average time difference between the refresh time and the real time is a measure of staleness. To check the robustness of the synchronization procedure, we require in addition that one asset in each pair has to lead the other one, and then taking the average estimates given by the two synchronization scheme. The comparison results are reported in Table 2.5 (lower diagonal). The estimates do not differ significantly from the estimates given by the Refresh Time scheme.

2.6 Empirical Study II: Crude Oil and S&P 500 Futures

To study the correlation between commodities and stock markets, we use high-frequency S&P 500 and Crude Oil futures as instruments due to their representativeness and availability. The S&P 500 futures are traded on the Chicago Mercantile Exchange (CME) from 8:30 a.m. to 3:15 p.m Central Time. The regular-size contracts are liquidly traded via the open outcry market, whereas the electronic market is in charge of mini-contracts. The Crude Oil futures used to be the most popular energy contract in the New York Mercantile Exchange (NYMEX), which has become part of the CME group recently. Since June 2006, the 24-hour electronic market for Crude Oil has started to take over the open outcry market. Nevertheless, the most active trading period is between 9:00 a.m. and 2:30 p.m. Eastern Time (after Feb 2007), when the open outcry market is open. We only consider the relatively liquid S&P500 futures traded via open outcry and the Crude Oil futures traded electronically. It is reasonable to assume that the microstructure noises across the two markets are uncorrelated. The sample period ranges from Feb 1, 2007 to Dec 18, 2009, covering the 2008 financial crisis.

Figure 2.5 describes the daily realized volatility for S&P 500 and Crude Oil respectively, and Figure 2.6 plots their daily correlation over the sampling period. Apparently, Crude Oil, the primary component of most commodity indices, is more volatile than the aggregated stock market, indicating that the commodity markets are potentially riskier

and more attractive than the equity markets. As recent financial innovation recreates commodities as a new trading avenue for investors, the price of an individual commodity is no longer determined by its own supply and demand, which partially leads to larger correlations between the aggregated financial markets and commodity markets, see Tang and Xiong (2010). However, the correlation is time-varying and may sometimes behave abnormally for a relatively short period. For instance, shortly after the Bear Sterns' fire sale, the returns of the two markets were negatively correlated, indicating that commodities and in particular Crude Oil future, might be used to hedge against the stock market downturn. However, in the climax of the financial crisis in October 2010, both markets started to move at the same pace, leading to a positive correlation of the returns. Such short term pattern may not be captured, at least punctually, using lower frequency data, since years of historical data may dilute the short term abnormality.

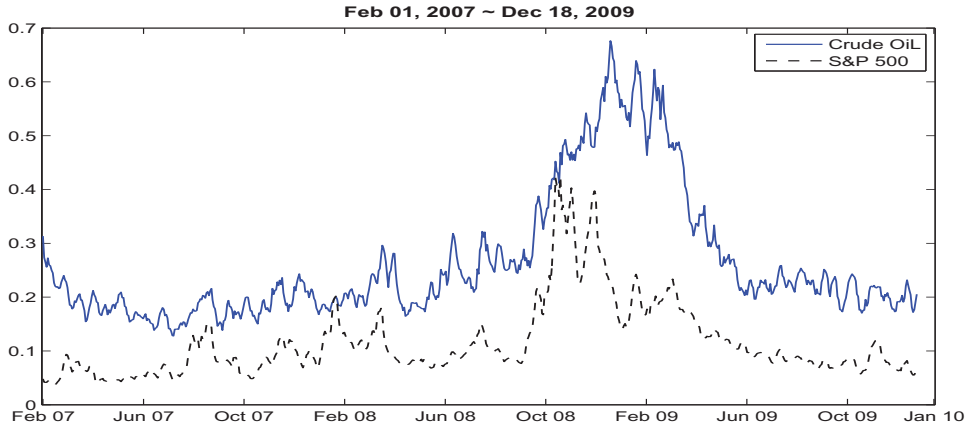


Figure 2.5: The 5-day moving average of the daily annualized volatilities of the S&P 500 and Crude Oil Futures

2.7 Conclusions

This chapter proposes quasi-likelihood based estimators of high frequency covariance and correlation with noisy and asynchronous tick-by-tick data as well as a generalized synchronization scheme. This is so far the only estimator that can deal with both noise and asynchronism while maintaining optimal convergence rates. Moreover, the feature of being

Figure 2.6: The 5-day moving average of the daily correlation between the S&P 500 and Crude Oil Futures

bandwidth-free makes this estimator advantageous in practice over other nonparametric estimators, by removing an essentially arbitrary degree of freedom. The proposed synchronization scheme is conservative in deleting data. Empirically, the chapter documents the time-varying correlations between six major currencies, and between commodities and stock markets during the recent financial crisis.

2.8 Appendix

2.8.1 Proof of Theorem 11

Without loss of generality, we assume that the drift terms are zero. Consider two new processes $\tilde{X}_1 \pm \tilde{X}_2$. Apparently, each of them is a one dimensional Itô process with additive i.i.d. noise. According to Theorem 7, $\widehat{Var}(\tilde{X}_1 + \tilde{X}_2)$, for example, is a consistent and efficient estimator of $\frac{1}{T}\int_0^T \sigma_{1t}^2 + \sigma_{2t}^2 + 2\rho_t\sigma_{1t}\sigma_{2t}dt$ with convergence rate $n^{1/4}$. Therefore, the covariance and correlation estimators are $n^{1/4}$-consistent, that is,

$$\widehat{Cov}(\tilde{X}_1, \tilde{X}_2) - \frac{1}{T}\int_0^T \rho_t\sigma_{1t}\sigma_{2t}dt = O_P(n^{-\frac{1}{4}})$$

$$\widehat{Corr}(\tilde{X}_1, \tilde{X}_2) - \frac{\int_0^T \rho_t \sigma_{1t} \sigma_{2t} dt}{\sqrt{\int_0^T \sigma_{1t}^2 dt} \sqrt{\int_0^T \sigma_{2t}^2 dt}} = O_P(n^{-\frac{1}{4}})$$

Also, the central limit theorem for $\widehat{Var}(\tilde{X}_1 + \tilde{X}_2)$ clearly holds and the asymptotic variance is given by Remark 1 with a_0 replaced by $(a_{1,0}^2 + 2\theta a_{1,0} a_{2,0} + a_{2,0}^2)^{\frac{1}{2}}$, where θ is the correlation between U_{1t} and U_{2t}. In the synchronous case, we can identify the correlation between the noises. Hence, we can work with a more general case here.

Therefore, in order to obtain the central limit theorem for the covariance and correlation estimators, it is sufficient to derive the asymptotic covariances of the four QMLE estimators:

$$V^{\alpha,\beta} = Acov\big(\widehat{Var}(\alpha\tilde{X}_1 + \tilde{X}_2), \widehat{Var}(\tilde{X}_1 - \beta\tilde{X}_2)\big)$$

where α and β can be $0, 1$ and -1.

We start with $\alpha = \beta = 1$. Suppose (σ_1^2, a_1^2) and (σ_2^2, a_2^2) are the parameters in two respective quasi-likelihood functions. It can be shown that the QMLE for the volatility and the noise variance are asymptotically independent, so we only need to consider the score functions with respect to σ_k^2, denoted as $\Psi_{k,n}$, where $k = 1, 2$. Let the inverse of the covariance matrix in each likelihood (1.2) be $\mathbf{\Omega}_k^{-1} = (\omega_k^{i,j})$. For any process A, denote $\Delta_i^n A = A_{\tau_i} - A_{\tau_{i-1}}$. Define,

$$\Psi_{k,n}(\omega) = \frac{1}{2\sqrt{n}}\{\frac{\partial \log(\det \mathbf{\Omega}_k)}{\partial \sigma_k^2} + \boldsymbol{Y}'_k \frac{\partial \mathbf{\Omega}_k^{-1}}{\partial \sigma_k^2} \boldsymbol{Y}_k\}$$
$$\bar{\Psi}_{k,n}(\omega) = \frac{1}{2\sqrt{n}}\{\frac{\partial \log(\det \mathbf{\Omega}_k)}{\partial \sigma_k^2} + tr(\frac{\partial \mathbf{\Omega}_k^{-1}}{\partial \sigma_k^2} \mathbf{\Sigma}_k)\}$$

where the i^{th} components of $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are

$$Y_{1,i} = \Delta_i^n \tilde{X}_1 + \Delta_i^n \tilde{X}_2$$
$$Y_{2,i} = \Delta_i^n \tilde{X}_1 - \Delta_i^n \tilde{X}_2$$

$\mathbf{\Sigma}_1 = (\Sigma_{1,i,j})$ and $\mathbf{\Sigma}_2 = (\Sigma_{2,i,j})$ are specified in the following way:

$$\begin{aligned}
\Sigma_{1,i,i} &= \int_{\tau_{i-1}}^{\tau_i} \sigma_{1,t}^2 + \sigma_{2,t}^2 + 2\rho_t\sigma_{1t}\sigma_{2t}dt + 2a_1^{*2}. \\
\Sigma_{2,i,i} &= \int_{\tau_{i-1}}^{\tau_i} \sigma_{1,t}^2 + \sigma_{2,t}^2 - 2\rho_t\sigma_{1t}\sigma_{2t}dt + 2a_2^{*2}. \\
\Sigma_{1,i,i-1} &= \Sigma_{1,i,i+1} = -a_1^{*2} \\
\Sigma_{2,i,i-1} &= \Sigma_{2,i,i+1} = -a_2^{*2}
\end{aligned}$$

where $a_1^{*2} = a_{1,0}^2 + a_{2,0}^2 + 2\theta a_{1,0}a_{2,0}$ and $a_2^{*2} = a_{1,0}^2 + a_{2,0}^2 - 2\theta a_{1,0}a_{2,0}$.

Also, let $\epsilon_{1,i} = \Delta_i^n U_1 + \Delta_i^n U_2$ and $\epsilon_{2,i} = \Delta_i^n U_1 - \Delta_i^n U_2$. Thus, the difference between $\Psi_{1,n}$ and $\bar{\Psi}_{1,n}$ is

$$\begin{aligned}
&\Psi_{1,n} - \bar{\Psi}_{1,n} \\
=&\frac{1}{2\sqrt{n}}\{\boldsymbol{Y}_1'\frac{\partial\mathbf{\Omega}_1^{-1}}{\partial\sigma_1^2}\boldsymbol{Y}_1 - tr(\frac{\partial\mathbf{\Omega}_1^{-1}}{\partial\sigma_1^2}\mathbf{\Sigma}_1)\} \\
=&\frac{1}{2\sqrt{n}}\{\sum_{i=1}^n \frac{\partial\omega_1^{ii}}{\partial\sigma_1^2}((\Delta_i^n X_1 + \Delta_i^n X_2)^2 - \int_{\tau_{i-1}}^{\tau_i} \sigma_{1t}^2 + \sigma_{2t}^2 + 2\rho_t\sigma_{1t}\sigma_{2t}dt) \\
&+ \sum_{i=1}^n\sum_{j\neq i}^n \frac{\partial\omega_1^{ij}}{\partial\sigma_1^2}(\Delta_i^n X_1 + \Delta_i^n X_2)(\Delta_j^n X_1 + \Delta_j^n X_2) \\
&+ 2\sum_{i=1}^n\sum_{j=1}^n \frac{\partial\omega_1^{ij}}{\partial\sigma_1^2}\epsilon_{1,j}(\Delta_i^n X_1 + \Delta_i^n X_2) \\
&+ \sum_{i=1}^n\sum_{j=1}^n \frac{\partial\omega_1^{ij}}{\partial\sigma_1^2}(\epsilon_{1,i}\epsilon_{1,j} - E\epsilon_{1,i}\epsilon_{1,j})\} \\
:=&\frac{1}{2\sqrt{n}}(M_1^1 + M_2^1 + M_3^1 + M_4^1)
\end{aligned} \tag{2.8}$$

$\Psi_{2,n} - \bar{\Psi}_{2,n}$ can be decomposed in a similar way. It follows from Theorem 7.1 in Jacod (2007) that

$$Acov(n^{-\frac{1}{4}}M_2^1, n^{-\frac{1}{4}}M_2^2) = \frac{a_2^2\sigma_1^2 + a_1^2\sigma_2^2 + 3a_1a_2\sigma_1\sigma_2}{2\sqrt{T}\sigma_1^3\sigma_2^3(\sigma_1a_2 + \sigma_2a_1)^3}\int_0^T (\sigma_{1t}^2 - \sigma_{2t}^2)^2dt$$

Further, we can show that $n^{-\frac{1}{4}}(M_1^1 + M_2^1)$ and $n^{-\frac{1}{4}}(M_1^2 + M_2^2)$ jointly $\sigma(X)$-stable convergence in law, due to the fact that $n^{-\frac{1}{4}}M_1^1$ and $n^{-\frac{1}{4}}M_1^2$ are asymptotically negligible.

Note that

$$\sum_{i=1}^{n}(\Delta_i^n X_1 + \Delta_i^n X_2)(\Delta_i^n X_1 - \Delta_i^n X_2) \xrightarrow{\mathcal{P}} \int_0^T (\sigma_{1t}^2 - \sigma_{2t}^2)dt$$

By conditioning on the filtration $\sigma(X)$ and standard central limit theorem, it follows that

$$Acov(n^{-\frac{1}{4}}M_3^1, n^{-\frac{1}{4}}M_3^2) = \frac{(a_{1,0}^2 - a_{2,0}^2)}{\sqrt{T}\sigma_1\sigma_2(\sigma_1 a_2 + \sigma_2 a_1)^3}\int_0^T (\sigma_{1t}^2 - \sigma_{2t}^2)dt$$

Last, the contribution of the noise terms is

$$\begin{aligned}
&Acov(n^{-\frac{1}{4}}M_4^1, n^{-\frac{1}{4}}M_4^2) \\
&= \lim_{n\to\infty}\sum_{i,j=1}^{n}\sum_{k,l=1}^{n}\frac{\partial\omega_1^{ij}}{\partial\sigma_1^2}\frac{\partial\omega_2^{kl}}{\partial\sigma_2^2}\Big(\text{cum}(\epsilon_{1,i},\epsilon_{1,j},\epsilon_{2,k},\epsilon_{2,l}) + 2\text{cov}(\epsilon_{1,i},\epsilon_{1,j})\text{cov}(\epsilon_{2,k},\epsilon_{2,l})\Big) \\
&:= \lim_{n\to\infty} V_1(\frac{\partial\omega_1}{\partial\sigma_1^2},\frac{\partial\omega_2}{\partial\sigma_2^2}) + V_2(\frac{\partial\omega_1}{\partial\sigma_1^2},\frac{\partial\omega_2}{\partial\sigma_2^2})
\end{aligned}$$

where

$$\begin{aligned}
V_1(v,\omega) &= \sum_{i,j,k,l=1}^{n} v^{ij}\omega^{kl}\text{cum}(\epsilon_{1,i},\epsilon_{1,j},\epsilon_{2,k},\epsilon_{2,l}) \\
V_2(v,\omega) &= \sum_{i,j,k,l=1}^{n} v^{ij}\omega^{kl}2\text{cov}(\epsilon_{1,i},\epsilon_{1,j})\text{cov}(\epsilon_{2,k},\epsilon_{2,l}) \\
&= 2(a_{1,0}^2 - a_{2,0}^2)^2\sum_{i=1}^{N}\sum_{j=1}^{N}\{v^{ij}(\omega^{j-1,i-1} + \omega^{j-1,i+1} - 2\omega^{j-1,i} + \omega^{j+1,i-1} \\
&\quad + \omega^{j+1,i+1} - 2\omega^{j+1,i} - 2(\omega^{j,i-1} + \omega^{j,i+1} - 2\omega^{j,i}))\}
\end{aligned}$$

And it follows from a similar proof of the Lemma 1 in Aït-Sahalia et al. (2005) that,

$$\begin{aligned}
&\text{cum}(\epsilon_{1,i},\epsilon_{1,j},\epsilon_{2,k},\epsilon_{2,l}) \\
&= \begin{cases} 2S(U_1,U_2), & \text{if } i=j=k=l; \\ (-1)^{s(i,j,k,l)}S(U_1,U_2), & \text{if } \max(i,j,k,l) = \min(i,j,k,l) + 1; \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

where $S(U_1, U_2) = \text{cum}_4[U_1] + \text{cum}_4[U_2] - 2\text{cov}(U_1^2, U_2^2) + 4\text{cov}^2(U_1, U_2)$ and $s(i,j,k,l)$ denotes the number of indices among (i,j,k,l) that are equal to $\min(i,j,k,l)$.

Note that for $k = 1, 2$, we have

$$\begin{aligned} V_k(\frac{\partial\omega_1}{\partial\sigma_1^2}, \frac{\partial\omega_2}{\partial\sigma_2^2}) = &(\frac{\partial\eta_1}{\partial\sigma_1^2})(\frac{\partial\eta_2}{\partial\sigma_2^2})V_k(\frac{\partial\omega_1}{\partial\eta_1}, \frac{\partial\omega_2}{\partial\eta_2}) + \frac{1}{\gamma_1^2\gamma_2^2}(\frac{\partial\gamma_1^2}{\partial\sigma_1^2}\frac{\partial\gamma_2^2}{\partial\sigma_2^2})V_k(\omega_1, \omega_2) \\ &- \frac{1}{\gamma_1^2}\frac{\partial\eta_2}{\partial\sigma_2^2}\frac{\partial\gamma_1^2}{\partial\sigma_1^2}V_k(\omega_1, \frac{\partial\omega_2}{\partial\eta_2}) - \frac{1}{\gamma_2^2}\frac{\partial\eta_1}{\partial\sigma_1^2}\frac{\partial\gamma_2^2}{\partial\sigma_2^2}V_k(\frac{\partial\omega_1}{\partial\eta_1}, \omega_2) \end{aligned}$$

where η_i and γ_i are given in (2.9) and (2.10), with σ^2 and a^2 replaced by σ_i^2 and a_i^2. Following a similar calculation as in Lemma 2 of Xiu (2010), we can obtain

$$\begin{aligned} V_1(\frac{\partial\omega_1}{\partial\sigma_1^2}, \frac{\partial\omega_2}{\partial\sigma_2^2}) &= O(1) \\ V_2(\frac{\partial\omega_1}{\partial\sigma_1^2}, \frac{\partial\omega_2}{\partial\sigma_2^2}) &= \frac{(a_{1,0}^2 - a_{2,0}^2)^2\sqrt{T}}{2a_1a_2(a_2\sigma_1 + a_1\sigma_2)^3}n^{\frac{1}{2}} + o(n^{\frac{1}{2}}) \end{aligned}$$

Therefore,

$$Acov(n^{-\frac{1}{4}}M_4^1, n^{-\frac{1}{4}}M_4^2) = \frac{(a_{1,0}^2 - a_{2,0}^2)^2\sqrt{T}}{2a_1a_2(a_2\sigma_1 + a_1\sigma_2)^3}$$

Combining the above results with Remark 1 and applying Lemma 1 and Proposition 5 in Barndorff-Nielsen et al. (2008b), we can obtain the central limit theorem for two-dimensional case:

$$\begin{pmatrix} n^{\frac{1}{4}}(\Psi_{1,n} - \bar{\Psi}_{1,n}) \\ n^{\frac{1}{4}}(\Psi_{2,n} - \bar{\Psi}_{2,n}) \end{pmatrix} \xrightarrow{\mathcal{L}_X} MN\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} U_{11} & U_{12} \\ U_{12} & U_{22} \end{pmatrix}\Big)$$

where

$$\begin{aligned} U_{12} =&\frac{a_2^2\sigma_1^2 + a_1^2\sigma_2^2 + 3a_1a_2\sigma_1\sigma_2}{8\sqrt{T}\sigma_1^3\sigma_2^3(\sigma_1a_2 + \sigma_2a_1)^3}\int_0^T(\sigma_{1t}^2 - \sigma_{2t}^2)^2dt \\ &+ \frac{(a_{1,0}^2 - a_{2,0}^2)}{4\sqrt{T}\sigma_1\sigma_2(\sigma_1a_2 + \sigma_2a_1)^3}\int_0^T(\sigma_{1t}^2 - \sigma_{2t}^2)dt + \frac{(a_{1,0}^2 - a_{2,0}^2)^2\sqrt{T}}{8a_1a_2(a_2\sigma_1 + a_1\sigma_2)^3} \\ U_{11} =&\frac{5\int_0^T(\sigma_{1t}^2 + 2\rho_t\sigma_{1t}\sigma_{2t} + \sigma_{2t}^2)^2dt}{64a_1\sigma_1^7\sqrt{T}} + \frac{(a_{1,0}^2 + a_{2,0}^2 + 2\theta a_{1,0}a_{2,0})^2\sqrt{T}}{64a_1^5\sigma_1^3} \end{aligned}$$

$$+\frac{(a_{1,0}^2+a_{2,0}^2+2\theta a_{1,0}a_{2,0})\int_0^T \sigma_{1t}^2+\sigma_{2t}^2+2\rho_t\sigma_{1t}\sigma_{2t}dt}{32\sigma_1^5 a_1^3\sqrt{T}}$$

$$U_{22}=\frac{5\int_0^T(\sigma_{1t}^2-2\rho_t\sigma_{1t}\sigma_{2t}+\sigma_{2t}^2)^2dt}{64a_2\sigma_2^7\sqrt{T}}+\frac{(a_{1,0}^2+a_{2,0}^2-2\theta a_{1,0}a_{2,0})^2\sqrt{T}}{64a_2^5\sigma_2^3}$$

$$+\frac{(a_{1,0}^2+a_{2,0}^2-2\theta a_{1,0}a_{2,0})\int_0^T \sigma_{1t}^2+\sigma_{2t}^2-2\rho_t\sigma_{1t}\sigma_{2t}dt}{32\sigma_2^5 a_2^3\sqrt{T}}$$

Therefore, a direct application of the Delta method yields

$$\begin{pmatrix} n^{\frac{1}{4}}(\hat{\sigma}_{1,n}^2-\frac{1}{T}\int_0^T \sigma_{1t}^2+2\rho_t\sigma_{1t}\sigma_{2t}+\sigma_{2t}^2dt) \\ n^{\frac{1}{4}}(\hat{\sigma}_{2,n}^2-\frac{1}{T}\int_0^T \sigma_{1t}^2-2\rho_t\sigma_{1t}\sigma_{2t}+\sigma_{2t}^2dt) \end{pmatrix} \xrightarrow{\mathcal{L}_X} MN\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix},\begin{pmatrix} V^{1,-1} & V^{1,1} \\ V^{1,1} & V^{-1,1} \end{pmatrix}\Big)$$

where

$$V^{1,-1}=T^{-\frac{1}{2}}\Big(5\frac{q_1^*}{\sigma_1^*}+3\sigma_1^{*3}\Big)a_1^*$$

$$V^{-1,1}=T^{-\frac{1}{2}}\Big(5\frac{q_2^*}{\sigma_2^*}+3\sigma_2^{*3}\Big)a_2^*$$

$$V^{1,1}=\frac{16a_1^*a_2^*}{T(a_2^*\sigma_1^*+a_1^*\sigma_2^*)^3}\Big(\frac{a_2^{*2}\sigma_1^{*2}+a_1^{*2}\sigma_2^{*2}+3a_1^*a_2^*\sigma_1^*\sigma_2^*}{2\sqrt{T}}\int_0^T(\sigma_{1t}^2-\sigma_{2t}^2)^2dt$$

$$+\frac{(a_{1,0}^2-a_{2,0}^2)\sigma_1^{*2}\sigma_2^{*2}}{\sqrt{T}}\int_0^T(\sigma_{1t}^2-\sigma_{2t}^2)dt+\frac{(a_{1,0}^2-a_{2,0}^2)^2\sigma_1^{*3}\sigma_2^{*3}\sqrt{T}}{2a_1^*a_2^*}\Big)$$

with $\sigma_1^{*2}=\frac{1}{T}\int_0^T \sigma_{1t}^2+\sigma_{2t}^2+2\rho_t\sigma_{1t}\sigma_{2t}dt, \sigma_2^{*2}=\frac{1}{T}\int_0^T \sigma_{1t}^2+\sigma_{2t}^2-2\rho_t\sigma_{1t}\sigma_{2t}dt, q_1^*=\frac{1}{T}\int_0^T(\sigma_{1t}^2+2\rho_t\sigma_{1t}\sigma_{2t}+\sigma_{2t}^2)^2dt$, and $q_2^*=\frac{1}{T}\int_0^T(\sigma_{1t}^2-2\rho_t\sigma_{1t}\sigma_{2t}+\sigma_{2t}^2)^2dt$.

Therefore,

$$n^{\frac{1}{4}}(\widehat{Cov}(\tilde{X}_1,\tilde{X}_2)-\frac{1}{T}\int_0^T \rho_t\sigma_{1t}\sigma_{2t}dt)\xrightarrow{\mathcal{L}_X} MN(0,V)$$

where

$$V=\begin{pmatrix} \frac{1}{4} & -\frac{1}{4} \end{pmatrix}\begin{pmatrix} V^{1,-1} & V^{1,1} \\ V^{1,1} & V^{-1,1} \end{pmatrix}\begin{pmatrix} \frac{1}{4} \\ -\frac{1}{4} \end{pmatrix}$$

Similarly,

$$V^{\alpha,\beta}=\frac{16a_1^\alpha a_2^\beta}{2T^{\frac{3}{2}}(a_2^\beta\sigma_1^\alpha+a_1^\alpha\sigma_2^\beta)^3}\Big(((a_2^\beta)^2(\sigma_1^\alpha)^2+(a_1^\alpha)^2(\sigma_2^\beta)^2$$

$$
\begin{aligned}
&+ 3a_1^{\alpha}a_2^{\beta}\sigma_1^{\alpha}\sigma_2^{\beta}\Big)\int_0^T(\alpha\sigma_{1t}^2-\beta\sigma_{2t}^2+(1-\alpha\beta)\rho_t\sigma_{1t}\sigma_{2t})^2dt\\
&+ 2(\alpha a_{1,0}^2-\beta a_{2,0}^2+(1-\alpha\beta)\theta a_{1,0}a_{2,0})(\sigma_1^{\alpha})^2(\sigma_2^{\beta})^2\\
&\cdot\int_0^T(\alpha\sigma_{1t}^2-\beta\sigma_{2t}^2+(1-\alpha\beta)\rho_t\sigma_{1t}\sigma_{2t})dt\\
&+ (a_1^{\alpha}a_2^{\beta})^{-1}(\alpha a_{1,0}^2-\beta a_{2,0}^2+(1-\alpha\beta)\theta a_{1,0}a_{2,0})^2(\sigma_1^{\alpha})^3(\sigma_2^{\beta})^3T\Big)
\end{aligned}
$$

where

$a_1^{\alpha}=(\alpha^2a_{1,0}^2+a_{2,0}^2+2\alpha\theta a_{1,0}a_{2,0})^{\frac{1}{2}}$, $a_2^{\beta}=(a_{1,0}^2+\beta^2a_{2,0}^2-2\beta\theta a_{1,0}a_{2,0})^{\frac{1}{2}}$, $\sigma_1^{\alpha}=(\frac{1}{T}\int_0^T\alpha^2\sigma_{1t}^2+\sigma_{2t}^2+2\alpha\rho_t\sigma_{1t}\sigma_{2t}dt)^{\frac{1}{2}}$, and $\sigma_2^{\beta}=(\frac{1}{T}\int_0^T\sigma_{1t}^2+\beta^2\sigma_{2t}^2-2\beta\rho_t\sigma_{1t}\sigma_{2t}dt)^{\frac{1}{2}}$.

Denote the asymptotic variances of the $\widehat{Var}(\tilde{X}_i)$ as V^i, for $i=1,2$. According to Theorem 7,

$$
V^i=\frac{5a_{i,0}\int_0^T\sigma_{i,t}^4dt}{T(\int_0^T\sigma_{i,t}^2dt)^{\frac{1}{2}}}+\frac{3a_{i,0}(\int_0^T\sigma_{i,t}^2dt)^{\frac{3}{2}}}{T^2}
$$

Therefore, using the Delta method again, we obtain

$$
n^{\frac{1}{4}}(\widehat{Corr}(\tilde{X}_1,\tilde{X}_2)-\frac{\int_0^T\rho_t\sigma_{1t}\sigma_{2t}dt}{\sqrt{\int_0^T\sigma_{1t}^2dt}\sqrt{\int_0^T\sigma_{2t}^2dt}})\xrightarrow{\mathcal{L}_X}MN(0,\tilde{V})
$$

and $\tilde{V}=\boldsymbol{e}\boldsymbol{\Lambda}\boldsymbol{e}'$, where

$$
\boldsymbol{\Lambda}=\begin{pmatrix}
V^1 & . & . & .\\
V^{1,0} & V^{1,-1} & . & .\\
V^{-1,0} & V^{1,1} & V^{-1,1} & .\\
V^{0,0} & V^{0,-1} & V^{0,1} & V^2
\end{pmatrix}
$$

and

$$
\begin{aligned}
\boldsymbol{e}=(&-\frac{T\int_0^T\rho_t\sigma_{1t}\sigma_{2t}dt}{2(\int_0^T\sigma_{1t}^2dt)^{\frac{3}{2}}(\int_0^T\sigma_{2t}^2dt)^{\frac{1}{2}}},\frac{T}{4(\int_0^T\sigma_{1t}^2dt)^{\frac{1}{2}}(\int_0^T\sigma_{2t}^2dt)^{\frac{1}{2}}},\\
&-\frac{T}{4(\int_0^T\sigma_{1t}^2dt)^{\frac{1}{2}}(\int_0^T\sigma_{2t}^2dt)^{\frac{1}{2}}},-\frac{T\int_0^T\rho_t\sigma_{1t}\sigma_{2t}dt}{2(\int_0^T\sigma_{1t}^2dt)^{\frac{1}{2}}(\int_0^T\sigma_{2t}^2dt)^{\frac{3}{2}}}).
\end{aligned}
$$

2.8.2 Proof of Theorem 13

In order to simplify the notation and without loss of generality, we assume that $\tilde{X}_{it}$ is observed at $\{t_{i,j}, 0 \leq j \leq n, i = 1, 2\}$, where $t_{1,j-1} \leq t_{2,j-1} \leq \tau_{j-1} < t_{1,j} \leq t_{2,j}$, for each $0 \leq j \leq n$. Otherwise, we can just swap the subscripts 1 and 2, if needed. Denote $\bar{n} = T/\bar{\Delta}$. Let $\Omega^{-1} = (\omega^{ij})$. Of course, we can add one more condition: $K^{-1} \leq \sigma_{it}^2 \leq K, \forall\ t \in [0, T]$, since one can always relax the constraint by localization scheme.

$$\begin{aligned}\boldsymbol{\Psi}_n &= (\Psi_n^1(\omega), \Psi_n^2(\omega))' \\ &= \Big(\frac{1}{2\sqrt{\bar{n}}}\{\frac{\partial \log(\det \boldsymbol{\Omega})}{\partial \sigma^2} + \boldsymbol{Y}'\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma^2}\boldsymbol{Y}\}, \frac{1}{2\bar{n}}\{\frac{\partial \log(\det \boldsymbol{\Omega})}{\partial a^2} + \boldsymbol{Y}'\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial a^2}\boldsymbol{Y}\}\Big)'\end{aligned}$$

where $Y_i = X_{1t_{1,i}} + X_{2t_{2,i}} - X_{1t_{1,i-1}} - X_{2t_{2,i-1}} + \epsilon_i$ and $\epsilon_i = U_{1t_{1,i}} + U_{2t_{2,i}} - U_{1t_{1,i-1}} - U_{2t_{2,i-1}}$. Also,

$$\bar{\boldsymbol{\Psi}}_n = \Big(\frac{1}{2\sqrt{\bar{n}}}\{\frac{\partial \log(\det \boldsymbol{\Omega})}{\partial \sigma^2} + tr(\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial \sigma^2}\boldsymbol{\Sigma}_0)\}, \frac{1}{2\bar{n}}\{\frac{\partial \log(\det \boldsymbol{\Omega})}{\partial a^2} + tr(\frac{\partial \boldsymbol{\Omega}^{-1}}{\partial a^2}\boldsymbol{\Sigma}_0)\}\Big)'$$

where $\boldsymbol{\Sigma}_0 = (\Sigma_{0,i,j})$ is given by

$$\begin{aligned}\Sigma_{0,i,i} &= \int_{t_{1,i-1}}^{t_{1,i}} \sigma_{1t}^2 dt + \int_{t_{2,i-1}}^{t_{2,i}} \sigma_{2t}^2 dt + \int_{t_{2,i-1}}^{t_{1,i}} 2\rho_t \sigma_{1t}\sigma_{2t} dt + 2u_1^2 \\ \Sigma_{0,i,i-1} &= \Sigma_{0,i-1,i} = -u_1^2 + \int_{t_{1,i-1}}^{t_{2,i-1}} \rho_t \sigma_{1t}\sigma_{2t} dt\end{aligned}$$

and $u_1^2 = a_{1,0}^2 + a_{2,0}^2$. The other coefficients are 0.

On the other hand, we can write

$$\omega^{ij} = \frac{(-\eta)^{|i-j|} - (-\eta)^{i+j} - (-\eta)^{2n-i-j+2} + (-\eta)^{2n-|i-j|+2}}{\gamma^2(1-\eta^2)(1-\eta^{2n+2})}$$

where

$$\eta = \frac{1}{2a^2}\{-2a^2 - \sigma^2\bar{\Delta} + \sqrt{\sigma^2\bar{\Delta}(4a^2 + \sigma^2\bar{\Delta})}\} \tag{2.9}$$

$$\gamma^2 = \frac{1}{2}\{2a^2 + \sigma^2\bar{\Delta} + \sqrt{\sigma^2\bar{\Delta}(4a^2 + \sigma^2\bar{\Delta})}\} \tag{2.10}$$

We consider the same decomposition of $\mathbf{\Psi}_n - \bar{\mathbf{\Psi}}_n$ as in formula (2.8). Obviously, data synchronization has no effect on the order of the variances for the last two terms; hence we are left with the following two asymptotic results that are in need of verification:

$$\sum_{i=1}^{n} \omega^{ii} \{ (\int_{t_{1,i-1}}^{t_{1,i}} \sigma_{1t} dW_{1t} + \int_{t_{2,i-1}}^{t_{2,i}} \sigma_{2t} dW_{2t})^2$$
$$- \int_{t_{1,i-1}}^{t_{1,i}} \sigma_{1t}^2 dt - \int_{t_{2,i-1}}^{t_{2,i}} \sigma_{2t}^2 dt - \int_{t_{2,i-1}}^{t_{1,i}} 2\rho_t \sigma_{1t} \sigma_{2t} dt \} = O_p(1)$$
$$\sum_{i=1}^{n} \sum_{j \neq i}^{n} \omega^{ij} (\int_{t_{1,i-1}}^{t_{1,i}} \sigma_{1t} dW_{1t} + \int_{t_{2,i-1}}^{t_{2,i}} \sigma_{2t} dW_{2t}) (\int_{t_{1,j-1}}^{t_{1,j}} \sigma_{1t} dW_{1t} + \int_{t_{2,j-1}}^{t_{2,j}} \sigma_{2t} dW_{2t})$$
$$- \sum_{i=1}^{n} (\omega^{i,i-1} + \omega^{i-1,i}) \int_{t_{1,i-1}}^{t_{2,i-1}} \rho_t \sigma_{1t} \sigma_{2t} dt = O_p(\bar{n}^{\frac{1}{4}})$$

By Lemma 1 in Xiu (2010) and using the tower property of the conditional expectation (conditioning on $\{\Delta_i\}$), we have for any $k = 1, 2$,

$$\sum_{i=1}^{n} \omega^{ii} \{ (\int_{t_{k,i-1}}^{t_{k,i}} \sigma_{kt} dW_{kt})^2 - \int_{t_{k,i-1}}^{t_{k,i}} \sigma_{kt}^2 dt \} = O_p(1)$$
$$\sum_{i=1}^{n} \sum_{j \neq i}^{n} \omega^{ij} (\int_{t_{k,i-1}}^{t_{k,i}} \sigma_{kt} dW_{kt} \int_{t_{k,j-1}}^{t_{k,j}} \sigma_{kt} dW_{kt}) = O_p(\bar{n}^{\frac{1}{4}})$$

Hence, it is sufficient to show that

$$\sum_{i=1}^{n} \omega^{ii} \{ \int_{t_{1,i-1}}^{t_{1,i}} \sigma_{1t} dW_{1t} \int_{t_{2,i-1}}^{t_{2,i}} \sigma_{2t} dW_{2t} - \int_{t_{2,i-1}}^{t_{1,i}} \rho_t \sigma_{1t} \sigma_{2t} dt \} = O_p(1)$$
$$\sum_{i=1}^{n} \omega^{i,i-1} \int_{t_{1,i-1}}^{t_{1,i}} \sigma_{1t} dW_{1t} \int_{t_{2,i-2}}^{t_{2,i-1}} \sigma_{2t} dW_{2t} - \sum_{i=1}^{n} \omega^{i,i-1} \int_{t_{1,i-1}}^{t_{2,i-1}} \rho_t \sigma_{1t} \sigma_{2t} dt = O_p(1)$$

These two equalities are in fact similar, since $\omega_{i,i-1} = \omega_{i,i} + o(1)$. It can be proved by noting that $\int_{t_{1,i-1}}^{t_{2,i-1}} \rho_t \sigma_{1t} \sigma_{2t} dt$ is the quadratic covariation of $\int_{t_{1,i-1}}^{t_{1,i}} \sigma_{1t} dW_{1t}$ and $\int_{t_{2,i-2}}^{t_{2,i-1}} \sigma_{2t} dW_{2t}$.

So far, we have obtained the point-wise convergence of $\mathbf{\Psi}_n - \bar{\mathbf{\Psi}}_n$ to 0. Following the same reasoning in the proofs of Theorems 2 and 4 in Xiu (2010), we can prove the stochastic equicontinuity of $\mathbf{\Psi}_n - \bar{\mathbf{\Psi}}_n$, which guarantees that

$$\sup_{\sigma^2, a^2} \| \mathbf{\Psi}_n(\sigma^2, a^2) - \bar{\mathbf{\Psi}}_n(\sigma^2, a^2) \| \xrightarrow{\mathrm{P}} 0.$$

Next, we verify the identifiability condition and find the roots of the $\bar{\Psi}_n$.

$$\begin{aligned}
\bar{\Psi}_n^2 &= \frac{1}{2\bar{n}}\{tr(\mathbf{\Omega}^{-1}\frac{\partial \mathbf{\Omega}}{\partial a^2}) + tr(\frac{\partial \mathbf{\Omega}^{-1}}{\partial a^2}\mathbf{\Sigma}_0)\} \\
&= \frac{1}{\bar{n}}\{(tr\mathbf{\Omega}^{-1} - tr\mathbf{\Omega}^{-1}\boldsymbol{J}) + u_1^2(tr\frac{\partial \mathbf{\Omega}^{-1}}{\partial a^2} - tr\frac{\partial \mathbf{\Omega}^{-1}}{\partial a^2}\boldsymbol{J}) \\
&+ \frac{1}{2}\sum_{i=1}^{n}\frac{\partial \omega^{ii}}{\partial a^2}(\int_{t_{1,i-1}}^{t_{1,i}}\sigma_{1t}^2 dt + \int_{t_{2,i-1}}^{t_{2,i}}\sigma_{2t}^2 dt + \int_{t_{2,i-1}}^{t_{1,i}} 2\rho_t\sigma_{1t}\sigma_{2t}dt)dt\} \\
&+ \frac{1}{2}\sum_{i=1}^{n}(\frac{\partial \omega^{i-1,i}}{\partial a^2} + \frac{\partial \omega^{i,i-1}}{\partial a^2})\int_{t_{1,i-1}}^{t_{2,i-1}}\rho_t\sigma_{1t}\sigma_{2t}dt
\end{aligned}$$

where $\boldsymbol{J} = (J_{ij})$ where $J_{i-1,i} = 1$, and the other components of $\boldsymbol{J}$ is 0.

$$\begin{aligned}
tr\mathbf{\Omega}^{-1} - tr(\mathbf{\Omega}^{-1}\boldsymbol{J}) &= \frac{n(1+\eta)(1+\eta^{2n+1}) - \eta(1+\eta^{2n}) - \frac{2\eta^2(1-\eta^{2n-1})}{1-\eta}}{\gamma^2(1-\eta^2)(1-\eta^{2n+2})} \\
&= \frac{n}{2a^2} - \frac{\sqrt{Ta^2\sigma^2}}{4a^4}n^{\frac{1}{2}} + O(1) \qquad (2.11)
\end{aligned}$$

Let $\omega^m = \omega^{n^{2/3},n^{2/3}}$. Note that $\omega^{i,j} = \omega^{j,i}$, and for any $n^{2/3} \leq i \leq n - n^{2/3}$,

$$\begin{aligned}
\omega^{ii} &= \omega^m(1+o(1)) = \omega^{i,i-1}(1+o(1)) \\
\frac{\partial \omega^{ii}}{\partial a^2} &= \frac{\partial \omega^{i,i-1}}{\partial a^2}(1+o(1)) = -\frac{1}{2a^2}\omega^{ii}(1+o(1)) \\
\frac{\partial \omega^{ii}}{\partial \sigma^2} &= \frac{\partial \omega^{i,i-1}}{\partial \sigma^2}(1+o(1)) = -\frac{1}{2\sigma^2}\omega^{ii}(1+o(1))
\end{aligned}$$

For $i \leq n^{2/3}$ and $i \geq n - n^{2/3}$, they are dominated by ω^m, and the integration is over an interval which shrinks at the rate $n^{-1/3}$. So

$$\begin{aligned}
&\frac{1}{2}\sum_{i=1}^{n}\frac{\partial \omega^{ii}}{\partial a^2}(\int_{t_{1,i-1}}^{t_{1,i}}\sigma_{1t}^2 dt + \int_{t_{2,i-1}}^{t_{2,i}}\sigma_{2t}^2 dt + \int_{t_{2,i-1}}^{t_{1,i}} 2\rho_t\sigma_{1t}\sigma_{2t}dt) \\
&+ \frac{1}{2}\sum_{i=1}^{n}(\frac{\partial \omega^{i-1,i}}{\partial a^2} + \frac{\partial \omega^{i,i-1}}{\partial a^2})\int_{t_{1,i-1}}^{t_{2,i-1}}\rho_t\sigma_{1t}\sigma_{2t}dt \\
=&\frac{\partial \omega^m}{\partial a^2}\int_0^T(\sigma_{1t}^2 + \sigma_{2t}^2 + 2\rho_t\sigma_{1t}\sigma_{2t})dt(1+o_p(1)) \\
=&-\frac{\bar{n}^{\frac{1}{2}}(a^2)^{-\frac{3}{2}}}{4\sqrt{\sigma^2 T}}\int_0^T(\sigma_{1t}^2 + \sigma_{2t}^2 + 2\rho_t\sigma_{1t}\sigma_{2t})dt(1+o_p(1)) \qquad (2.12)
\end{aligned}$$

By calculation, we can also obtain

$$\begin{aligned}\bar{\Psi}_n^2 =&(\frac{1}{2a^2}-\frac{u_1^2}{2a^4})\\&+(\frac{3u_1^2\sqrt{\sigma^2T}}{8a^5}-\frac{\sqrt{\sigma^2T}}{4a^3}-\frac{\int_0^T(\sigma_{1t}^2+\sigma_{2t}^2+2\rho_t\sigma_{1t}\sigma_{2t})dt}{8a^3\sqrt{\sigma^2T}})\bar{n}^{-\frac{1}{2}}+o_p(\bar{n}^{-\frac{1}{2}})\end{aligned}$$

Hence, if we solve $\bar{\Psi}_n^2=0$, we get

$$a_n^{*2}=u_1^2+(\frac{3u_1^2\sqrt{\sigma_n^{*2}T}}{4a_n^*}-\frac{\sqrt{\sigma_n^{*2}T}a_n^*}{2}-\frac{a_n^*\int_0^T(\sigma_{1t}^2+\sigma_{2t}^2+2\rho_t\sigma_{1t}\sigma_{2t})dt}{4\sqrt{\sigma_n^{*2}T}})n^{-\frac{1}{2}}+o_p(\bar{n}^{-\frac{1}{2}})$$

On the other hand, let $\mathbf{\Gamma}$ be the symmetric trigonal matrix, whose i^{th} element on the diagonal is $\Gamma_i:=\int_{t_{1,i-1}}^{t_{1,i}}\sigma_{1t}^2dt+\int_{t_{2,i-1}}^{t_{2,i}}\sigma_t^2dt+\int_{t_{2,i-1}}^{t_{1,i}}2\rho_t\sigma_{1t}\sigma_{2t}dt-\sigma^2\bar{\Delta}$, and the element off the diagonal is $\int_{t_{1,i-1}}^{t_{2,i-1}}\rho_t\sigma_{1t}\sigma_{2t}dt$ then

$$\begin{aligned}&\bar{\Psi}_n^1(\sigma^2,a^2)\\=&\frac{1}{2\sqrt{\bar{n}}}\{tr(\mathbf{\Omega}^{-1}\frac{\partial\mathbf{\Omega}}{\partial\sigma^2})+\frac{\partial tr(\mathbf{\Omega}^{-1}\mathbf{\Sigma}_0)}{\partial\sigma^2}\}\\=&\frac{1}{2\sqrt{\bar{n}}}\{tr(\mathbf{\Omega}^{-1}\frac{\partial\mathbf{\Omega}}{\partial\sigma^2})+\frac{\partial tr(\mathbf{\Omega}^{-1}(\mathbf{\Omega}+(2\boldsymbol{I}-\boldsymbol{J}-\boldsymbol{J}')(a^2-u_1^2)+\mathbf{\Gamma}))}{\partial\sigma^2}\}\\=&\frac{1}{2\sqrt{\bar{n}}}\{\sum_{i=1}^n(\frac{\partial\omega^{i,i-1}}{\partial\sigma^2}+\frac{\partial\omega^{i,i-1}}{\partial\sigma^2})\Gamma_{i,i-1}+\sum_{i=1}^n\frac{\partial\omega^{ii}}{\partial\sigma^2}\Gamma_{ii}\}-\frac{\sqrt{T}}{8a^3\sigma}(a^2-u_1^2)+O_p(\bar{n}^{-\frac{1}{2}})\\=&-\frac{1}{8a\sigma^3\sqrt{T}}(\int_0^T\sigma_{1t}^2+\sigma_{2t}^2+2\rho_t\sigma_{1t}\sigma_{2t}dt-\sigma^2T)-\frac{\sqrt{T}}{8a^3\sigma}(a^2-u_1^2)+O_p(\bar{n}^{-\frac{1}{2}})\end{aligned}\tag{2.13}$$

Set $\bar{\Psi}_n^1(\sigma^2,a^2)=0$, that is,

$$\int_0^T\sigma_{1t}^2+\sigma_{2t}^2+2\rho_t\sigma_{1t}\sigma_{2t}dt-\sigma_n^{*2}T=-\frac{\sigma_n^{*2}T}{a_n^{*2}}(a_n^{*2}-u_1^2)+O_p(\bar{n}^{-\frac{1}{2}})=O_p(\bar{n}^{-\frac{1}{2}})$$

Therefore, all the desired equalities remain as in Xiu (2010), except that we replace a_0^2 with u_1^2 and $\int_0^T\sigma_t^2dt$ with $\int_0^T\sigma_{1t}^2+\sigma_{2t}^2+2\rho_t\sigma_{1t}\sigma_{2t}dt$. Thus, the identifiability condition holds, and hereby we have $\hat{\sigma}^2-\sigma_n^{*2}=o_p(1)$ and $\hat{a}^2-a_n^{*2}=o_p(1)$. The desired convergence rates follow from Taylor expansion of $\mathbf{\Psi}_n-\bar{\mathbf{\Psi}}_n$ as in the proof of Theorem 11.

Table 2.2: Simulation Comparisons of the Estimates

		$\hat{\Sigma}_{11}$	$\hat{\Sigma}_{12}$	$\hat{\Sigma}_{22}$
$\bar{\Delta} = 2s$	RC	0.1603	0.0583	0.0904
$n = 11700$		(0.0032)	(0.0018)	(0.0018)
	QMLE	0.1603	0.0584	0.0906
$\bar{n}_1 = 5853$		(0.0242)	(0.0112)	(0.0074)
$\bar{n}_2 = 7801$	MRK	0.1967	0.0584	0.0920
$\bar{n} = 5573$		(0.0466)	(0.0147)	(0.0147)
	TSCV	0.1565	0.0569	0.0883
		(0.0306)	(0.0148)	(0.0146)
$m_1 = 5151$	HY		0.0584	
$m_2 = 5307$			(0.0028)	
Span $= 1.92s$	QMLE*		0.0582	
			(0.0108)	
		$\hat{\Sigma}_{11}$	$\hat{\Sigma}_{12}$	$\hat{\Sigma}_{22}$
$\bar{\Delta} = 30s$	RC	0.1600	0.0581	0.0901
$n = 781$		(0.0121)	(0.0068)	(0.0068)
	QMLE	0.1606	0.0582	0.0904
$\bar{n}_1 = 392$		(0.0509)	(0.0241)	(0.0178)
$\bar{n}_2 = 522$	MRK	0.2116	0.0575	0.0923
$\bar{n} = 372$		(0.0732)	(0.0270)	(0.0264)
	TSCV	0.1504	0.0543	0.0847
		(0.0536)	(0.0256)	(0.0247)
$m_1 = 345$	HY		0.0581	
$m_2 = 355$			(0.0106)	
Span$= 28.7s$	QMLE*		0.0575	
			(0.0232)	

Note: This table reports the summary statistics for the simulation results. $\bar{\Delta}$ is the sampling frequency for the original data. n is the size of the whole sample. $\bar{n}_i$ is the average number of observations for the ith asset. $\bar{n}$ is the average number of synchronized observations using the Refresh Time scheme. m_1 and m_2 are the effective sample sizes for two series respectively, using the HY method. Span measures the average difference between the refresh time and the real observation time. $\hat{\Sigma}_{ij}$ is the component of the estimated covariance matrix. The data in parenthesis are the RMSE. RC is the benchmark method applied to the no noise and the whole data (Synchronized). The Monte Carlo sample size is 10000.

Table 2.3: Summary Statistics of the Log Returns of the Foreign Exchange Futures

FX	Avg No	Avg Freq	Mean	Std Err	1st Lag	2nd Lag	3rd Lag
AUD	6142	14.07s	9.97e-08	1.53e-04	−0.0769	0.0082	0.0031
CAD	5898	14.65s	2.25e-08	1.24e-04	−0.1186	0.0033	0.0007
EUR	17097	5.05s	2.16e-08	5.69e-05	−0.1141	0.0099	0.0014
GBP	10508	8.22s	1.35e-08	7.62e-05	−0.0651	0.0123	0.0037
JPY	11685	7.39s	−2.18e-08	8.64e-05	−0.1066	0.0089	0.0005
CHF	6284	13.75s	2.35e-08	1.07e-04	−0.0846	0.0067	−0.0025

Note: The second and third columns are respectively the average numbers of trades and average trading frequencies per day. The fourth and fifth column are the average log-returns and their association standards. The auto-correlation coefficients of lags 1, 2 and 3 are depicted at the sixth, seventh and eighth columns.

Table 2.4: Comparison Between Refresh Time and HY Synchronization Scheme

		AUD	CAD	EUR	GBP	JPY	CHF
Refresh Time	% of AUD		56.9%	87.2%	73.7%	70.0%	58.6%
	% of CAD	59.2%		86.6%	74.3%	74.9%	58.8%
	% of EUR	31.3%	21.7%		52.2%	53.7%	34.4%
	% of GBP	43.1%	41.7%	84.9%		64.3%	47.3%
	% of JPY	40.5%	37.8%	78.6%	57.9%		42.7%
	% of CHF	57.3%	55.1%	93.6%	79.1%	79.4%	
HY	% of AUD		59.1%	85.5%	73.5%	76.5%	60.8%
	% of CAD	61.3%		84.7%	74.0%	74.5%	60.9%
	% of EUR	35.0%	33.2%		53.4%	55.1%	37.6%
	% of GBP	46.4%	44.9%	81.6%		64.5%	50.0%
	% of JPY	44.3%	41.3%	76.8%	58.9%		46.0%
	% of CHF	59.7%	57.7%	90.7%	77.7%	78.1%	

Note: In this table, we compare the Refresh Time and HY synchronization scheme by calculating the percentage of data reserved.

Table 2.5: Summary of the Correlation Estimates

	AUD	CAD	EUR	GBP	JPY	CHF
AUD		0.2592	0.3783	0.3044	0.0061	0.2994
CAD	0.2584		0.2708	0.2169	0.0203	0.2282
EUR	0.3774	0.2715		0.4703	0.2138	0.6409
GBP	0.3011	0.2173	0.4724		0.1079	0.4051
JPY	0.0064	0.0201	0.2131	0.1083		0.2656
CHF	0.2955	0.2272	0.6475	0.4032	0.2635	

Note: This table reports the average correlations among the six currencies. The numbers on the upper diagonal are based on Refresh Time scheme, while the numbers on the lower diagonal are the average of two synchronization schemes, which in addition, require one asset to lead the other one.

Table 2.6: Comparisons of the correlation estimates

Active Markets	Asia, Australia, Europe	Europe, US, Canada	US, Canada
Chicago Time	5PM-7AM	7AM-11AM	11AM-4PM
AUD — CAD	0.2476	0.2766	0.2342
AUD — EUR	0.3735	0.4042	0.3153
AUD — GBP	0.3001	0.3233	0.2534
AUD — JPY	-0.0041	0.0486	-0.0321
AUD — CHF	0.3146	0.3114	0.2042
CAD — EUR	0.2618	0.2885	0.2314
CAD — GBP	0.2125	0.2294	0.1858
CAD — JPY	0.0122	0.0469	-0.0156
CAD — CHF	0.2298	0.2391	0.1747
EUR — GBP	0.4555	0.4948	0.4648
EUR — JPY	0.1801	0.2663	0.1939
EUR — CHF	0.6685	0.6341	0.5754
GBP — JPY	0.0692	0.1746	0.0929
GBP — CHF	0.3996	0.4306	0.3582
JPY — CHF	0.2198	0.3213	0.2641

Note: This table compares the correlations of each FX pairs in three sessions within a day.

CHAPTER 3

DISSECTING AND DECIPHERING EUROPEAN OPTION PRICES USING CLOSED-FORM SERIES EXPANSION

We seek a closed-form series approximation of European option prices under a variety of diffusion models. The proposed convergent series are derived using either the Hermite polynomial approach or the undetermined coefficients method. Departing from the usual option pricing routine in the literature, our model assumptions are fairly general, with no requirements for affine dynamics or explicit characteristic functions. Moreover, the closed-form expansions provide a distinct insight into how and on which order the model parameters affect option prices, in particular for close-to-maturity options. Such explicit formulae are advantageous over alternative numerical solutions of partial differential equations or simulation methods in regard to real-time calibration and hedging with contingent claims. With closed-form expansions, we explicitly translate model features into option prices, such as stochastic interest rate, mean-reverting drift, and self-exciting or skewed jumps. Numerical examples illustrate the accuracy of this approach.

3.1 Introduction

How are the drift, volatility and jump components of underlying risk neutral dynamics translated into option prices? On which order does each component matter as options approach expiration? Is it possible to separate their first-order contributions to option prices? Are there economic insight behind the impact of each component? These questions are important not only for financial engineers who are trying to construct effective models that better capture market movement, but also relevant to market participants who are hedging the risk of their portfolios with contingent claims. We address these questions by seeking closed-form series expansions of option prices, which fill the gap between closed-form solutions and numerical methods to option valuation.

Even for one-factor continuous-time dynamic models, closed-form option pricing formulae are very sparse. Most models with closed-form solutions are either confined to the log-normal class including the famous Black-Scholes-Merton formula (Black and Scholes (1973) and Merton (1976)), or the Bessel process class, e.g. CIR and CEV models discussed in Cox (1975), Cox et al. (1985) and Cox et al. (1976), or can be reduced to these classes after transformation (see Goldenberg (1991)). These formulae, most of which were derived decades ago, are still playing important roles in the financial industry nowadays. Recently, the Fourier-transform-based approach has evolved rapidly, substantially enlarging the class of models that have closed-form option pricing formulae. Explicit formulae have been derived for particular diffusions with stochastic volatility (Heston (1993)) and jumps (Bates (1996)), along with stochastic interest rates (Scott (1997)). Carr and Madan (1998) further develop this approach, extending the scope to general models driven by Lévy processes. Nevertheless, the Fourier-transform-based approach relies on explicit characteristic functions, naturally arising from Lévy-Khintchine representations, or on closed-form solutions to ordinary differential systems for affine jump diffusion models (Duffie et al. (2000)). Numerical methods of solving partial differential equations or simulation-based methods are developed to tackle option-valuation problems that do not have closed-form solutions. Unfortunately, when trying to estimate or calibrate these models, numerical methods are computationally expensive because optimization on top of these pricing algorithms is ex-

tremely time-consuming. In such circumstances, a closed-form expansion formula may simplify the task to a great extent.

In this paper, we complement the literature by providing a closed-form expansion for European-type option prices. Unlike numerical methods, the expansion approach offers a number of advantages. Instead of simply calculating a number, it gives more insight into how parameters influence prices and to what extent, which even closed-form solutions cannot offer. Furthermore, expansion formulae are smooth, so that differentiation becomes equally tractable, hence permitting real-time calibration and hedging. Comparative statics results are also straightforward to derive with closed-form formulae. The proposed method works for one-factor diffusion models, potentially inhomogeneous with jumps, and certain multivariate models, and allows nonlinear dynamics, which may not have explicit characteristic functions.

In the same spirit, Kimmel (2009) obtains explicit analytic series for bond prices under diffusion models. Nevertheless, an explicit expansion formula for option prices is more cumbersome to obtain, due to the non-smoothness of the payoff function at maturity. Recently, Kristensen and Mele (2010) propose an approach to approximating the option price by expanding the difference between the true model price and the Black-Scholes price. Their approach avoids the singularity with the help of an auxiliary closed-form pricing formula. Nevertheless, this expansion is less informative regarding how option prices are determined near expiration. In contrast, we propose a series expansion with a well-designed initial term that suffices to capture this singularity and the rest part approximated by power series. As a result, the structure of option prices becomes transparent. A similar strategy has been successfully applied for transition density or likelihood expansion for homogeneous diffusions in Aït-Sahalia (2002), Aït-Sahalia (2008) and Aït-Sahalia (1999), for jump diffusions (Yu (2007)), and inhomogeneous diffusions (Egorov et al. (2003)). The advantage of series expansion over alternative simulation-based approaches has been documented in Jensen and Poulsen (2002) and Hurn et al. (2007). Related works also include Aït-Sahalia and Kimmel (2009), Aït-Sahalia and Kimmel (2007), and Bakshi and Ju (2005).

To demonstrate the economic value of this approach, we provide closed-form expansions for a variety of models in asset pricing. These examples are designed to study

some specific features in the underlying dynamics, such as stochastic interest rate, mean-reverting drift, self-exciting or skewed jumps. Based on the closed-form expansions, the structure of options written on these models becomes transparent, shedding light on how these characteristics are translated into option prices and what the magnitude of their effects are.

The paper is organized as follows. Section 3.2 introduces two approaches to closed-form expansions for binary options. Section 3.3 discusses vanilla option pricing based on general time-inhomogeneous models, jump diffusion models and certain multivariate models. Section 3.4 provides several examples to show how to translate some features of underlying dynamics into option prices. Section 3.5 concludes. The appendix contains the mathematical proofs.

3.2 Closed-Form Expansion of Option Prices

A typical European-style claim offers the holder the right but not the obligation to either buy (Call) or sell (Put) a predetermined contingent payoff at maturity. For example, vanilla call options allow the holder to purchase the underlying security at the strike price if the exercise value of the underlying settles above it at expiration. As recent financial innovation is spanning the market, more exotic options are created and flourished, such as binary call options, whose exercise values are either some fixed amount of cash (cash-or-nothing) or the value of the underlying security itself (asset-or-nothing), if the option expires in-the-money. Unlike vanilla options, most binary options were only traded over-the-counter before June 2008, when the CBOE started offering the standardized exchange-traded cash-or-nothing binary options on the S&P 500 index and its implied volatility (VIX) with continuous quotations. In the following, we discuss explicit expansions for both vanilla and binary options.

We start with a simple case in which the underlying security X_t follows a scalar diffusion process and the interest rate is a fixed constant r. Suppose there exists an equivalent

martingale measure Q under which the dynamics of X_t satisfies:

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dW_t^Q \tag{3.1}$$

If X_t itself is a tradable asset, then $\mu(X_t;\theta) = rX_t$, but we do not need to impose this constraint here.

As is well known, European option price $\Psi(\Delta, x)$ with maturity T and strike K satisfy the Feynman-Kac partial differential equation,

$$-\frac{\partial\Psi(\Delta,x;\theta)}{\partial\Delta} + \mu(x;\theta)\frac{\partial\Psi(\Delta,x;\theta)}{\partial x} + \frac{1}{2}\sigma^2(x;\theta)\frac{\partial^2\Psi(\Delta,x;\theta)}{\partial x^2} - r\Psi(\Delta,x;\theta) = 0 \tag{3.2}$$

with the initial condition determined by the contingent payoff $f(x)$ at expiration:

$$\Psi(0,x;\theta) = f(x)1_{\{x>K\}} \tag{3.3}$$

where $\Delta = T - t$ denotes the time to maturity and x is the price level at time t.

Equivalently, the option price $\Psi(\Delta, x)$ can also be written as the discounted expectation of future payoff under risk neutral measure:

$$\Psi(\Delta,x) = e^{-r\Delta}E(f(X_T)1_{\{X_T>K\}}|X_t = x) = e^{-r\Delta}\int_{D_X\cap[K,\infty)} f(z)p_X(\Delta,z|x)dz \tag{3.4}$$

where $p_X(\Delta, z|x)$ denotes the state price density or transition density, which also satisfies (3.2). For convenience, the parameter vector θ is omitted without ambiguity, and the domain of X_t is denoted as D_X. Most financial models in practice have $D_X = (0,\infty)$, or $(-\infty,\infty)$.

Our goal is to develop a closed-form series expansion of the option prices in the time variable. Such expansions are able to demonstrate which factors are more important to option prices as time approaches the maturity. Apparently, a straightforward approach is to directly postulate an appropriate series and verify it by plugging it into the equation. For example, a Taylor series around $\Delta = 0$ is able to solve a similar equation for zero-coupon bond, as shown by Kimmel (2009). Unlike pricing a zero-coupon bond, the "hockey stick"

option payoff introduces singularity in the initial condition, which precludes the possibility of analytic expansion for the option price. Nevertheless, it is possible to propose an irregular leading term that suffices to "absorb" the singularity, leaving the rest part approximated by power series. We can imagine there may be too many candidates for the leading term. Without any theoretical guidance, seeking the right one (provided it is unique) is amount to find a needle in a haystack. It turns out that we can adopt a bottom-up strategy to build the entire expansion from scratch using the Hermite polynomials.

3.2.1 The Hermite Polynomials Approach

To introduce this method, we start with expanding the price of a cash-or-nothing binary call option with $f(x) = 1$. In fact, it is clear from Equation (3.4) that, barring from the time discounting, the option price is equal to the probability of expiring in-the-money, hence expanding it is equivalent to approximating the cumulative transition distribution function. Aït-Sahalia (2002) discusses the transition density expansion using the Hermite polynomials. These two problems are amount to solving the same partial differential equation with different initial conditions. The initial value of the transition density at $\Delta = 0$ is a δ-function, or Dirac mass, whereas for the distribution function, the initial value is a step function, or Heaviside function. Our strategy is to begin with expanding the transition density, and then integrate it to obtain the expansion for distribution function.

We follow similar steps as in Aït-Sahalia (2002). First, we transform the underlying process X to Y, whose transition density becomes "closer" to the density of normal distribution. Next, we perform another transformation from Y to Z, which is sufficiently "close" to a $N(0, 1)$ variable. We then find a convergent density expansion for Z with standard normal density serving as a leading term. Further, we obtain the density expansion for X by applying the Jacobian formula, and finally integrate it over the region $x > K$ to calculate the binary option price.

To fix ideas, we change the variable X to Y so that the diffusion term of Y is standard-

ized:

$$Y_t = \gamma(X_t) = \int_{\cdot}^{X_t} \frac{1}{\sigma(s)} ds$$

The lower bound of the integration does not play a role, and hence is omitted here. It is straightforward to find that Y_t satisfies

$$dY_t = \mu_Y(Y_t)dt + dW_t^Q$$

For any given $\Delta > 0$ and any given y in D_Y, we make the second transformation:

$$Z_{t+\Delta} = \frac{Y_{t+\Delta} - y}{\sqrt{\Delta}}$$

The above transformation makes it possible to approximate the density of Z given $Y_t = y$, to the Jth order in the following way:

$$p_Z^{(J)}(\Delta, z|y) = \phi(z) \sum_{j=0}^{J} \eta_j(\Delta, y) H_j(z)$$

where $\phi(z)$ is the standard normal density, and $H_j(z)$s are the Hermite polynomials satisfying: $H_0(z) = 1$, and for any $j \geq 1$,

$$H_j(z) = \phi(z)^{-1} \frac{d^j}{dz^j} \phi(z)$$

It is natural to come up with Hermite polynomials, as they are constructed using the standard normal density function, orthogonal to each other, and hence qualified to be a potential basis in the function space. Due to the orthogonality of $H_j(z)$, the coefficients $\eta_j(\Delta, y)$ is given by

$$\eta_j(\Delta, y) = \frac{1}{j!} \int_{-\infty}^{\infty} H_j(z) p_Z(\Delta, z|y) dz$$

Let $p_Y(\Delta, \omega|y)$ denote the transition density of $Y_{t+\Delta}|Y_t$, therefore the density of $Z_{t+\Delta} =$

z given $Y_t = y$ can also be expressed as:

$$p_Z(\Delta, z|y) = \Delta^{\frac{1}{2}} p_Y(\Delta, \Delta^{\frac{1}{2}} z + y|y)$$

Further, Jacobian formula yields the transition density of $X_{t+\Delta} = s$ given $X_t = x$:

$$p_X(\Delta, s|x) = \sigma(s)^{-1} \times \Delta^{-\frac{1}{2}} p_Z(\Delta, \Delta^{-\frac{1}{2}}(\gamma(s) - \gamma(x))|\gamma(x))$$

Since we have established the Jth order density expansion for $p_Z(\Delta, z|y)$, the above equation immediately provides the Jth order expansion for $p_X(\Delta, s|x)$, denoted as $p_X^{(J)}(\Delta, s|x)$.

Given transition density expansion, it is straightforward to construct the cumulative distribution function by integration, which also yields the binary option price $\Psi(\Delta, x)$ using the probability representation of the solution to (3.2). Loosely speaking, allowing the interchange of integral and infinite summation immediately yields the expansion of $\Psi(\Delta, x)$:

$$\begin{aligned}
&\Psi(\Delta, x) = e^{-r\Delta} \int_K^\infty p_X^{(\infty)}(\Delta, s|x) ds \\
=&e^{-r\Delta} \int_{\frac{\gamma(K)-\gamma(x)}{\sqrt{\Delta}}}^{\frac{\gamma(\infty)-\gamma(x)}{\sqrt{\Delta}}} p_Z^{(\infty)}(\Delta, z|\gamma(x)) dz \\
\sim&e^{-r\Delta} \sum_{j=0}^{\infty} \eta_j(\Delta, \gamma(x)) \int_{\frac{\gamma(K)-\gamma(x)}{\sqrt{\Delta}}}^{\frac{\gamma(\infty)-\gamma(x)}{\sqrt{\Delta}}} \phi(z) H_j(z) dz \\
\sim& - e^{-r\Delta} \Big(\Phi(z) + \phi(z) \sum_{j=0}^{\infty} \eta_{j+1}(\Delta, \gamma(x)) H_j(z) \Big) \Big|_{z=\frac{\gamma(K)-\gamma(x)}{\sqrt{\Delta}}}
\end{aligned}$$

where $\Phi(z)$ is the standard normal cumulative distribution function. The last step utilizes the fact that for $j \geq 1$,

$$\int \phi(z) H_j(z) = \frac{1}{\sqrt{2\pi}} \int \frac{d^j}{dz^j} [e^{-\frac{z^2}{2}}] dz = \frac{1}{\sqrt{2\pi}} \frac{d^{j-1}}{dz^{j-1}} [e^{-\frac{z^2}{2}}] = \phi(z) H_{j-1}(z) \tag{3.5}$$

and the conjecture that that the integral evaluated at the upper limit converges to zero. The following theorem guarantees that the proposed series is a legitimate candidate to approximating $\Psi(\Delta, x)$.

Theorem 14. *Under Assumptions 1-3 given in Appendix 3.6.1, there exists $\bar{\Delta} > 0$ (could be ∞), such that for every $\Delta \in (0, \bar{\Delta})$, the following sequence*

$$\Psi^{(J)}(\Delta, x) = e^{-r\Delta}\Big(\Phi(\frac{\gamma(x)-\gamma(K)}{\sqrt{\Delta}}) + \phi(\frac{\gamma(x)-\gamma(K)}{\sqrt{\Delta}})\sum_{j=0}^{J}(-1)^{j+1}\eta_{j+1}(\Delta,\gamma(x))H_j(\frac{\gamma(x)-\gamma(K)}{\sqrt{\Delta}})\Big)$$

converges to $\Psi(\Delta, x)$ uniformly in x over any compact set in D_X (and θ over parameter space Θ), where $\Psi(\Delta, x)$ solves the Feynman-Kac partial differential equation (3.2) with the initial condition $\Psi(0, x) = 1_{\{x>K\}}$, for any $K \in D_X$.

To obtain explicit forms of $\eta_j(\Delta, y)$, we may recycle those coefficients calculated in Aït-Sahalia (2002), (e.g. (4.4)-(4.9) on page 238). In fact,

$$\eta_j(\Delta, y) = \frac{1}{j!}E\Big(H_j(\Delta^{-\frac{1}{2}}(Y_{t+\Delta} - y))|Y_t = y\Big)$$

and for any polynomial $g(Y_{t+\Delta}, y)$, apply the Taylor series expansion in Δ:

$$E\Big(g(Y_{t+\Delta}, y)|Y_t = y\Big) = \sum_{i=0}^{I}\mathcal{L}_Y^i \cdot g(y, y)\frac{\Delta^i}{i!} + E\Big(\mathcal{L}_Y^{I+1} \cdot g(Y_{t+\Delta}, y)|Y_t = y\Big)\frac{\Delta^{I+1}}{(I+1)!}$$

where $\mathcal{L}_Y$ is the infinitesimal generator of the process Y_t. Therefore, the Taylor series of $\eta_j(\Delta, x)$ up to the Ith order can be calculated explicitly. If we gather all terms with respect to increasing orders of Δ, and let J go to ∞, we may obtain an alternative expression of the expansion:

$$\Psi(\Delta, x) = e^{-r\Delta}\Big(\Phi(\frac{\gamma(x)-\gamma(K)}{\sqrt{\Delta}}) + \phi(\frac{\gamma(x)-\gamma(K)}{\sqrt{\Delta}})\exp\big(-\int_{\gamma(K)}^{\gamma(x)}\mu_Y(s)ds\big)\sum_{j=0}^{\infty}\chi_j(\gamma(x),\gamma(K))\frac{\Delta^j}{j!}\Big) \tag{3.6}$$

where

$$\chi_0(y, \omega) = (y-\omega)^{-1}(\exp(\int_{\omega}^{y}\mu_Y(s)ds) - 1)$$

and for $j \geq 1$,

$$\chi_j(y,\omega) = j(y-\omega)^{-j-1}\int_\omega^y (u-\omega)^j(\lambda_Y(u)\chi_{j-1}(u,\omega) + \frac{1}{2}\frac{\partial^2\chi_{j-1}(u,\omega)}{\partial u^2})du$$
$$\lambda_Y(y) = -\frac{1}{2}(\mu_Y^2(y) + \frac{\partial\mu_Y(y)}{\partial y})$$

which can be calculated explicitly once and for all, using Mathematica.

The expansion formula immediately uncovers that the volatility function specified in the model plays the leading role in determining the option prices especially in the short run, and that the normal cumulative and density functions are sufficient to capture the entire singularity from the contingent payoff.

Theorem 14 establishes a convergent series expansion of the cash-or-nothing binary call option prices. Binary put options with payoff $1_{\{x<K\}}$ can be expanded similarly. In regard to vanilla options with payoff $(x-K)^+$, we may reduce the pricing problem to the one that has already been solved above. More specifically, the drift term in most financial models is of the affine form $\mu(x) = \kappa(\alpha - x)$. In such cases, we can decompose the contingent payoff $x - K$ at expiration to be $\alpha - K$ and $x - \alpha$. The former payoff can be replicated trivially using cash-or-nothing binary options. For the latter part, we can perform the transformation: $\Psi(\Delta, x) = (x-\alpha)\hat{\Psi}(\Delta, x)$. The Feynman Kac equation (3.2) is simplified to a similar one for $\hat{\Psi}(\Delta, x)$ with constant "interest rate" $\kappa + r$:

$$-\frac{\partial\hat{\Psi}(\Delta,x)}{\partial\Delta} + (\kappa(\alpha-x) + \frac{\sigma^2(x)}{x-\alpha})\frac{\partial\hat{\Psi}(\Delta,x)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2\hat{\Psi}(\Delta,x)}{\partial x^2} - (\kappa+r)\hat{\Psi}(\Delta,x) = 0 \tag{3.7}$$

with the desired initial condition $\hat{\Psi}(0,x) = 1_{\{x>K\}}$.

To fully explore the potential of the expansion method and pricing vanilla options under more complex models, we propose a more convenient approach, which we call the lucky guess approach or the method of undetermined coefficients.

3.2.2 The Lucky Guess Approach

Alternatively, inspired by the Hermite polynomials approach, we can postulate an appropriate series expansion form, plug it into the equation and solve for the coefficients directly. This top-down approach may produce the same convergent sequence, and is sometimes more effective and convenient for general models and payoff functions. Apparently, it would become a bold guess and most likely an unlucky one if the leading term is not derived from the first method. Therefore, the two proposed methods are always combined together.

To introduce this approach, we once again solve the partial differential equation (3.2) with the initial condition $\Psi(0,x) = 1_{\{x>K\}}$ first, leaving general cases aside for the moment. We postulate that the solution has the following form:

$$\Psi(\Delta,x) = e^{-r\Delta}\Big(\Phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}) + \frac{\sqrt{\Delta}}{\sqrt{2\pi}}\exp(-\frac{C^{(-2)}(x)}{\Delta})\sum_{k=0}^{\infty} C^{(k)}(x)\Delta^k\Big) \tag{3.8}$$

Even without insight from the Hermite polynomial approach, this expression is straightforward to come up with for this simple case. The discount factor $e^{-r\Delta}$ is added to account for the $-r\Psi(\Delta,x)$ term in (3.2), and this trick works whenever the coefficient in front of $\Psi(\Delta,x)$ in (3.2) is constant. The first term in brackets plays the role of smoothing out the singularity in the initial condition in that $\Phi(\cdot)$ is an infinitely smooth approximation of the Heaviside function, and its value depends on the sign of $C^{(-1)}(x)$, when Δ approaches zero. The second term in brackets resembles the transition density expansions given by Aït-Sahalia (2002).

Matching coefficients for terms with the same order of Δ, we have the following recursive equations:

$$C^{(-1)}(x) = \int_K^x \frac{1}{\sigma(s)}ds \tag{3.9}$$

$$C^{(-2)}(x) = \frac{1}{2}\Big(\int_K^x \frac{1}{\sigma(s)}ds\Big)^2 \tag{3.10}$$

For $k \geq -1$,

$$C^{(k+1)}(x)\Big(\frac{1}{2} + (k+1) + \mathcal{L}C^{(-2)}(x)\Big) + \sigma^2(x)\frac{dC^{(k+1)}(x)}{dx}\frac{dC^{(-2)}(x)}{dx} = \mathcal{L}C^{(k)}(x) \qquad (3.11)$$

and the boundary condition at $x = K$ is embedded in (3.11) due to $\frac{dC^{(-2)}(x)}{dx}|_{x=K} = 0$. $\mathcal{L}$ is the infinitesimal generator of the process X_t.

Theorem 15. *Suppose that $\Psi(\Delta, x)$ is the price of a cash-or-nothing binary option with strike K, underlying price level x, and time-to-maturity Δ. The coefficients in the expansion (3.8) for $\Psi(\Delta, x)$ are given explicitly by (3.9), (3.10) and for $k \geq -1$,*

$$\begin{aligned} C^{(k+1)}(x) = \int_K^x \Big\{ &(\sigma(s)C^{(-1)}(s))^{-1}\mathcal{L}C^{(k)}(s) \\ &\cdot \exp\Big(\int_x^s (\sigma(u)C^{(-1)}(u))^{-1}((k+2) + C^{-1}(u)\mathcal{L}C^{-1}(u))du\Big)\Big\}ds \end{aligned} \qquad (3.12)$$

It is easy to verify that the two expansion formulae (3.6) and (3.8) agree with each other. For more complicated models, we often use the first method to find the leading term, then postulate an appropriate expression and plug it into the equation. The next section provides extensions showing how to derive vanilla option price expansion formulae for more general models such as time inhomogeneous diffusions with potential jumps and certain multivariate models.

3.3 Extensions

3.3.1 Jump Diffusion Models

We investigate jump diffusion models in this section, in which jumps are supposed to be large and infrequent with daily observations. In practice, such models are employed to explain abnormally large daily returns due to economic catastrophes, unexpected news and other rare events. We assume that the underlying state variable in the risk neutral world follows a jump diffusion:

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t^Q + dZ_t \qquad (3.13)$$

where Z_t is a finite activity pure jump process with jump intensity $\lambda(X_t)$, and jump size density $\nu(z) : R \to [0, 1]$.

A European call option with payoff $f(X_T)1_{\{X_T>K\}}$ at maturity has price $\Psi(\Delta, x) : [0, T] \times \mathbb{R} \to \mathbb{R}$, which satisfies the following integro-differential equation:

$$\begin{aligned} 0 =& -\frac{\partial \Psi(\Delta, x)}{\partial \Delta} + \mu(x)\frac{\partial \Psi(\Delta, x)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 \Psi(\Delta, x)}{\partial x^2} - r(x)\Psi(\Delta, x) \\ &+ \lambda(x)\int_{-\infty}^{\infty}\Big(\Psi(\Delta, x+z) - \Psi(\Delta, x)\Big)\nu(z)dz \\ =&(-\frac{\partial}{\partial \Delta} + \mathcal{A} - r(x))\Psi(\Delta, x) \end{aligned} \tag{3.14}$$

with the initial condition

$$\Psi(0, x) = f(x)1_{\{x>K\}} \tag{3.15}$$

where the operator $\mathcal{A}$ is given by

$$\mathcal{A}g(x) = \mu(x)\frac{\partial g(x)}{\partial x} + \frac{1}{2}\frac{\partial^2 g(x)}{\partial x^2} + \lambda(x)\int_{-\infty}^{\infty}\Big(g(x+z) - g(x)\Big)\nu(z)dz$$

Hereafter, we denote the operator without the integral part as $\mathcal{L}$, and y^{inv} as the inverse of function y.

In this case, the Bayes' rule acts as a guide for finding an appropriate form. Note that

$$\begin{aligned} p(N_\Delta \geq 0|X_t = x) &= O(1) \\ p(N_\Delta = 1|X_t = x) &= O(\Delta) \\ p(N_\Delta > 1|X_t = x) &= o(\Delta) \end{aligned}$$

and by Bayes' rule,

$$\begin{aligned} &E(f(X_T)1_{\{X_T>K\}}|X_t = x) \\ =&\sum_{n=0}^{\infty} E(f(X_T)1_{\{X_T>K\}}|X_t = x, N_\Delta = n)P(N_\Delta = n|X_t = x) \end{aligned}$$

Therefore, as the option approaches expiration, jumps in the underlying variable occur on the order of $O(\Delta)$, so that the dominating term remains to be the part contributed by the conditional risk-neutral density on no jumps. The following form is hereby postulated:

$$\Psi(\Delta, x) = \Phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}) \sum_{k=0}^{\infty} B^{(k)}(x)\Delta^k + \sqrt{\Delta}\phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}) \sum_{k=0}^{\infty} C^{(k)}(x)\Delta^k + \sum_{k=1}^{\infty} D^{(k)}(x)\Delta^k \tag{3.16}$$

The first term resembles the previous case, where an additional series is attached for non-constant $r(x)$ and stochastic intensity $\lambda(x)$. The last two terms have been used for the transition density approximation by Yu (2007).

Taking the proposed formula into the equation (3.14), and matching coefficients in terms of Δ and $\phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})$, we can derive Theorem 16.

Theorem 16. *Assume $\Psi(\Delta, x)$ satisfies the option pricing equations (3.14) and (3.15). The coefficients in the expansion (3.16) satisfy the following restrictions:*

$$0 = B^{(0)} - f(x) \tag{3.17}$$

$$0 = (k+1)B^{(k+1)} + (r(x) + \lambda(x))B^{(k)} - \mathcal{L}B^{(k)} \tag{3.18}$$

$$0 = -C^{(0)}\left(1 + C^{(-1)}\mathcal{L}C^{(-1)}\right) - \sigma(x)C^{(-1)}\frac{dC^{(0)}(x)}{dx} + B^{(0)}\mathcal{L}C^{(-1)} + \sigma(x)\frac{dB^{(0)}(x)}{dx} \tag{3.19}$$

$$0 = -C^{(k+1)}\left(k+2 + C^{(-1)}\mathcal{L}C^{(-1)}\right) - \sigma(x)C^{(-1)}\frac{dC^{(k+1)}(x)}{dx} + \sigma(x)\frac{dB^{(k+1)}(x)}{dx} + (\mathcal{L} - r(x) - \lambda(x))C^{(k)} + B^{(k+1)}\mathcal{L}C^{(-1)} \tag{3.20}$$

$$0 = -D^{(1)} + \lambda(x)h_0(x,0) \tag{3.21}$$

$$0 = -(k+2)D^{(k+2)} + \lambda(x)\sum_{r=0}^{k+1}\frac{1}{(2r)!}\frac{\partial^{2r}h_{k+1-r}(x,\omega)}{\partial\omega^{2r}}|_{\omega=0}M_{2r} + (\mathcal{A} - r(x))D^{(k+1)} + \lambda(x)\sum_{r=0}^{k}\frac{1}{(2r)!}M_{2r}\frac{\partial^{2r}g_{k-r}(x,\omega)}{\partial\omega^{2r}}|_{\omega=0} \tag{3.22}$$

where

$$h_k(x,\omega) = \int_{(C^{(-1)})^{inv}(\omega)-x}^{\infty} B^{(k)}(x+z)\nu(z)dz$$
$$g_k(x,\omega) = C^{(k)}((C^{(-1)})^{inv}(\omega))\nu\Big((C^{(-1)})^{inv}(\omega)-x\Big)\Big(|\frac{dC^{(-1)}(x)}{dx}|_{(C^{(-1)})^{inv}(\omega)}\Big)^{-1}$$

and M_s is the sth moment of the standard normal.

Equations (3.19) and (3.20) are ordinary differential equations, which have similar closed-form solutions as in Theorem 15, and therefore omitted here. The other equations can be solved one after another by induction. For jump size distribution supported on the interval $[0,\infty)$, e.g. exponential distribution, the derivation is similar, but the formulae are more complicated as given in the Appendix 3.6.5.

From (3.16) and Theorem 16, we can distinguish the order of the impact by drift, volatility and jump components on the price of close-to-maturity vanilla options respectively:

Remark. *For vanilla call options with $f(x) = x - K$ under such jump diffusion model, we can deduce that the approximation up to the order $\Delta^{\frac{3}{2}}$ is*

$$\begin{aligned}\Psi(\Delta,x) =& \Phi\left(\Delta^{-\frac{1}{2}}\int_K^x \frac{1}{\sigma(s)}ds\right)\Big((x-K)+B^{(1)}(x)\Delta\Big)\\ &+(x-K)(\int_K^x \frac{1}{\sigma(s)}ds)^{-1}\phi\left(\Delta^{-\frac{1}{2}}\int_K^x \frac{1}{\sigma(s)}ds\right)\Delta^{\frac{1}{2}}+D^{(1)}(x)\Delta+O(\Delta^{\frac{3}{2}})\end{aligned}$$

where the drift and jumps are relevant to $B^{(1)}(x)$ and $D^{(1)}(x)$. Therefore, volatility determines the leading terms, followed by jumps and drift part which affect first order terms.

3.3.2 Time Inhomogeneous Diffusion Models

Time inhomogeneous scalar models, including Derman and Kani (1994), Dupire (1994), Rubinstein (1995) and Ho and Lee (1986), may be the most popular toolkits for practitioners. These models can be employed to perfectly match the cross sectional option prices or term structure of interest rate using a convenient binomial or trinomial tree approximation instead of specifying function forms for the drift or volatility. Nevertheless, Dumas et al. (1998) point out empirically that a naive model guided approach outperforms these tree

models in terms of out-of-sample prediction. The following series expansion approach is suitable for a generic parametric model, which may lead to better predictive results.

Specifically, suppose that the underlying state variable X_t satisfies

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t^Q \tag{3.23}$$

We consider a general pricing equation:

$$\frac{\partial \Psi(t,x)}{\partial t} + \mu(t,x)\frac{\partial \Psi(t,x)}{\partial x} + \frac{1}{2}\sigma^2(t,x)\frac{\partial^2 \Psi(t,x)}{\partial x^2} - r(t,x)\Psi(t,x) = 0 \tag{3.24}$$

with the terminal condition

$$\Psi(T,x) = 1_{\{x>K\}} \tag{3.25}$$

This general model nests a variety of one-factor derivative pricing models. For example, we can interpret $\Psi(t,x)$ as the price of a binary option based on time inhomogeneous models. Also, pricing any option with sufficiently smooth non-zero contingent payoff $f(x)$ can be included in this setting, by a simple transformation $\hat{\Psi}(t,x) = f(x)^{-1}\Psi(t,x)$, since $\hat{\Psi}(t,x)$ must satisfy a similar equation with the terminal condition (3.25). In addition, this model can be used to price interest rate derivatives if X_t is interpreted as the underlying factor that determines interest rate.

Following the similar idea introduced previously, we may directly postulate a formal series, and then determine coefficients by brutal calculations. Here we borrow some inspiration from transition density expansion based on the Hermite polynomials discussed in Egorov et al. (2003). Denote

$$Y_t = \gamma(t, X_t) = \int_{\cdot}^{X_t} \frac{1}{\sigma(t,u)}du$$

Let $\Delta = T - t$, and

$$Z_s = \frac{Y_s - y}{\sqrt{\Delta}}$$

So the transition density $p_X(\omega, s|x, t)$ can be expressed as

$$p_X(\omega, s|x, t) = \sigma(s, \omega)^{-1}\Delta^{-\frac{1}{2}} p_Z(\Delta^{-\frac{1}{2}}(\gamma(s, \omega) - \gamma(t, x)), s|\gamma(t, x), t)$$

with $p_Z(z, s|y, t)$ the transition density of Z_s given $Y_t = y$.

Under similar assumptions, the conditional density of Z_s can be approximated using the Hermite polynomials:

$$p_Z^{(J)}(z, s|y, t) = \phi(z) \sum_{k=0}^{J} \eta_j(s, y, t) H_k(z)$$

where

$$\eta_j(s, y, t) = \frac{1}{j!} \int_{-\infty}^{\infty} H_j(z) p_Z(z, s|y, t) dz$$

If $r(t, x) = r$ is constant, we can derive the Jth order closed-form expansion of $\Psi(t, x)$ as

$$\begin{aligned}\Psi^{(J)}(t, x) =& e^{-r\Delta}\Big(\Phi(\frac{\gamma(t, x) - \gamma(t + \Delta, K)}{\sqrt{\Delta}}) + \phi(\frac{\gamma(t, x) - \gamma(t + \Delta, K)}{\sqrt{\Delta}}) \\ & \cdot \sum_{j=0}^{J} (-1)^{j+1} \eta_{j+1}(t, \gamma(t, x), t + \Delta) H_j(\frac{\gamma(t, x) - \gamma(t + \Delta, K)}{\sqrt{\Delta}})\Big)\end{aligned}$$

Hereby, we postulate the following formula for solutions to general cases:

$$\Psi(t, x) = \Phi(\frac{C^{(-1)}(t, x)}{\sqrt{\Delta}}) \sum_{k=0}^{\infty} B^{(k)}(t, x)\Delta^k + \Delta^{\frac{1}{2}} \phi(\frac{C^{(-1)}(t, x)}{\sqrt{\Delta}}) \sum_{k=0}^{\infty} C^{(k)}(t, x)\Delta^k \qquad (3.26)$$

where an additional series is added to deal with an arbitrary function $r(t, x)$. By calculation, we can obtain:

Theorem 17. *Suppose that $\Psi(t, x)$ solves the pricing equation (3.24) with initial condition (3.25). The coefficients of the expansion formula (3.26) satisfy*

$$\begin{aligned}B^{(0)}(t, x) =& 1 \\ C^{(-1)}(t, x) =& \gamma(t, x) - \gamma(t + \Delta, K)\end{aligned}$$

$$C^{(0)}(t,x) = \int_K^x \Big\{(\sigma(t,\omega)C^{(-1)}(t,\omega))^{-1}\Big((\frac{\partial}{\partial t}+\mathcal{L})C^{(-1)}(t,\omega)\Big)\cdot \exp\Big(\int_x^\omega (\sigma(t,u)C^{(-1)}(t,u))^{-1}(1+C^{(-1)}(t,u)(\frac{\partial}{\partial t}+\mathcal{L})C^{(-1)}(t,u))du\Big)\Big\}d\omega$$

and for $k \geq 0$, $B^{(k)}$ and $C^{(k)}$satisfy the following induction equations:

$$\begin{aligned}
B^{(k+1)}(t,x) =& (k+1)^{-1}(\frac{\partial}{\partial t}+\mathcal{L}-r(t,x))B^{(k)}(t,x)\\
C^{(k+1)}(t,x) =& \int_K^x \Big\{(\sigma(t,\omega)C^{(-1)}(t,\omega))^{-1}\Big(B^{(k+1)}(t,\omega)(\frac{\partial}{\partial t}+\mathcal{L})C^{(-1)}(t,\omega)\\
&+\sigma(t,\omega)\frac{\partial B^{(k+1)}(t,\omega)}{\partial \omega}+(\frac{\partial}{\partial t}+\mathcal{L}-r(t,\omega))C^{(k)}(t,\omega)\Big)\\
&\cdot \exp\Big(\int_x^\omega (\sigma(t,u)C^{(-1)}(t,u))^{-1}((k+2)+C^{(-1)}(t,u)(\frac{\partial}{\partial t}\\
&+\mathcal{L})C^{(-1)}(t,u))du\Big)\Big\}d\omega
\end{aligned}$$

Remark. *Consider X_t to be a short-term interest rate process, then*

$$B(t,x;T) = \sum_{k=0}^{\infty} B^{(k)}(t,x)\Delta^k$$

satisfies the partial differential equation (3.24) with terminal condition $B(T,x;T) = 1$, hence $B(t,x;T)$ is the zero-coupon bond price with maturity T.

Kimmel (2009) discusses this type of expansion for Bond prices and provides a modification of the expansion that may be analytic in the time variable.

3.3.3 Multivariate Diffusion Models

In this section, we discuss the possibility of extending the previous procedure to multivariate models. Assume that underlying factors follow

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t^Q \tag{3.27}$$

where X_t is an n-dimensional vector of state variables and W_t is a n-dimensional independent Brownian motion under Q measure. The potential correlation between state variables are therefore embedded in the volatility matrix. Denote $V(x) = \sigma(x)\sigma(x)^\tau = (v_{ij}(x))$ to be

the covariance matrix. The Feynman-Kac partial differential equation is:

$$-\frac{\partial \Psi(\Delta, x)}{\partial \Delta} + \sum_{i=1}^{n} \mu_i(x) \frac{\partial \Psi(\Delta, x)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} v_{ij}(x) \frac{\partial^2 \Psi(\Delta, x)}{\partial x_i x_j} - r(x)\Psi(\Delta, x) = 0 \quad (3.28)$$

The first obstacle of extending previous procedure is the existence of $\gamma(X)$ in the multivariate case, which is guaranteed for univariate models. Aït-Sahalia (2008) defines the concept of reducibility: a diffusion X is said to be reducible if and only if there exists an invertible mapping γ, such that $Y_t = \gamma(X_t)$ satisfies:

$$dY_t = \mu_Y(Y_t)dt + dW_t^Q$$

It turns out that reducibility alone is insufficient for obtaining closed-form formulae for option prices as in the previous cases. In fact, if we follow the Hermite polynomials approach, the leading term of the density of Z should be standard multivariate normal, hence the leading term of binary cash-or-nothing option price is

$$\Psi^{(0)}(\Delta, x) = (2\pi\Delta)^{-\frac{n}{2}} \int dy_2 \dots dy_n \int_D dy_1 \exp(-\frac{\|\gamma(y) - \gamma(x)\|^2}{2\Delta}) \cdot |\det(\sigma^{-1}(y))|$$

where D denotes the region in which the option expires in the money. Unlike scalar cases, the multivariate integral may not be simplified to a one-dimensional integral by changing variables, owing to the fact that the domain of integration is rather complicated once transformed by $\gamma(\cdot)$. If x_1 is the price of the underlying that determines the option price, namely $D = 1_{\{y_1 > K\}}$, then degenerate cases that can be solved explicitly requires $V(x)$ to satisfy the following condition besides reducibility:

$$V(x) = \begin{pmatrix} v_{11}(x_1) & V_{12}(x) \\ V_{12}(x)^\tau & V_{22}(x) \end{pmatrix} \quad (3.29)$$

where $v_{11}(x_1)$ is a scalar that depends on x_1 alone. In such case, an appropriate solution

may have the following form:

$$\Psi(\Delta, x) = \Phi\left(\frac{1}{\sqrt{\Delta}}\int_K^{x_1}\frac{1}{\sqrt{v_{11}(u)}}du\right)\sum_{j=0}^{\infty}B^{(j)}(x)\Delta^j + \sqrt{\Delta}\phi\left(\frac{1}{\sqrt{\Delta}}\int_K^{x_1}\frac{1}{\sqrt{v_{11}(u)}}du\right)\sum_{j=0}^{\infty}C^{(j)}(x)\Delta^j$$

The formulae that determine the coefficients are similar to those given in Theorem 16, but in vector form:

Theorem 18. *For models with covariance matrices of the form (3.29) and option payoff at expiration $f(x)1_{\{x_1>K\}}$, the coefficients for option price expansion satisfy the following restrictions:*

$$\begin{aligned}
0 =& B^{(0)} - f(x)\\
0 =& (k+1)B^{(k+1)} + r(x)B^{(k)} - \mathcal{L}B^{(k)}\\
0 =& -C^{(0)}\left(1 + C^{(-1)}\mathcal{L}C^{(-1)}\right) - C^{(-1)}\frac{\partial C^{(-1)}}{\partial x^\tau}\cdot V(x)\cdot\frac{\partial C^{(0)}}{\partial x} + B^{(0)}(x)\mathcal{L}C^{(-1)}\\
&+ \frac{\partial C^{(-1)}}{\partial x^\tau}\cdot V(x)\cdot\frac{\partial B^{(0)}}{\partial x}\\
0 =& -C^{(k+1)}\left(k+2 + C^{(-1)}\mathcal{L}C^{(-1)}\right) - C^{(-1)}\frac{\partial C^{(-1)}}{\partial x^\tau}\cdot V(x)\cdot\frac{\partial C^{(k+1)}}{\partial x} + B^{(k+1)}\mathcal{L}C^{(-1)}\\
&+ \frac{\partial C^{(-1)}}{\partial x^\tau}\cdot V(x)\cdot\frac{\partial B^{(k+1)}}{\partial x} + (\mathcal{L} - r(x))C^{(k)}
\end{aligned}$$

where $C^{(-1)}(x) = \Delta^{-1/2}\int_K^{x_1} v_{11}(s)^{-1/2}ds$, $\mathcal{L} = \mu(x)^\tau\cdot\frac{\partial}{\partial x} + \frac{1}{2}Trace(V(x)\cdot\frac{\partial^2}{\partial x\partial x^\tau})$, and $\cdot$ denotes matrix multiplication.

In regard to irreducible cases, the leading terms may not be normal cumulative distribution function and its density. In fact, these leading terms are closely related to the concept of geodesics and Riemannian manifolds in differential geometry. In general multivariate cases, the partial differential equation may correspond to nontrivial manifold such as Poincaré hyperbolic surface, whose geodesics could have rather complex structure. Asymptotic expansions are derived in such cases to approximate option prices, see e.g. Henry-Labordere (2005) and Hagan et al. (2005) for asymptotic expansions under some stochastic volatility models.

3.4 Translating Underlying Model Structures into Option Prices

We provide several examples of series expansion for option prices in this section, in order to shed light on some insight that the expansion formulae offer. The corresponding figures illustrate the accuracy of approximation by comparing the expansion formulae with closed-form solutions or Monte Carlo simulations. With parameters that are calibrated from the market prices, the expansion formulae approximate the true prices very well.

3.4.1 The Benchmark Model

We at first provide the expansions for the benchmark Black-Scholes model.

Example 1 (Black-Scholes Model)**.** *In the Black-Scholes model, the stock price follows:*

$$dX_t = (r-\delta)X_t dt + \sigma X_t dW_t^Q$$

The vanilla option price has a closed-form expression given by Black and Scholes (1973):

$$\Psi(\Delta, x) = xe^{-\delta\Delta}\Phi(d_1) - Ke^{-r\Delta}\Phi(d_2)$$

where

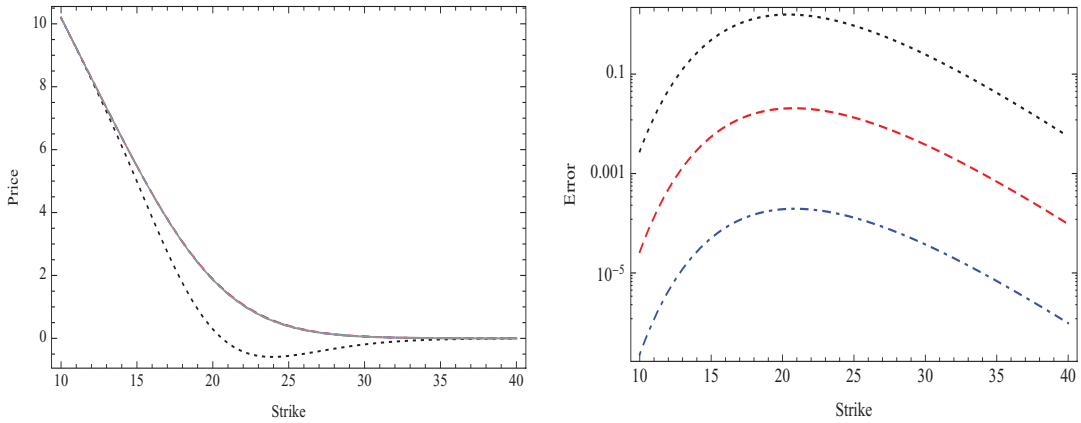
$$\begin{aligned} d_1 =& \frac{\log(\frac{x}{K}) + (r - \delta + \frac{1}{2}\sigma^2)\Delta}{\sigma\sqrt{\Delta}} \\ d_2 =& d_1 - \sigma\sqrt{\Delta} \end{aligned}$$

The following closed-form expansions are plotted in Figure 3.1.

$$\begin{aligned} \Psi(\Delta, x) =& \Phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} B^{(k)}(x)\Delta^k + \sqrt{\Delta}\phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} C^{(k)}(x)\Delta^k \\ B^{(k)}(x) =& \frac{(-1)^k}{k!}(x\delta^k - Kr^k), k \geq 0 \\ C^{(-1)}(x) =& \frac{\log\left(\frac{x}{K}\right)}{\sigma} \\ C^{(0)}(x) =& \frac{\sigma(x-K)}{\log\left(\frac{x}{K}\right)}, \textit{if } x \neq K; \textit{or } K\sigma, \textit{if } x = K. \end{aligned}$$

$$C^{(1)}(x) = \frac{\sigma\left((K-x)\sigma^2 + K^{\frac{1}{2}+\frac{r}{\sigma^2}-\frac{\delta}{\sigma^2}} x^{\frac{1}{2}-\frac{r}{\sigma^2}+\frac{\delta}{\sigma^2}} \sigma^2 \log\left(\frac{x}{K}\right) + (Kr - x\delta)\log\left(\frac{x}{K}\right)^2\right)}{\log\left(\frac{x}{K}\right)^3}, \text{ if } x \neq K; \text{ or}$$
$$K\left(\frac{(r-\delta)^2}{2\sigma} - \frac{(r+\delta)\sigma}{2} - \frac{\sigma^3}{24}\right), \text{ if } x = K.$$

Figure 3.1: Black-Scholes Model



Note: The black dotted line, red dashed line and blue dotted-dash line illustrate the $O(\Delta^{1/2})$, $O(\Delta^{3/2})$ and $O(\Delta^{5/2})$ order approximations respectively. The grey line denotes the true prices. Y-axis of the right panel is on a logarithmic scale. The parameters are: $\sigma = 0.2, r = 4\%, \delta = 0.01, x = 20$, and $\Delta = 1$.

3.4.2 The Role of Elasticity of Variance

A direct generalization of Black-Scholes model is the following Constant Elasticity of Variance (CEV) model which introduces one additional parameter γ. The elasticity of variance with respect to the state variable X is $2\gamma - 2$. When $\gamma \to 1$, the expansion formulae match the above ones.

Example 2 (CEV model)**.** *The CEV model assumes that*

$$dX_t = (r-\delta)X_t dt + \sigma X_t^{\gamma} dW_t^Q$$

We consider the case with $\gamma > 1$, so that the probability of hitting the boundary is 0. The closed-form

pricing formula is available in Cox (1975):

$$\Psi(\Delta, x) = xe^{-\delta\Delta}(1 - \chi^2(c, -b, a)) - Ke^{-r\Delta}\chi^2(a, 2 - b, c)$$

where $\chi^2(x, d, n)$ denotes the cumulative distribution function of non-central χ^2 distribution with noncentrality parameter n and degree of freedom d, and

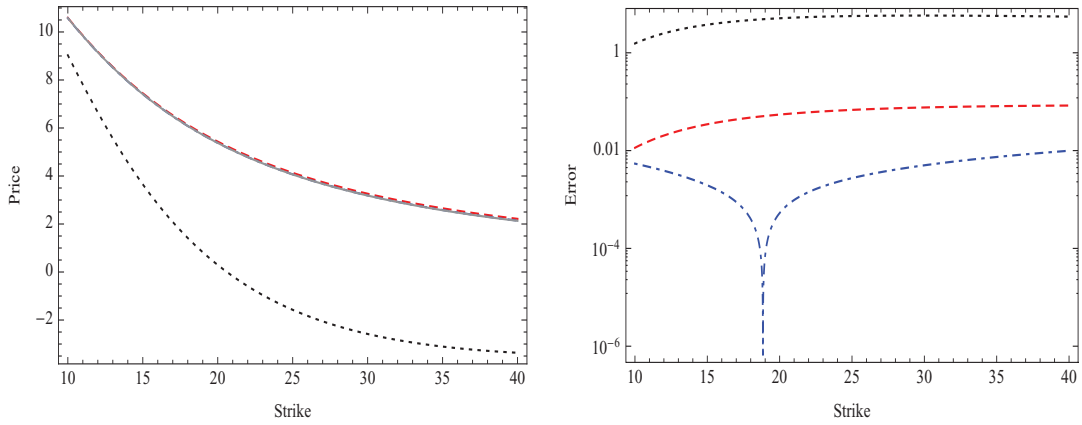
$$\begin{aligned} a =& \frac{(Ke^{-(r-\delta)\Delta})^{2(1-\gamma)}}{(1-\gamma)^2 v}, \qquad b = \frac{1}{1-\gamma}, \qquad c = \frac{x^{2(1-\gamma)}}{(1-\gamma)^2 v}, \\ v =& \frac{\sigma^2}{2(r-\delta)(\gamma-1)}(e^{2(r-\delta)(\gamma-1)\Delta} - 1) \end{aligned}$$

The following closed-form expansions are plotted in Figure 3.2.

$$\begin{aligned} \Psi(\Delta, x) =& \Phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} B^{(k)}(x)\Delta^k + \sqrt{\Delta}\phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} C^{(k)}(x)\Delta^k \\ B^{(k)}(x) =& \frac{(-1)^k}{k!}(x\delta^k - Kr^k), k \geq 0 \\ C^{(-1)}(x) =& -\frac{K^{1-\gamma} - x^{1-\gamma}}{\sigma - \gamma\sigma} \\ C^{(0)}(x) =& \frac{K^\gamma(K-x)x^\gamma(-1+\gamma)\sigma}{K^\gamma x - Kx^\gamma}, \text{ if } x \neq K; \text{ or } K^\gamma\sigma, \text{ if } x = K. \\ C^{(1)}(x) =& \frac{(Kx)^\gamma(-1+\gamma)\sigma}{(-K^\gamma x + Kx^\gamma)^3}\Bigg(K^{1+2\gamma}rx^2 + K^3rx^{2\gamma} - K^{2\gamma}x^3\delta - K^2x(2r(Kx)^\gamma + x^{2\gamma}\delta) \\ & + e^{\frac{(Kx)^{-2\gamma}(K^{2\gamma}x^2 - K^2x^{2\gamma})(r-\delta)}{2(-1+\gamma)\sigma^2}}K^{1+\frac{3\gamma}{2}}x^{5\gamma/2}(-1+\gamma)\sigma^2 - e^{\frac{(Kx)^{-2\gamma}(K^{2\gamma}x^2 - K^2x^{2\gamma})(r-\delta)}{2(-1+\gamma)\sigma^2}}K^{5\gamma/2} \\ & x^{1+\frac{3\gamma}{2}}(-1+\gamma)\sigma^2 - x(Kx)^{2\gamma}(-1+\gamma)^2\sigma^2 + K(Kx)^\gamma(2x^2\delta + (Kx)^\gamma(-1+\gamma)^2\sigma^2)\Bigg), \text{if } x \neq K; \text{or} \\ & \left(\frac{K^{2-\gamma}(r-\delta)^2}{2\sigma} - \frac{K^\gamma(r+\delta)\sigma}{2} + \frac{K^{-2+3\gamma}(-2+\gamma)\gamma\sigma^3}{24}\right), \text{if } x = K. \end{aligned}$$

The elasticity parameter γ appears in every order of the expansion, as the volatility exerts an inherent influence on option prices. Schroder (1989) develops an algorithm to approximate the option prices based on this CEV model, whereas the above closed-form formulae offer an alternative method, which is much simpler.

Figure 3.2: CEV Model



Note: The black dotted line, red dashed line and blue dotted-dash line illustrate the $O(\Delta^{1/2})$, $O(\Delta^{3/2})$ and $O(\Delta^{5/2})$ order approximations respectively. The grey line denotes the true prices. Y-axis of the right panel is on a logarithmic scale. The parameters are: $\sigma = 0.2, r = 4\%, \delta = 0.01, x = 20, \Delta = 1$, and $\gamma = 1.4$.

3.4.3 The Influence of Stochastic Interest Rate

It implies immediately from the above formulae that a constant interest rate influences option prices on the order of $O(\Delta)$. This is intuitive as interest rate appears in the discount factor, which needs time to take effect. A natural question that then arises is how stochastic interest rate affects option prices. We compare the benchmark model with the following two-factor model which features stochastic interest rate.

Example 3 (Stochastic Interest Rate)**.** *Assuming that the stock price X_t and interest rate r_t satisfy a 2-dimensional diffusion model driven by two independent Brownian motions:*

$$\begin{aligned} dX_t &= (r_t - \delta)X_t dt + \sigma X_t dW_t \\ dr_t &= \beta(\alpha - r_t) + \kappa\sqrt{r_t} dB_t \end{aligned}$$

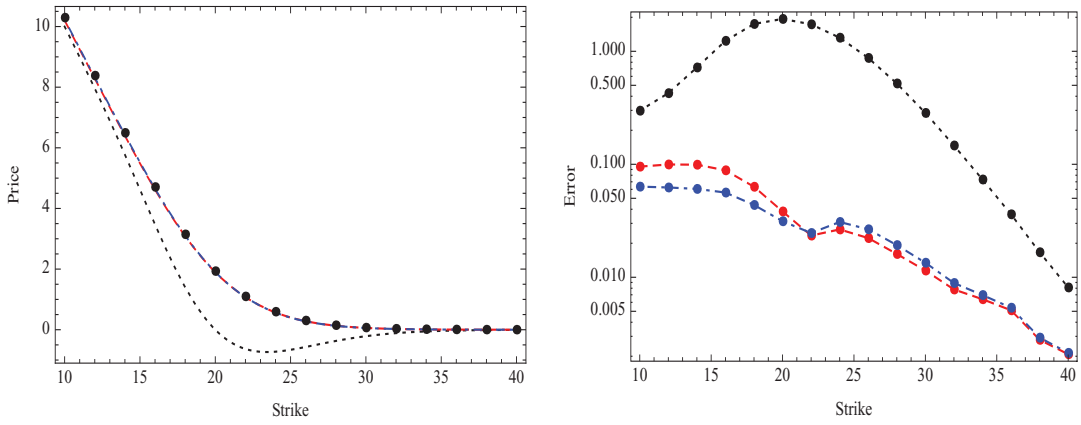
The following closed-form expansions are plotted in Figure 3.3.

$$\Psi(\Delta, x, r) = \Phi\left(\frac{C^{(-1)}(x,r)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} B^{(k)}(x,r)\Delta^k + \sqrt{\Delta}\phi\left(\frac{C^{(-1)}(x,r)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} C^{(k)}(x,r)\Delta^k$$

$$B^{(0)}(x,r) = x - K$$

$$
\begin{aligned}
B^{(1)}(x,r) =& Kr - x\delta \\
B^{(2)}(x,r) =& -\frac{1}{2}\left(Kr^2 + Kr\beta - K\alpha\beta - x\delta^2\right) \\
C^{(-1)}(x,r) =& \frac{\log\left(\frac{x}{K}\right)}{\sigma} \\
C^{(0)}(x,r) =& \frac{\sigma\left(x-K\right)}{\log\left(\frac{x}{K}\right)}, \textit{if } x \neq K; \textit{or } K\sigma, \textit{if } x = K. \\
C^{(1)}(x,r) =& \frac{\sigma\left((K-x)\sigma^2 + K^{\frac{1}{2}+\frac{r}{\sigma^2}-\frac{\delta}{\sigma^2}} x^{\frac{1}{2}-\frac{r}{\sigma^2}+\frac{\delta}{\sigma^2}} \sigma^2 \log\left(\frac{x}{K}\right) + (Kr - x\delta)\log\left(\frac{x}{K}\right)^2\right)}{\log\left(\frac{x}{K}\right)^3}, \textit{if } x \neq K; \textit{or} \\
& K\left(\frac{(r-\delta)^2}{2\sigma} - \frac{(r+\delta)\sigma}{2} - \frac{\sigma^3}{24}\right), \textit{if } x = K.
\end{aligned}
$$

Figure 3.3: Stochastic Interest Rate Model



Note: The black dotted line, red dashed line and blue dotted-dash line illustrate the $O(\Delta^{1/2})$, $O(\Delta^{3/2})$ and $O(\Delta^{5/2})$ order approximations respectively. The black dots denote the true prices. Y-axis of the right panel is on a logarithmic scale. The parameters are: $\sigma = 0.2, \alpha = 0.07, \kappa = 0.02, \beta = 0.26, r = 4\%, \delta = 0.01, x = 20$, and $\Delta = 1$. The number of sample path is 5000 in Monte Carlo, each of which is simulated using Euler's method with 2520 time intervals.

Apparently, the stochastic characteristics (β, κ) are not revealed up to the order of $O(\Delta^2)$, rendering the model indistinguishable from the benchmark, unless the current interest rate level deviates from its long-run mean ($r \neq \alpha$).

3.4.4 The Effect of Mean-Reversion

Mean-reversion is a common phenomenon in financial markets. For example, empirical evidence in the literature has pointed out that volatility and short rates dynamics often

exhibit such feature. When it comes to pricing options written on them, the mean-reversion feature may be carried over to their risk neutral dynamics, since neither volatility nor short rates are directly tradable assets.

Does mean-reversion in the underlying state variable affect short-term options written on it? Consider a call option written on the volatility dynamics. Intuitively, as volatility drifts away above its long-run mean, it has a stronger tendency of moving downwards, decreasing its potential future higher value. Therefore, out-of-the-money options in such cases may be traded at a discount. When volatility is below its long-run mean level, in-the-money options may be traded at a premium.

Identifying this effect is not an easy task, not to mention measuring the exact amount of premium. It is possible, however, to characterize it using closed-form expansions. Intuitively, since it takes time for volatility to return to its long-run mean level, (in other words, the mean-reversion effect needs time to manifest), it is likely that the mean-reversion would not impact the leading order of the option prices. In fact, this is evident from the expansion for cash-or-nothing binary option prices, which also reflect the probability of moving above the pre-specified strike levels.

We expand the cash-or-nothing binary volatility option with the Square-Root (SQR) model to demonstrate the mean-reversion effect. This type of volatility model has been considered in e.g. Heston (1993), Pan (2002) and Mencia and Sentana (2009).

Example 4 (Square-Root model)**.** *The SQR model for volatility assumes that*

$$dV_t = \beta(\alpha - V_t)dt + \sigma\sqrt{V_t}dW_t^Q$$

The true transition density is given by Cox et al. (1985), from which we can calculate the binary option price by integration:

$$\Psi(\Delta, v) = \int_K^\infty p(u, \Delta|v)du$$

where

$$p(u,\Delta|v) = ce^{-y-x}\Big(\frac{x}{y}\Big)^{\frac{q}{2}} I_q\Big(2\sqrt{xy}\Big), \qquad c = \frac{2\beta}{\sigma^2(1-e^{-\beta\Delta})}, \qquad x = cu, \qquad y = cve^{-\beta\Delta}$$

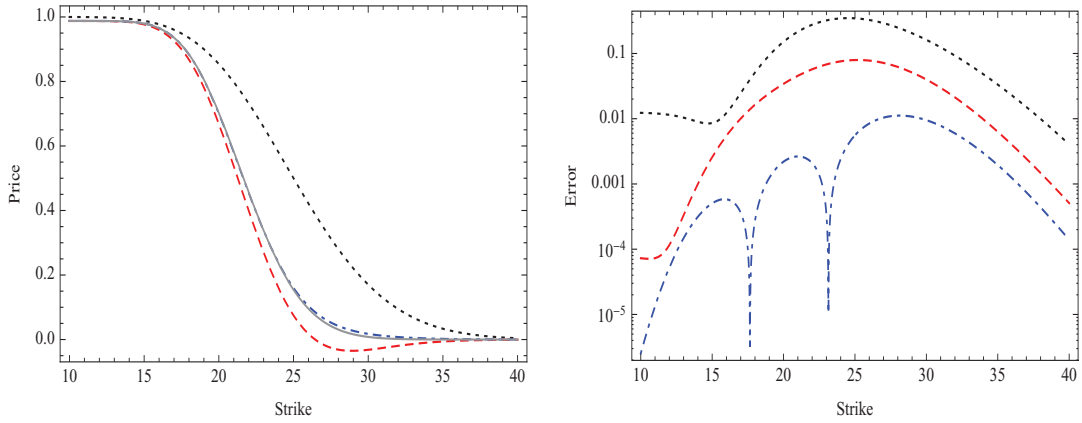
and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q.

Closed-form expansions for a binary call option with payoff $1_{\{V_t>K\}}$ are given below. Figure 3.4 illustrates the accuracy of expansions.

$$\begin{aligned}
\Psi(\Delta,v) =& \Phi\left(\frac{C^{(-1)}(v)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} B^{(k)}(v)\Delta^k + \sqrt{\Delta}\phi\left(\frac{C^{(-1)}(v)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} C^{(k)}(v)\Delta^k \\
B^{(k)}(v) =& \frac{(-1)^k r^k}{k!} \\
C^{(-1)}(v) =& \frac{2\big(\sqrt{v}-\sqrt{K}\big)}{\sigma} \\
C^{(0)}(v) =& -\frac{e^{-\frac{K\beta}{\sigma^2}} v^{-\frac{\alpha\beta}{\sigma^2}}\Big(-e^{\frac{v\beta}{\sigma^2}} K^{\frac{\alpha\beta}{\sigma^2}} v^{1/4} + e^{\frac{K\beta}{\sigma^2}} K^{1/4} v^{\frac{\alpha\beta}{\sigma^2}}\Big)\sigma}{2K^{1/4}\Big(\sqrt{K}-\sqrt{v}\Big)}, \text{if } v \neq K; \text{or } \frac{-4K\beta+4\alpha\beta-\sigma^2}{4\sqrt{K}\sigma}, \text{if } v = K. \\
C^{(1)}(v) =& \frac{e^{\frac{v\beta}{\sigma^2}} v^{\frac{1}{4}-\frac{\alpha\beta}{\sigma^2}}}{192(\sqrt{K}-\sqrt{v})^2\sigma}\Bigg(e^{-\frac{K\beta}{\sigma^2}} K^{-\frac{3}{4}+\frac{\alpha\beta}{\sigma^2}}\Big(-16K^2\beta^2+48\alpha^2\beta^2-3\sigma^4+48K(2\alpha\beta^2-(2r+\beta)\sigma^2)\Big) \\
& - \frac{1}{K^{1/4}(-\sqrt{K}v+v^{3/2})}\Big(24e^{-\frac{v\beta}{\sigma^2}} K^{1/4} v^{\frac{3}{4}+\frac{\alpha\beta}{\sigma^2}}\sigma^2(4Kr-8\sqrt{K}r\sqrt{v}+4rv+\sigma^2) \\
& - e^{-\frac{K\beta}{\sigma^2}} K^{\frac{\alpha\beta}{\sigma^2}}\sqrt{v}(\sqrt{v}(16v^2\beta^2-96v\alpha\beta^2-48\alpha^2\beta^2+96rv\sigma^2+48\alpha\beta\sigma^2+15\sigma^4) \\
& + \sqrt{K}(-16v^2\beta^2+48\alpha^2\beta^2-48\alpha\beta\sigma^2+9\sigma^4+96v(\alpha\beta^2-r\sigma^2)))\Big)\Bigg), \text{if } v \neq K; \text{or} \\
& \frac{1}{384K^{3/2}\sigma^3}\Bigg(64K^3\beta^3-64\alpha^3\beta^3+48\alpha^2\beta^2\sigma^2+4\alpha\beta\sigma^4-3\sigma^6-48K^2\beta\Big(4\alpha\beta^2-(8r+\beta)\sigma^2\Big) \\
& + 12K\Big(4\alpha\beta-\sigma^2\Big)\Big(4\alpha\beta^2-(8r+\beta)\sigma^2\Big)\Bigg) \\
& \text{if } v = K.
\end{aligned}$$

The $O(1)$ order term $\Phi\left(\frac{2\left(\sqrt{v}-\sqrt{K}\right)}{\sigma\sqrt{\Delta}}\right)$ in the expansion reflects the effect of moneyness. If the current volatility level v is higher than the strike price K, the chance of expiring in-the-money is dominant, otherwise the probability of expiring out-of-the-money is larger. When the option is at-the-money, the chance of moving upward or downward is equal up to the $O(1)$ order. This term therefore represents the probability as if the underlying volatility process were not mean-reverting. The Δ appeared in the denominator controls the first order temporal effect, that is, the shorter the time-to-maturity is, the more weight

Figure 3.4: Square-Root Model



Note: The black dotted line, red dashed line and blue dotted-dash line illustrate the $O(\Delta^{1/2})$, $O(\Delta^{3/2})$ and $O(\Delta^{5/2})$ order approximations respectively. The grey line denotes the true prices. Y-axis of the right panel is on a logarithmic scale. The parameters are: $\sigma = 0.2, \alpha = 0.2, \beta = 4, r = 5\%, v = 0.25$, and $\Delta = 90/360$. The current level of the VIX is equal to $100 \cdot v$.

the moneyness effect accounts for. As Δ decreases to zero, the leading term approaches to the indicator function.

The effect of mean-reversion is manifested at the $O(\Delta^{\frac{1}{2}})$ order for cash-or-nothing binary options. The effect is more transparent regarding at-the-money options. Fixing K at the current variance level v, the probability becomes increasingly large as α increases due to the pulling effect of mean-reversion when $\beta > 0$. The term $-\sigma/(4K^{1/2})$ is a higher order adjustment to the $O(1)$ term that may have nothing to do with mean-reversion effect. Excluding this adjustment term, simple calculation derives that as long as $\alpha > (v - K)/(\log v - \log K)$, the option may be traded at a premium due to mean-reversion, otherwise, the option may be traded at a discount. Moreover, the tradeoff level for α is always between v and K regardless of whether the option is in-the-money or out-of-the-money.

Similarly, for asset-or-nothing binary options, the mean-reversion effect also appears in $C^{(0)}$ on the order of $O(\Delta^{\frac{1}{2}})$, although the effect is more intricate since the payoff at expiration is also involved. More interestingly, when it comes to vanilla options, the mean-reversion effect by each component cancels out on the order of $O(\Delta^{\frac{1}{2}})$. The same is true for the Double Mean-Reversion (DMR) model, which includes an additional stochastic factor

for the mean level in the SQR model, as discussed in Amengual (2008), Mencia and Sentana (2009), and Egloff et al. (2010).

Example 5 (Double Mean-Reversion model)**.** *The DMR model for volatility assumes that*

$$\begin{aligned} dV_t =& \beta(y_t - V_t)dt + \sigma\sqrt{V_t}dW_t^Q \\ dy_t =& \xi(\alpha - y_t)dt + \kappa\sqrt{y_t}dB_t^Q \end{aligned}$$

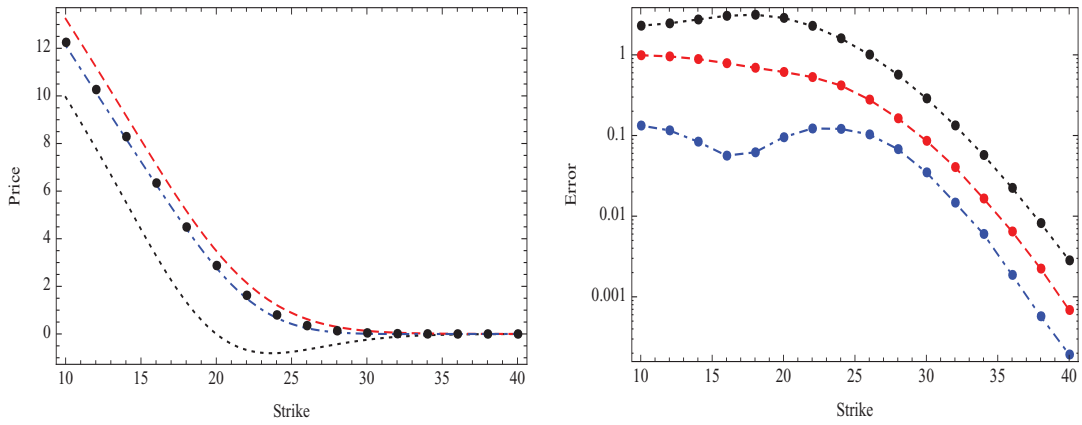
where $E(dW_t \cdot dB_t) = 0$.

The following closed-form expansions for a vanilla call option are plotted in Figure 3.5.

$$\begin{aligned} \Psi(\Delta, v, y) =& \Phi\left(\frac{C^{(-1)}(v,y)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} B^{(k)}(v,y)\Delta^k + \sqrt{\Delta}\phi\left(\frac{C^{(-1)}(v,y)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} C^{(k)}(v,y)\Delta^k \\ B^{(0)}(v,y) =& v - K \\ B^{(1)}(v,y) =& -r(v-K) - (v-y)\beta \\ B^{(2)}(v,y) =& -\frac{1}{2}\Big(Kr^2 - r^2v + 2r(-v+y)\beta + \beta(-\alpha\xi - v\beta + y(\beta+\xi))\Big) \\ C^{(-1)}(v,y) =& \frac{2(\sqrt{v}-\sqrt{K})}{\sigma} \\ C^{(0)}(v,y) =& \frac{\sigma(\sqrt{v}+\sqrt{K})}{2} \\ C^{(1)}(v,y) =& -\frac{1}{8(\sqrt{K}-\sqrt{v})^2}e^{-\frac{K\beta}{\sigma^2}}v^{-\frac{y\beta}{\sigma^2}}\sigma\Bigg(-2e^{\frac{v\beta}{\sigma^2}}K^{\frac{1}{4}+\frac{y\beta}{\sigma^2}}v^{1/4}\sigma^2 + e^{\frac{K\beta}{\sigma^2}}v^{\frac{y\beta}{\sigma^2}}\Big(4K^{3/2}r - 4Kr\sqrt{v} + \sqrt{v}\Big(4rv \\ & + 4v\beta - 4y\beta + \sigma^2\Big) + \sqrt{K}\Big(-4rv - 4v\beta + 4y\beta + \sigma^2\Big)\Big)\Bigg), \text{if } v \neq K; \text{ or} \\ & \frac{16K^2\beta^2 + 16y^2\beta^2 - \sigma^4 - 16K\Big(2y\beta^2 + (2r+\beta)\sigma^2\Big)}{32\sqrt{K}\sigma}, \text{if } v = K. \end{aligned}$$

The parameter β governs the speed of mean-reversion towards the short-run stochastic mean level y. Both β and y appear on the order of $O(\Delta)$, whereas the long-run mean-reversion rate, ξ, and the long-run mean level α influence the option prices on the order of $O(\Delta^2)$.

Figure 3.5: Double Mean-Reversion Model



Note: The black dotted line, red dashed line and blue dotted-dash line illustrate the $O(\Delta^{1/2})$, $O(\Delta^{3/2})$ and $O(\Delta^{5/2})$ order approximations respectively. The black dots denote the true prices. Y-axis of the right panel is on a logarithmic scale. The parameters are: $\sigma = 0.25, \alpha = 0.25, \beta = 2.5, r = 5\%, v = 0.2, y = 0.25, \xi = 4, \kappa = 0.2$, and $\Delta = 60/360$. The current level of the VIX is equal to $100 \cdot v$.

3.4.5 The Impact of Jumps

What is the price impact on options contributed by underlying jumps? Intuitively, the presence of jumps effectively influence the option prices by altering the volatility, skewness, and kurtosis of the underlying returns. The mean of jump size distribution seems irrelevant, as the jump compensator always zero it out so that the mean return of the risk-neutral dynamics is always equal to the risk free rate. We try to understand the role of jumps by investigating the option price expansion among alternative models.

Benchmark Jumps

The paradigm that incorporates discontinuous returns is the Merton's jump diffusion discussed in Merton (1976).

Example 6 (Merton's jump-diffusion model)**.** *Assume that*

$$\frac{dX_t}{X_t} = (r - (m-1)\lambda)dt + \sigma dW_t^Q + (e^J - 1)dN_t$$

where J has normal distribution $N\left(\log m - \frac{\nu^2}{2}, \nu^2\right)$ and N_t is a Poisson process with intensity λ.

The closed-form vanilla option pricing formula is given by Merton (1976):

$$\begin{aligned}
&\Psi(\Delta, x)\\
=&\sum_{k=0}^{\infty} \frac{(\lambda\Delta)^k}{k!}\left(x m^k e^{-\lambda m \Delta}\Phi\Big(\frac{\log(\frac{x}{K}) + (r - \lambda(m-1) + \sigma^2/2)\Delta + (\log m + \nu^2/2)k}{\sqrt{\sigma^2\Delta + k\nu^2}}\Big)\right.\\
&\left. - K e^{-(\lambda+r)\Delta}\Phi\Big(\frac{\log(\frac{x}{K}) + (r - \lambda(m-1) - \sigma^2/2)\Delta + (\log m - \nu^2/2)k}{\sqrt{\sigma^2\Delta + k\nu^2}}\Big)\right)
\end{aligned}$$

The following Closed-form expansions are plotted in Figure 3.6.

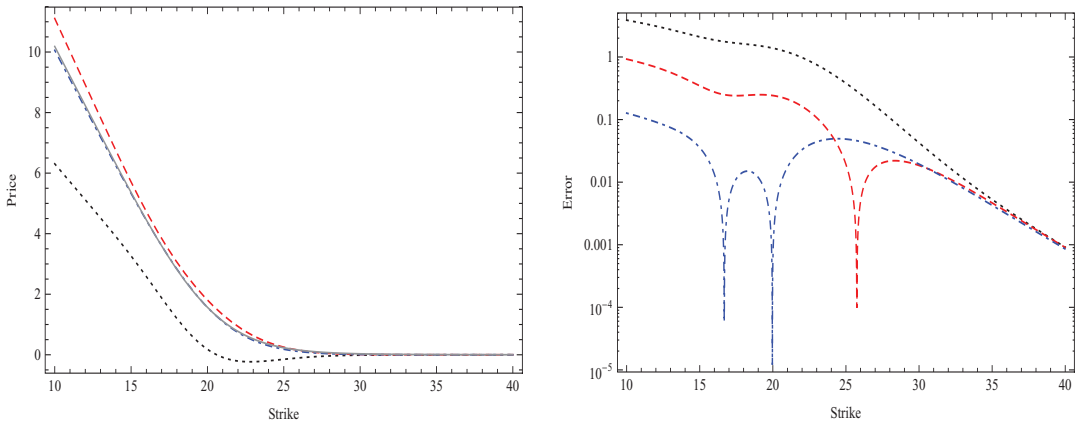
$$\begin{aligned}
\Psi(\Delta, x) =& \Phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} B^{(k)}(x)\Delta^k + \sqrt{\Delta}\phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty} C^{(k)}(x)\Delta^k + \sum_{k=1}^{\infty} D^{(k)}(x)\Delta^k\\
B^{(k)}(x) =& \frac{(-1)^k\left(x(m\lambda)^k - K(r+\lambda)^k\right)}{k!}, k \geq 0\\
C^{(-1)}(x) =& \frac{\log\left(\frac{x}{K}\right)}{\sigma}\\
C^{(0)}(x) =& \frac{\sigma(x-K)}{\log\left(\frac{x}{K}\right)}, \textit{if } x \neq K; \textit{or } K\sigma, \textit{if } x = K.\\
C^{(1)}(x) =& \frac{\sigma}{\log\left(\frac{x}{K}\right)^3}\left((K-x)\sigma^2 + K\left(\frac{x}{K}\right)^{\frac{1}{2} - \frac{r}{\sigma^2} + \frac{(-1+m)\lambda}{\sigma^2}}\sigma^2\log\left(\frac{x}{K}\right) + (-mx\lambda + K(r+\lambda))\log\left(\frac{x}{K}\right)^2\right),\\
&\textit{if } x \neq K; \textit{or}\\
&K\left(\frac{r^2}{2\sigma} + \frac{(-1+m)^2\lambda^2}{2\sigma} - \frac{(1+m)\lambda\sigma}{2} - \frac{\sigma^3}{24} - \frac{r\left(2(-1+m)\lambda + \sigma^2\right)}{2\sigma}\right), \textit{if } x = K.\\
D^{(1)}(x) =& mx\lambda\Phi\Big(\frac{\log(\frac{x}{K}) + \log(m) + \frac{1}{2}\nu^2}{\nu}\Big) - K\lambda\Phi\Big(\frac{\log(\frac{x}{K}) + \log(m) - \frac{\nu^2}{2}}{\nu}\Big)\\
D^{(2)}(x) =& \frac{1}{4}\lambda\Bigg(e^{-\frac{\nu^4 + 4\log(\frac{mx}{K})^2}{8\nu^2}}\sqrt{\frac{2mxK}{\pi}}\frac{\sigma^2}{\nu} - 2K\lambda\Phi\left(\frac{-\nu^2 + 2\log(m) + \log(x) - \log(K)}{\sqrt{2}\nu}\right)\\
&- 4m^2x\lambda\Phi\left(\frac{\nu^2 + 2\log(mx) - 2\log(K)}{2\nu}\right) + 2m^2x\lambda\Phi\left(\frac{\nu^2 + 2\log(m) + \log(x) - \log(K)}{\sqrt{2}\nu}\right)\\
&+ 4K(r+\lambda)\Phi\left(\frac{-\nu^2 + 2\log(mx) - 2\log(K)}{2\nu}\right)\Bigg)
\end{aligned}$$

These expansions are derived using Theorem 16 by considering an equivalent option written on the log price, so that the jump size is independent of the price level X. Apparently, up to the order of $O(\Delta^{1/2})$, the option prices are indistinguishable under Black-Scholes and Merton's Jump diffusion models. This is intuitive, as in expectation Poisson jumps occur with probability of the order $O(\Delta)$. The impact of jumps on option prices is

therefore on the order of $O(\Delta)$. More specifically, we can separate the first-order impact by jumps from the option price explicitly:

$$\underbrace{mx\lambda\left(\Phi\left(\frac{\log(\frac{x}{K})+\log(m)+\frac{1}{2}\nu^2}{\nu}\right)-\Phi\left(\frac{\log(\frac{x}{K})}{\sigma\sqrt{\Delta}}\right)\right)}_{\text{asset-or-nothing portion}}$$
$$\underbrace{-K\,\lambda\left(\Phi\left(\frac{\log(\frac{x}{K})+\log(m)-\frac{\nu^2}{2}}{\nu}\right)-\Phi\left(\frac{\log(\frac{x}{K})}{\sigma\sqrt{\Delta}}\right)\right)}_{\text{cash-or-nothing portion}}$$

Figure 3.6: Merton's Jump Diffusion Model



Note: The black dotted line, red dashed line and blue dotted-dash line illustrate the $O(\Delta^{1/2})$, $O(\Delta^{3/2})$ and $O(\Delta^{5/2})$ approximations respectively. The grey line denotes the true prices. Y-axis of the right panel is on a logarithmic scale. The parameters are: $\sigma = 0.2, r = 4\%, x = 20, \Delta = 0.5, m = 0.98, \nu = 0.15$, and $\lambda = 1.0$.

It implies immediately from the expansion that as jump intensity λ rises, the impact of jump becomes increasingly important. For at-the-money options, if the jump size has positive mean level that is $\log(m) > \nu^2/2$, then the probability of expiring in the money is boosted, so is the price of a cash-or-nothing binary portion, which agrees with the intuition. More generally, whether the jump impact on this portion is positive or not depends on the relative magnitude of $(\log(x/K) + \log(m) - \nu^2/2)/\nu$ and $\log(x/K)/(\sigma\sqrt{\Delta})$. When variance uncertainty parameter σ diminishes to zero, the price of jump impact is negative for in-the-money binary options, since if there were no jumps in reality, the option would

for sure expire in-the-money. When it comes to a vanilla option, it can be shown that this first order jump impact is always positive. Again, this is in agreement with our intuition that the presence of jump simply augments the volatility of the underlying returns, leading to an increase in the value of any options with convex payoff.

One-Sided and Skewed Jumps

To better capture the volatility smile and asymmetric leptokurtic feature of asset returns, Kou (2002) introduces an alternative jump diffusion model where the jumps J in log returns have asymmetric double exponential distribution with the density:

$$\nu(z) = p \cdot \eta_1 e^{-\eta_1 z} 1_{\{z \geq 0\}} + q \cdot \eta_2 e^{\eta_2 z} 1_{\{z < 0\}}$$

where $\eta_1 > 1$, $\eta_2 > 0$, $p + q = 1$, and $0 \leq p, q \leq 1$. The mean, variance, and skewness of the jump size in log returns are:

$$\gamma_1 = \frac{p}{\eta_1} - \frac{q}{\eta_2}, \gamma_2 = pq(\frac{1}{\eta_1} + \frac{1}{\eta_2})^2 + \frac{p}{\eta_1^2} + \frac{q}{\eta_2^2},$$
$$\gamma_3 = \frac{2(p^3 - 1)\eta_1^3 - 2(q^3 - 1)\eta_2^3 + 6pq\eta_1\eta_2(q\eta_2 - p\eta_1)}{(p\eta_2^2 + q\eta_1^2 + pq(\eta_1 + \eta_2)^2)^{\frac{3}{2}}}$$

By changing p, η_1, and η_2, this model is able to capture many empirical features of jumps, which may have disparate impacts on option prices. We analyze the first order impact through closed-form expansions:

Example 7 (Kou's jump-diffusion model)**.** *Assume that*

$$d\log(X_t) = \mu dt + \sigma dW_t^Q + JdN_t$$

where J has an asymmetric double exponential distribution with parameters p, η_1, and η_2, and N_t is a Poisson process with intensity λ. In this case, we have

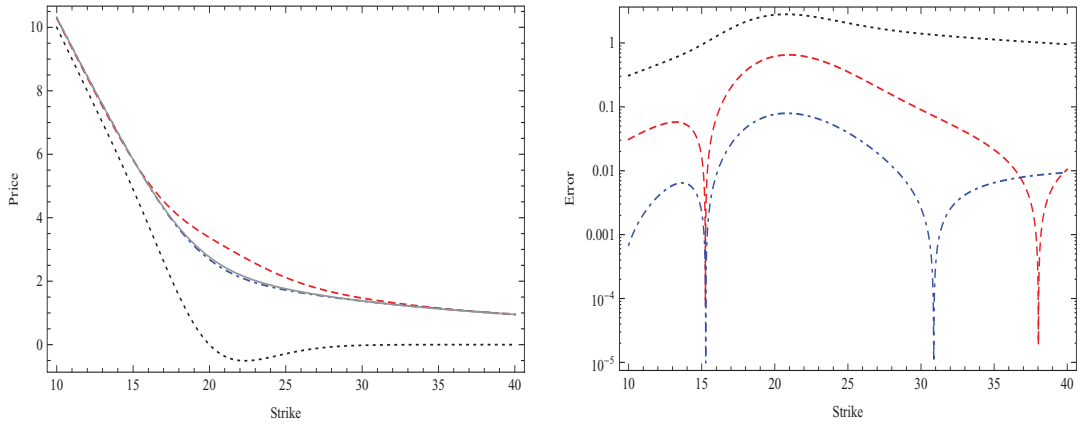
$$\mu = r - \frac{1}{2}\sigma^2 - \lambda\Big(\frac{p\eta_1}{\eta_1 - 1} + \frac{q\eta_2}{\eta_2 + 1} - 1\Big)$$

The closed-form vanilla option pricing formula is given by Theorem 2 in Kou (2002), and omitted here due to its complexity. We expand this model using expansion formulae given in the Appendix 3.6.5.

The following Closed-form expansions are plotted in Figure 3.7.

$$
\begin{aligned}
\Psi(\Delta,x) =& \Phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty}B^{(k)}(x)\Delta^k+\sqrt{\Delta}\phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty}C^{(k)}(x)\Delta^k\\
&+\left(1-\Phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\right)\sum_{k=1}^{\infty}D^{(k)}(x)\Delta^k\\
B^{(0)}(x) =& x-K\\
B^{(1)}(x) =& \frac{K(r(1+\eta_2)-(-1+p)(\frac{K}{x})^{\eta_2}\lambda)}{1+\eta_2}\\
B^{(2)}(x) =& -\frac{1}{2(-1+\eta_1)(1+\eta_2)^2(\eta_1+\eta_2)}\Bigg(K(-1+p)\Big(\frac{K}{x}\Big)^{\eta_2}\lambda^2\Big((-1+p)\eta_1^2(1+2\eta_2)+(-1+p)\eta_1(-1-\eta_2\\
&+2\eta_2^2)+\eta_2(1+p+2\eta_2+2p\eta_2+2p\eta_2^2)+(-1+p)(-1+\eta_1)\eta_2(1+\eta_2)(\eta_1+\eta_2)\log\Big(\frac{K}{x}\Big)\Big)-\frac{Kr^2}{2}\Bigg)\\
&+\frac{1}{2}K(-1+p)\Big(\frac{K}{x}\Big)^{\eta_2}\lambda(2r-\eta_2\sigma^2)\\
C^{(-1)}(x) =& \frac{\log\Big(\frac{x}{K}\Big)}{\sigma}\\
C^{(0)}(x) =& \frac{\sigma(x-K)}{\log(\frac{x}{K})}, \textit{if } x\neq K\textit{; or } K\sigma\textit{, if } x=K.\\
C^{(1)}(x) =& -\frac{\sigma}{(-1+\eta_1)(1+\eta_2)\log(\frac{K}{x})^3}\Bigg((K-x)(-1+\eta_1)(1+\eta_2)\sigma^2-\Big(\frac{K}{x}\Big)^{\frac{1}{2}+\frac{r}{\sigma^2}-\frac{(1+(-1+p)\eta_1+p\eta_2)\lambda}{(-1+\eta_1)(1+\eta_2)\sigma^2}}x(-1+\eta_1)\\
&(1+\eta_2)\sigma^2\log\Big(\frac{K}{x}\Big)-K\Big(-r(-1+\eta_1)(1+\eta_2)+\Big(-\Big(\frac{K}{x}\Big)^{\eta_2}(-1+\eta_1)+p\Big(\frac{x}{K}\Big)^{\eta_1}\Big(1-\Big(\frac{K}{x}\Big)^{\eta_1+\eta_2}\\
&+\Big(\frac{K}{x}\Big)^{\eta_1+\eta_2}\eta_1+\eta_2\Big)\Big)\lambda\Big)\log\Big(\frac{K}{x}\Big)^2\Bigg), \textit{if } x\neq K\textit{; or}\\
&-\frac{K\lambda\left(2r(1+(-1+p)\eta_1+p\eta_2)+(-1+(-2+p)\eta_2+\eta_1(1+p+2\eta_2))\sigma^2\right)}{2(-1+\eta_1)(1+\eta_2)\sigma}\\
&+\frac{K(1+(-1+p)\eta_1+p\eta_2)^2\lambda^2}{2(-1+\eta_1)^2(1+\eta_2)^2\sigma}-\frac{K\left(-12r^2+12r\sigma^2+\sigma^4\right)}{24\sigma}\\
D^{(1)}(x) =& \frac{Kp\Big(\frac{x}{K}\Big)^{\eta_1}\lambda}{-1+\eta_1}\\
D^{(2)}(x) =& \frac{Kp\Big(\frac{x}{K}\Big)^{\eta_1}\lambda^2}{2(-1+\eta_1)^2(1+\eta_2)(\eta_1+\eta_2)}\Bigg(2\eta_1-p\eta_1-4\eta_1^2+2p\eta_1^2+2\eta_1^3-2p\eta_1^3+p\eta_2-p\eta_1\eta_2-2p\eta_1^2\eta_2+p\eta_2^2\\
&-2p\eta_1\eta_2^2+p(-1+\eta_1)\eta_1(1+\eta_2)(\eta_1+\eta_2)\log\Big(\frac{K}{x}\Big)\Bigg)+\frac{1}{2}Kp\Big(\frac{x}{K}\Big)^{\eta_1}\lambda\Big(2r+\eta_1\sigma^2\Big)
\end{aligned}
$$

Figure 3.7: Kou's Jump Diffusion Model



Note: The black dotted line, red dashed line and blue dotted-dash line illustrate the $O(\Delta^{1/2})$, $O(\Delta^{3/2})$ and $O(\Delta^{5/2})$ approximations respectively. The grey line denotes the true prices. Y-axis of the right panel is on a logarithmic scale. The parameters are: $\sigma = 0.2, r = 4\%, x = 20, \Delta = 0.5, \eta_1 = 2.5, \eta_2 = 3.2, p = 0.4$, and $\lambda = 1.0$.

The first-order jump contribution can be separated out as well:

$$\lambda K\left(\frac{q}{1+\eta_2}\left(\frac{K}{x}\right)^{\eta_2}\Phi\Big(\frac{\log(\frac{x}{K})}{\sigma\sqrt{\Delta}}\Big)+\frac{p}{-1+\eta_1}\left(\frac{x}{K}\right)^{\eta_1}\Big(1-\Phi\Big(\frac{\log(\frac{x}{K})}{\sigma\sqrt{\Delta}}\Big)\Big)\right)$$

Similarly, no matter how much the portion of positive jumps accounts for, the jump impact on the vanilla option is always positive, as the primary contribution by jumps is to amplify the volatility. As λ decreases to zero, the model degenerates to the Black-Scholes case and the jump influence fades away.

Self-Exciting Jumps

To make the dynamics of jumps more interesting, we consider the following Hawkes jump-diffusion example where the jump intensity itself is another stochastic process driven by the same Poisson process. This jump process features contagion, or self-excitation, that is, the occurrence of a jump raises the odd of another jump in near future. This model was originally proposed by Hawkes (1971), and has recently been employed to model contagion phenomenon in finance by Aït-Sahalia et al. (2010a). We compare this model with Merton's jump diffusion by examining the first several terms in the expansion.

Example 8 (Hawkes' jump-diffusion model)**.** *Assume that the log price follows:*

$$d\log X_t = \mu dt + \sigma dW_t^Q + JdN_t$$
$$d\lambda_t = \alpha(\lambda_\infty - \lambda_t)dt + \beta dN_t$$

where J has normal distribution $N\big(\log m - \frac{\nu^2}{2}, \nu^2\big)$, and N_t is a self-exciting pure jump process with intensity λ_t following another stationary process with average intensity $\bar{\lambda} = E(\lambda_t) = \alpha\lambda_\infty/(\alpha - \beta)$. When $\beta = 0$ and $\lambda_0 = \lambda_\infty$, we have $\lambda_t = \bar{\lambda} = \lambda_\infty$, hence the model degenerates to the previous one. Similar to the Merton's model, we have

$$\mu = r - \frac{1}{2}\sigma^2 - (m-1)\bar{\lambda}$$

The partial differential equation corresponding to this model is

$$\begin{aligned} 0 =& -\frac{\partial\Psi(\Delta,x,\lambda)}{\partial\Delta} + (r-(m-1)\bar{\lambda})x\frac{\partial\Psi(\Delta,x,\lambda)}{\partial x} + \frac{1}{2}\sigma^2x^2\frac{\partial^2\Psi(\Delta,x,\lambda)}{\partial x^2} \\ &+ \alpha(\lambda_\infty - \lambda)\frac{\partial\Psi(\Delta,x,\lambda)}{\partial\lambda} + \lambda\int_{-\infty}^{\infty}\Big(\Psi(\Delta,xe^z,\beta+\lambda) - \Psi(\Delta,x,\lambda)\Big)\nu(z)dz \\ &- r\Psi(\Delta,x,\lambda) \end{aligned}$$

The following Closed-form expansions are plotted in Figure 3.8.

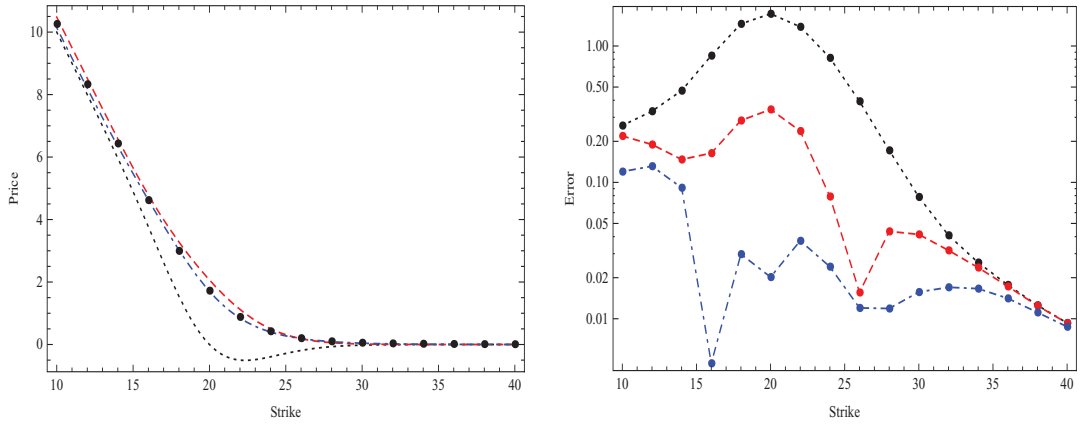
$$\begin{aligned} \Psi(\Delta,x,\lambda) =& \Phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty}B^{(k)}(x,\lambda)\Delta^k + \sqrt{\Delta}\phi\left(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}\right)\sum_{k=0}^{\infty}C^{(k)}(x,\lambda)\Delta^k + \sum_{k=1}^{\infty}D^{(k)}(x,\lambda)\Delta^k \\ B^{(0)}(x,\lambda) =& x - K \\ B^{(1)}(x,\lambda) =& -x\Big(\lambda + (m-1)\bar{\lambda}\Big) + K(r+\lambda) \\ B^{(2)}(x,\lambda) =& \frac{1}{2}\Big(x\Big((\lambda + (-1+m)\bar{\lambda})^2 + \alpha(\lambda-\lambda_\infty)\Big) - K\Big((r+\lambda)^2 + \alpha(\lambda-\lambda_\infty)\Big)\Big) \\ C^{(-1)}(x,\lambda) =& \frac{\log\left(\frac{x}{K}\right)}{\sigma} \\ C^{(0)}(x,\lambda) =& \frac{\sigma(x-K)}{\log\left(\frac{x}{K}\right)}, \text{if } x \neq K; \text{or } K\sigma, \text{if } x = K. \\ C^{(1)}(x,\lambda) =& \frac{1}{\log(\frac{x}{K})^3}\sigma\Bigg((K-x)\sigma^2 + K\Big(\frac{x}{K}\Big)^{\frac{1}{2}-\frac{r}{\sigma^2}+\frac{(-1+m)\bar{\lambda}}{\sigma^2}}\sigma^2\log\Big(\frac{x}{K}\Big) + \Big(K(r+\lambda) - x((\lambda+(-1+m)\bar{\lambda})\Big) \\ & \log\Big(\frac{x}{K}\Big)^2\Bigg), \text{if } x \neq K; \text{or} \end{aligned}$$

$$K\left(\frac{r^2}{2\sigma}+\frac{(-1+m)^2\bar{\lambda}^2}{2\sigma}-\frac{((-1+m)\bar{\lambda}+2\lambda)\sigma}{2}-\frac{\sigma^3}{24}-\frac{r\left(2(-1+m)\bar{\lambda}+\sigma^2\right)}{2\sigma}\right), \textit{if } x=K.$$

$$D^{(1)}(x,\lambda) = mx\lambda\Phi\Big(\frac{\log(\frac{x}{K})+\log(m)+\frac{1}{2}\nu^2}{\nu}\Big) - K\lambda\Phi\Big(\frac{\log(\frac{x}{K})+\log(m)-\frac{\nu^2}{2}}{\nu}\Big)$$

$$\begin{aligned}D^{(2)}(x,\lambda) =& \frac{1}{4}\Bigg(e^{-\frac{\nu^4+4\log(\frac{mx}{K})^2}{8\nu^2}}\sqrt{\frac{2Kmx}{\pi}}\frac{\lambda\sigma^2}{\nu}+2m^2x\lambda(\beta+\lambda)\Phi\Big(\frac{\nu^2-\log(K)+2\log(m)+\log(x)}{\sqrt{2}\nu}\Big)\\ &+2K\Big(2r\lambda+\lambda(\beta+2\lambda)+\alpha(\lambda-\lambda_\infty)\Big)\Phi\Big(\frac{-\nu^2-2\log(K)+2\log(mx)}{2\nu}\Big)\\ &-2K\lambda(\beta+\lambda)\Phi\Big(\frac{-\nu^2+2\log(m)-\log(K)+\log(x)}{\sqrt{2}\nu}\Big)\\ &-2mx(\lambda(\beta+2(\lambda+(-1+m)\bar{\lambda}))+\alpha(\lambda-\lambda_\infty))\Phi\Big(\frac{\nu^2-2\log(K)+2\log(mx)}{2\nu}\Big)\Bigg)\end{aligned}$$

Figure 3.8: Hawkes' Jump Diffusion Model



Note: The black dotted line, red dashed line and blue dotted-dash line illustrate the $O(\Delta^{1/2})$, $O(\Delta^{3/2})$ and $O(\Delta^{5/2})$ order approximations respectively. The black dots denote the true prices. Y-axis of the right panel is on a logarithmic scale. The parameters are: $\sigma = 0.2, \alpha = 15, \beta = 10, m = 0.98, \nu = 0.15, r = 4\%, \lambda = 1.0, \lambda_\infty = 0.8, x = 20$, and $\Delta = 0.5$. The number of sample path is 5000 in Monte Carlo, each of which is simulated using Euler's method with 2520 time intervals.

It appears from $B^{(1)}(x,\lambda)$ that the contagion feature ($\beta \neq 0$) appears in the very first order ($O(\Delta)$) impact of jumps on the option value, simply by altering the risk neutral expectation ($\bar{\lambda}$) of jump arrival rate. It seems counterintuitive at first thought, as one may expect the contagion to exhibit on a higher order after the first jump has occurred. However, since the contagion effect is expected, it is encrypted into option prices immediately without waiting for the occurrence of the first jump. When $\beta = 0$, and $\lambda = \lambda_\infty = \bar{\lambda}$, the expansions degenerate to those in Merton's case.

3.5 Concluding Remarks

We introduce a novel approach to expanding European option prices using closed-form convergent sequence in the time variable. The proposed approach works with general dynamics, without any requirement on affine structure or closed-form characteristics functions, bridging the gap between sparse closed-form solutions and less informative numerical methods. In addition to the computational benefits from closed-form expansion formulae, this approach provides insight into how model parameters affect option prices and their relative importance as the option contract approaches expiration, which even closed-form solutions may not offer.

3.6 Appendix

3.6.1 Assumptions

I adopt similar assumptions made by Aït-Sahalia (2002).

Assumption 19. *(Smoothness of the Coefficients)*
The functions $\mu(x)$ and $\sigma(x)$ are infinitely differentiable in $x \in D_X$.

Assumption 20. *(Nondegeneracy of the Diffusion)*
1. $D_X = (-\infty, +\infty)$: there exists a constant $c > 0$, such that $\sigma(x) > c$ for all $x \in D_X$.
2. $D_X = (0, +\infty)$: near 0, if $\lim_{x \to 0+} \sigma(x) = 0$, there exist constant ξ_0, $\omega \geq 0$ and $\rho \geq 0$ such that $\sigma(x) \geq \omega x^\rho$ for all $0 < x < \xi_0$; away from 0, for each $\xi > 0$, there exists a constant $c_\xi > 0$ such that $\sigma(x) \geq c_\xi$ for all $x \in [\xi, \infty]$.

The constraints on the boundary behavior are directly given to the transform diffusion Y_t for convenience. The domain of Y is given by $D_Y = (\underline{y}, \bar{y})$.

Assumption 21. *(Boundary Behavior)*
Let $\lambda_Y(y) = -\frac{1}{2}(\mu_Y^2(y) + \frac{\partial \mu_Y(y)}{\partial y})$. μ_Y and $\frac{\partial \mu_Y(y)}{\partial y}$ have at most polynomial growth near the boundaries and $\lim_{y \to \partial D_Y} \lambda(y) < +\infty$. Near $\bar{y} = +\infty$, $\mu_Y(y) \leq -Ky^\beta$ for some $\beta > 1$; near $\underline{y} = -\infty$, $\mu_Y(y) \geq K|y|^\beta$ for some $\beta > 1$; near $\underline{y} = 0$, $\mu_Y(y) \geq \kappa y^{-\alpha}$ for some $\alpha > 1$ and $\kappa > 0$; and near $\bar{y} = 0$, $\mu_Y(y) \leq -\kappa|y|^{-\alpha}$ for some $\alpha > 1$ and $\kappa > 0$.

3.6.2 Proof of Theorem 14

The proof is separated into two steps:

Step 1: Convergence of $\Psi^{(J)}(\Delta, x)$.

Choose $\bar{\Delta}$ as in Proposition 2 of Aït-Sahalia (2002). It suffices to show that

$$\sum_{j=0}^{J}(-1)^{j+1}\eta_{j+1}(\Delta, y)H_j(z)$$

absolutely converges, uniformly for y in any compact set of D_y and for $z \in \mathcal{R}$. Let $z = \Delta^{-\frac{1}{2}}(y - y)$. Define

$$\nu_j(\Delta, y) = \frac{1}{j!}\int_{-\infty}^{\infty} H_j(z)\frac{\partial p_Z(\Delta, z|y)}{\partial z}dz$$

As argued in Aït-Sahalia (2002) (page 252-253) that $\eta_j(\Delta, y)$ is well defined and satisfies that $\eta_j(\Delta, y) = \nu_{j+1}(\Delta, y)$, and by Theorem II in Stone (1927), there exists a constant K, such that for all $z \in \mathcal{R}$ and every integer j,

$$|H_j(z)| \leq K(j!)^{\frac{1}{2}}j^{-\frac{1}{4}}(1 + 2^{-\frac{5}{4}}|z|^{\frac{5}{2}}e^{\frac{z^2}{4}})$$

therefore, we have

$$\begin{aligned}
&|\eta_{j+1}(\Delta, y)H_j(z)| = |\nu_{j+2}(\Delta, y)||H_j(z)| \\
\leq &K(1 + 2^{-\frac{5}{4}}|z|^{\frac{5}{2}}e^{\frac{z^2}{4}})\Big((j!)^{\frac{1}{2}}j^{-\frac{1}{4}}((j+2)!)^{-\frac{1}{2}}((j+2)!)^{\frac{1}{2}}|\nu_{j+2}(\Delta, y)|\Big) \\
\leq &\frac{1}{2}K(1 + 2^{-\frac{5}{4}}|z|^{\frac{5}{2}}e^{\frac{z^2}{4}})\Big((j!)j^{-\frac{1}{2}}((j+2)!)^{-1} + (j+2)!\nu_{j+2}^2(\Delta, y)\Big)
\end{aligned}$$

Notice that $\sum(j!)j^{-\frac{1}{2}}((j+2)!)^{-1}$ converges. Also, Aït-Sahalia (2002) (page 253) shows that

$$\sum_{j=0}^{J} j!\nu_j^2(\Delta, y) \leq (2\pi)^{\frac{1}{2}}\int_{-\infty}^{+\infty} e^{\frac{z^2}{2}}(\frac{\partial p_Z(\Delta, z|y)}{\partial z})^2 dz < \infty$$

uniformly for y over any compact set of D_Y. Hence, $\Psi^{(J)}(\Delta, x)$ converges, and $\Psi(\Delta, x)$ is

well defined.

Step 2: $\lim_{\Delta\to 0}\Psi(\Delta,x)=1$, if $x>K$, or 0 if $x<K$.

In fact, we have already shown that

$$\begin{aligned}
&\phi(z)\sum_{j=0}^{\infty}|\eta_{j+1}(\Delta,y)H_j(z)|\\
\leq&\frac{1}{2}K\phi(z)(1+2^{-\frac{5}{4}}|z|^{\frac{5}{2}}e^{\frac{z^2}{4}})\Big(\sum_{j=0}^{\infty}(j!)j^{-\frac{1}{2}}((j+2)!)^{-1}+\sum_{j=0}^{\infty}(j)!\nu_j^2(\Delta,y)\Big)\\
\leq&\frac{1}{2}K\phi(z)(1+2^{-\frac{5}{4}}|z|^{\frac{5}{2}}e^{\frac{z^2}{4}})\Big(\sum_{j=0}^{\infty}(j!)j^{-\frac{1}{2}}((j+2)!)^{-1}+(2\pi)^{\frac{1}{2}}\int_{-\infty}^{+\infty}e^{\frac{z^2}{2}}(\frac{\partial p_Z(\Delta,z|y)}{\partial z})^2dz\Big)
\end{aligned}$$

Further, it is shown by Aït-Sahalia (2002) (page 253) that for some constant $b_i, i=0,\ldots,4$,

$$|\frac{\partial p_Z(\Delta,z|y)}{\partial z}|\leq b_0e^{-\frac{3z^2}{8}}R(|z|,|y|)e^{b_1|z||y|+b_2|z|+b_3|y|+b_4y^2} \tag{3.30}$$

where $R(|z|,|y|)$ is a polynomial of finite order in $|z|$ and y.

Plugging in $y=\gamma(x)$ and $z=\frac{\gamma(x)-\gamma(K)}{\sqrt{\Delta}}$, and let $\Delta\to 0$, it follows that

$$\phi(\frac{\gamma(x)-\gamma(K)}{\sqrt{\Delta}})\sum_{j=0}^{\infty}|\eta_{j+1}(\Delta,\gamma(x))H_j(\frac{\gamma(x)-\gamma(K)}{\sqrt{\Delta}})|\longrightarrow 0$$

Also, because $\gamma(x)$ is increasing in x, $\lim_{\Delta\to 0}\Phi(\frac{\gamma(x)-\gamma(K)}{\sqrt{\Delta}})=1_{\{x>K\}}$, which concludes the second step.

Step 3: $\Psi(\Delta,x)$ solves the Feynman-Kac PDE (3.2).

For convenience, we change variable $y=\gamma(x)$. Let $\tilde{\Psi}(\Delta,y)=\Psi(\Delta,x)$, and $\tilde{\Psi}^{(J)}(\Delta,y)=\Psi^{(J)}(\Delta,x)$. The Feynman-Kac PDE becomes

$$(-\frac{\partial}{\partial\Delta}+\mathcal{L}_Y)\tilde{\Psi}(\Delta,y)=0 \tag{3.31}$$

where

$$\mathcal{L}_Yf=\mu_Y(y)\frac{\partial f}{\partial y}+\frac{1}{2}\frac{\partial^2 f}{\partial y^2}$$

Taking derivative of $\eta_j(\Delta, y)$ with respect to Δ and y, and using the fact that $H_j'(z) = -jH_{j-1}(z)$, we have

$$\begin{aligned}
\frac{\partial \eta_j(\Delta,y)}{\partial \Delta} =& \frac{1}{j!}\int_{-\infty}^{+\infty} H_j'(\frac{\omega-y}{\sqrt{\Delta}})(-\frac{\omega-y}{2\Delta^{\frac{3}{2}}})p_Y(\Delta,\omega|y) + H_j(\frac{\omega-y}{\sqrt{\Delta}})\frac{\partial p_Y(\Delta,\omega|y)}{\partial \Delta}d\omega \\
\frac{\partial \eta_j(\Delta,y)}{\partial y} =& \frac{1}{j!}\int_{-\infty}^{+\infty} (-\frac{1}{\sqrt{\Delta}})H_j'(\frac{\omega-y}{\sqrt{\Delta}})p_Y(\Delta,\omega|y) + H_j(\frac{\omega-y}{\sqrt{\Delta}})\frac{\partial p_Y(\Delta,\omega|y)}{\partial y}d\omega \\
\frac{\partial^2 \eta_j(\Delta,y)}{\partial y^2} =& \frac{1}{j!}\int_{-\infty}^{+\infty} H_j'(\frac{\omega-y}{\sqrt{\Delta}})(-\frac{1}{\sqrt{\Delta}})\frac{\partial p_Y(\Delta,\omega|y)}{\partial y} + H_j(\frac{\omega-y}{\sqrt{\Delta}})\frac{\partial^2 p_Y(\Delta,\omega|y)}{\partial y^2}d\omega \\
&+ \frac{1}{(j-1)!}\int_{-\infty}^{+\infty} H_{j-1}'(\frac{\omega-y}{\sqrt{\Delta}})(-\frac{1}{\Delta})p_Y(\Delta,\omega|y) \\
&+ \frac{1}{\sqrt{\Delta}}H_{j-1}(\frac{\omega-y}{\sqrt{\Delta}})\frac{\partial p_Y(\Delta,\omega|y)}{\partial y}d\omega
\end{aligned}$$

By tedious calculation using the fact that $H_{j+1}(z) = H_j'(z) - zH_j(z)$, we have

$$(-\frac{\partial}{\partial \Delta} + \mathcal{L}_Y)\tilde{\Psi}^{(J)}(\Delta, y) \tag{3.32}$$

$$= \frac{1}{\sqrt{\Delta}}\phi(\frac{y-\gamma(K)}{\sqrt{\Delta}})(-1)^{J+1}\Big(\mu_Y(y)\eta_{J+1}(\Delta,y) + \frac{\partial \eta_{J+1}(\Delta,y)}{\partial y}\Big)H_{J+1}(\frac{y-\gamma(K)}{\sqrt{\Delta}}) \tag{3.33}$$

We have shown that as $J \to 0$,

$$|\eta_{J+1}(\Delta,y)H_{J+1}(\frac{y-\gamma(K)}{\sqrt{\Delta}})| \longrightarrow 0$$

uniformly in y over any compact set of D_Y.

Similarly as above, notice that

$$\frac{\partial \eta_{J+1}(\Delta,y)}{\partial y} = \frac{1}{(J+2)!}\int_{-\infty}^{+\infty} H_{J+2}(\omega)\frac{\partial^2 p_Z(\Delta,\omega|y)}{\partial \omega \partial y}d\omega$$

it is then sufficient to show that

$$\int_{-\infty}^{+\infty} e^{\frac{\omega^2}{2}}(\frac{\partial^2 p_Z(\Delta,\omega|y)}{\partial \omega \partial y})^2 d\omega < \infty$$

which can be implied from a similar bound as in (3.30).

Because of (3.32), and uniform convergence of $\tilde{\Psi}^{(J)}(\Delta, z)$ to $\tilde{\Psi}(\Delta, z)$, it follows that $\tilde{\Psi}(\Delta, z)$ satisfies the PDE (3.31), which concludes the proof.

3.6.3 Proof of Theorem 15

We plug (3.8) into equation (3.2), and obtain

$$\begin{aligned}
&(-\frac{\partial}{\partial\Delta}+\mathcal{L}-r)\Big(e^{-r\Delta}\Phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\Big)\\
=&\frac{1}{\sqrt{2\pi}}\exp(-r\Delta-\frac{(C^{(-1)}(x))^2}{2\Delta})\Big\{\frac{1}{2}C^{(-1)}(x)\Big(1-\sigma^2(x)(\frac{dC^{(-1)}(x)}{dx})^2\Big)\Delta^{-\frac{3}{2}}\\
&+\mathcal{L}C^{(-1)}(x)\Delta^{-\frac{1}{2}}\Big\}
\end{aligned}$$

and

$$\begin{aligned}
&(-\frac{\partial}{\partial\Delta}+\mathcal{L}-r)\Big(e^{-r\Delta}\frac{\sqrt{\Delta}}{\sqrt{2\pi}}\exp(-\frac{C^{(-2)}(x)}{\Delta})\sum_{k=0}^{\infty}C^{(k)}(x)\Delta^k\Big)\\
=&\frac{1}{\sqrt{2\pi}}\exp(-r\Delta-\frac{C^{(-2)}(x)}{\Delta})\Big(\psi_1(\Delta,x)+\psi_2(\Delta,x)+\psi_3(\Delta,x)\Big)
\end{aligned}$$

where

$$\begin{aligned}
\psi_1(\Delta,x)=&\Delta^{-\frac{3}{2}}\Big(-C^{(-2)}(x)+\frac{1}{2}\sigma^2(x)(\frac{dC^{(-2)}(x)}{dx})^2\Big)\sum_{k=0}^{\infty}C^{(k)}(x)\Delta^k\\
\psi_2(\Delta,x)=&\Delta^{\frac{1}{2}}\sum_{k=0}^{\infty}\Big(-\frac{1}{2}C^{(k+1)}(x)-(k+1)C^{(k+1)}(x)-\mathcal{L}C^{(-2)}(x)C^{(k+1)}(x)\\
&-\sigma^2(x)\frac{dC^{(-2)}(x)}{dx}\frac{dC^{(k+1)}(x)}{dx}+\mathcal{L}C^{(k)}(x)\Big)\Delta^k\\
\psi_3(\Delta,x)=&\Delta^{-\frac{1}{2}}\Big(-C^{(0)}(x)(\frac{1}{2}+\mathcal{L}C^{(-2)}(x))-\frac{dC^{(-2)}(x)}{dx}\frac{dC^{(0)}(x)}{dx}\sigma^2(x)\Big)
\end{aligned}$$

Requiring that the coefficients of $\Delta^{-\frac{3}{2}}$ are zeros, we have

$$-C^{(-2)}(x)+\frac{1}{2}\sigma^2(x)(\frac{dC^{(-2)}(x)}{dx})^2=0 \tag{3.34}$$

then, imposing $C^{(-2)}(K)=0$ gives

$$C^{(-2)}(x)=\frac{1}{2}(\int_K^x\frac{1}{\sigma(s)}ds)^2 \tag{3.35}$$

and also,

$$1 - \sigma^2(x)(\frac{dC^{(-1)}(x)}{dx})^2 = 0 \tag{3.36}$$

so with $C^{(-1)}(K) = 0$ and imposing increasing monotonicity, we arrive at

$$C^{(-1)}(x) = \int_K^x \frac{1}{\sigma(s)} ds \tag{3.37}$$

3.6.4 Proof of Theorem 16

Taking the postulated terms of (3.16) into the equation (3.14) respectively, we obtain

$$\begin{aligned}
&(-\frac{\partial}{\partial\Delta} + \mathcal{A} - r(x))\Big(\Phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\sum_{k=0}^{\infty} B^{(k)}(x)\Delta^k\Big) \\
=&\Phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\sum_{k=0}^{\infty}\Big(-(k+1)B^{(k+1)}(x) - (r(x)+\lambda(x))B^{(k)}(x) + \mathcal{L}B^{(k)}(x)\Big)\Delta^k \\
+&\Delta^{-\frac{1}{2}}\phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\sum_{k=0}^{\infty}\Big(B^{(k)}(x)\mathcal{L}C^{(-1)}(x) + (\frac{dB^{(k)}(x)}{dx})\sigma^2(x)(\frac{dC^{(-1)}(x)}{dx})\Big)\Delta^k \\
+&\lambda(x)\sum_{k=0}^{\infty}\int_{-\infty}^{\infty}\Phi(\frac{C^{(-1)}(x+z)}{\sqrt{\Delta}})B^{(k)}(x+z)\nu(z)dz\Delta^k
\end{aligned} \tag{3.38}$$

and

$$\begin{aligned}
&(-\frac{\partial}{\partial\Delta} + \mathcal{A} - r(x))\Big(\Delta^{\frac{1}{2}}\phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\sum_{k=0}^{\infty} C^{(k)}(x)\Delta^k\Big) \\
=&\phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\Big(\psi_1(\Delta,x) + \psi_2(\Delta,x) + \psi_3(\Delta,x)\Big) + \psi_4(\Delta,x)
\end{aligned}$$

where

$$\begin{aligned}
\psi_1(\Delta,x) =&\Delta^{-\frac{3}{2}}\sum_{k=0}^{\infty}\frac{1}{2}(C^{(-1)}(x))^2C^{(k)}(x)\Big((\frac{dC^{(-1)}(x)}{dx})^2\sigma^2(x) - 1\Big)\Delta^k \\
\psi_2(\Delta,x) =&\Delta^{\frac{1}{2}}\sum_{k=0}^{\infty}\Big(-C^{(k+1)}(x)\Big(1 + C^{(-1)}(x)\mathcal{L}C^{(-1)}(x)\Big) - (k+1)C^{(k+1)}(x) \\
&- C^{(-1)}(x)(\frac{dC^{(k+1)}(x)}{dx})\sigma^2(x)(\frac{dC^{(-1)}(x)}{dx}) + (\mathcal{L} - r(x) - \lambda(x))C^{(k)}(x)\Big)\Delta^k \\
\psi_3(\Delta,x) =&\Delta^{-\frac{1}{2}}\Big(-C^{(0)}(x)\Big(1 + C^{(-1)}(x)\mathcal{L}C^{(-1)}(x)\Big)
\end{aligned}$$

$$
\begin{aligned}
&- C^{(-1)}(x)(\frac{dC^{(0)}(x)}{dx})\sigma^2(x)(\frac{dC^{(-1)}(x)}{dx})\Big)\\
\psi_4(\Delta,x) =&\lambda(x)\frac{\Delta^{\frac{1}{2}}}{\sqrt{2\pi}}\sum_{k=0}^{\infty}\int_{-\infty}^{\infty}\exp(-\frac{(C^{(-1)}(x+z))^2}{2\Delta})C^{(k)}(x+z)\nu(z)dz\Delta^k
\end{aligned}
$$

Let $D^{(0)}(x)=0$. The contribution of the last term is

$$
(-\frac{\partial}{\partial\Delta}+\mathcal{A}-r)\Big(\sum_{k=1}^{\infty}D^{(k)}(x)\Delta^k\Big)=\sum_{k=0}^{\infty}\Big(-(k+1)D^{(k+1)}(x)+(\mathcal{A}-r(x))D^{(k)}(x)\Big)\Delta^k
$$

Note that we can reuse the formulae

We consider each integral term in $\psi_4(\Delta,x)$. Imposing that $\frac{dC^{(-1)}(x)}{dx}>0$, and by the inverse mapping theorem, we assume that $C^{(-1)}(x)$ is invertible with the other variables fixed. By a similar argument as in Yu (2007), we change variable such that $\omega=C^{(-1)}(x+z)$, which is a function of z. Let M_r the rth moment of standard normal variable.

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\exp(-\frac{(C^{(-1)}(x+z))^2}{2\Delta})C^{(k)}(x+z)\nu(z)dz\\
=&\int_{-\infty}^{\infty}\exp(-\frac{\omega^2}{2\Delta})C^{(k)}\Big(((C^{(-1)})^{inv}(\omega))\Big)\nu((C^{(-1)})^{inv}(\omega)-x)/|\frac{dC^{(-1)}}{dx}|_{(C^{(-1)})^{inv}(\omega)}d\omega\\
=&\int_{-\infty}^{\infty}\exp(-\frac{\omega^2}{2\Delta})g_k(x,\omega)d\omega\\
=&\int_{-\infty}^{\infty}\exp(-\frac{\omega^2}{2\Delta})\sum_{r=0}^{\infty}\frac{1}{r!}\frac{\partial^r g_k}{\partial\omega^r}|_{\omega=0}\omega^r d\omega\\
=&\sum_{r=0}^{\infty}\frac{1}{r!}\frac{\partial^r g_k}{\partial\omega^r}|_{\omega=0}\int_{-\infty}^{\infty}\exp(-\frac{\omega^2}{2\Delta})\omega^r d\omega\\
=&\sqrt{2\pi}\sum_{r=0}^{\infty}\frac{1}{r!}M_r\frac{\partial^r g_k}{\partial\omega^r}|_{\omega=0}\Delta^{\frac{r+1}{2}}\\
=&\sqrt{2\pi}\sum_{r=0}^{\infty}\frac{1}{(2r)!}M_{2r}\frac{\partial^{2r} g_k}{\partial\omega^{2r}}|_{\omega=0}\Delta^{r+\frac{1}{2}}
\end{aligned}
$$

Therefore,

$$
\psi_4(\Delta,x)=\Delta\lambda(x)\sum_{k=0}^{\infty}\Delta^k\sum_{r=0}^{k}\frac{1}{(2r)!}M_{2r}\frac{\partial^{2r}g_{k-r}}{\partial\omega^{2r}}|_{\omega=0}
$$

and

$$g_k(x,\omega) = C^{(k)}((C^{(-1)})^{inv}(\omega))\nu\Big(((C^{(-1)})^{inv}(\omega) - x)\Big)\Big(|\frac{dC^{(-1)}}{dx}|_{(C^{(-1)})^{inv}(\omega)}\Big)^{-1}$$

Similarly, we can deduce that

$$\begin{aligned}
&\sum_{k=0}^{\infty}\int_{-\infty}^{\infty}\Phi(\frac{C^{(-1)}(x+z)}{\sqrt{\Delta}})B^{(k)}(x+z)\nu(z)dz\Delta^k \\
=&\sum_{k=0}^{\infty}\frac{\Delta^{-\frac{1}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-\frac{\omega^2}{2\Delta})d\omega\int_{(C^{(-1)})^{inv}(\omega)-x}^{\infty}B^{(k)}(x+z)\nu(z)dz\Delta^k \\
=&\sum_{k=0}^{\infty}\sum_{r=0}^{k}\frac{1}{(2r)!}\frac{\partial^{2r}h_{k-r}(x,\omega)}{\partial\omega^{2r}}|_{\omega=0}M_{2r}\Delta^k
\end{aligned}$$

where

$$h_k(x,\omega) = \int_{(C^{(-1)})^{inv}(\omega)-x}^{\infty}B^{(k)}(x+z)\nu(z)dz$$

Matching the coefficients yields the desired equations.

3.6.5 Expansion Formulae for Jump Diffusions with Positive Jump Sizes

When the jump size density is supported on $[0,\infty)$, or is non-smooth at origin, e.g. double exponential distributions, the expansion formulae are more involved. In fact, we may rewrite the postulated formula in the following way:

$$\begin{aligned}
\Psi(\Delta,x) = &\Phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\sum_{k=0}^{\infty}B^{(k)}(x)\Delta^k + \sqrt{\Delta}\phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\sum_{k=0}^{\infty}C^{(k)}(x)\Delta^k \\
&+(1-\Phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}))\sum_{k=1}^{\infty}D^{(k)}(x)\Delta^k
\end{aligned}$$

Without loss of generality, we consider the $[0,\infty)$ case and focus on terms that are different from the previous case:

$\psi_4(\Delta,x)$

$$
\begin{aligned}
=&\lambda(x)\frac{\Delta^{\frac{1}{2}}}{\sqrt{2\pi}}\sum_{k=0}^{\infty}\int_0^{\infty}\exp(-\frac{(C^{(-1)}(x+z))^2}{2\Delta})C^{(k)}(x+z)\nu(z)dz\Delta^k\\
=&\lambda(x)\phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\sum_{k=0}^{\infty}G_0(k,x)\Delta^{k+\frac{3}{2}}+\lambda(x)(1-\Phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}}))\sum_{k=0}^{\infty}G_1(k,x)\Delta^{k+1}
\end{aligned}
$$

where

$$
\begin{aligned}
G_0(k,x)=&\sum_{j=0}^{k}\sum_{l=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^l(C^{(-1)}(x))^{l+s}}{(s+2j+1)!!(s)!!(l)!}\frac{\partial^{s+2j+l+1}g_{k-j}}{\partial\omega^{s+2j+l+1}}|_{\omega=C^{(-1)}(x)}\\
G_1(k,x)=&\sum_{j=0}^{k}\sum_{r=0}^{\infty}\frac{(-1)^rC^{(-1)}(x)^r}{(r)!(2j)!!}\frac{\partial^{r+2j}g_{k-j}}{\partial\omega^{r+2j}}|_{\omega=C^{(-1)}(x)}
\end{aligned}
$$

Also,

$$
\begin{aligned}
&\sum_{k=0}^{\infty}\int_0^{\infty}\Phi(\frac{C^{(-1)}(x+z)}{\sqrt{\Delta}})B^{(k)}(x+z)\nu(z)dz\Delta^k\\
=&\sum_{k=0}^{\infty}\int_0^{\infty}B^{(k)}(x+z)\nu(z)dz\Delta^k-\sum_{k=0}^{\infty}\int_{C^{(-1)}(x)}^{\infty}(1-\Phi(\frac{\omega}{\sqrt{\Delta}}))(-\frac{\partial h_k(x,\omega)}{\partial\omega})d\omega\Delta^k\\
=&\phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\sum_{k=0}^{\infty}H_0(k,x)\Delta^{k+\frac{1}{2}}+\Phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})\sum_{k=0}^{\infty}(H_2(k,x)-H_1(k-1,x))\Delta^k\\
&+\sum_{k=0}^{\infty}\Big(h_k(x,C^{(-1)}(x))-H_2(k,x)+H_1(k-1,x)\Big)\Delta^k
\end{aligned}
$$

where

$$
\begin{aligned}
H_0(k,x)=&\sum_{j=0}^{k}\sum_{l=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^l(C^{(-1)}(x))^{l+s}}{(s+2j+1)!!(s)!!(l)!}\frac{\partial^{s+2j+l+1}h_{k-j}}{\partial\omega^{s+2j+l+1}}|_{\omega=C^{(-1)}(x)}\\
H_1(k,x)=&\sum_{j=0}^{k}\sum_{r=0}^{\infty}\frac{(-1)^r(C^{(-1)}(x))^r}{(r)!(2j+2)!!}\frac{\partial^{r+2j+2}h_{k-j}}{\partial\omega^{r+2j+2}}|_{\omega=C^{(-1)}(x)}\\
H_2(k,x)=&\sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^s(C^{(-1)}(x))^{r+1+s}}{(r+1)!s!}\frac{\partial^{r+1+s}h_k}{\partial\omega^{r+1+s}}|_{\omega=C^{(-1)}(x)}
\end{aligned}
$$

The above calculations rely on the following formula: for any integer $k\geq 0$,

$$
\int_A^{\infty}\omega^re^{-\frac{\omega^2}{2\Delta}}d\omega
$$

$$= \begin{cases} 2^k \Delta^{k+1} e^{-\frac{A^2}{2\Delta}} \sum_{l=0}^{k} \frac{k!}{l!} \Big(\frac{A^{2l}}{2^l \Delta^l} \Big) \\ \text{if } r = 2k+1, \\ 2^{k+\frac{1}{2}} \Delta^{k+\frac{3}{2}} \Big\{ \frac{(2k+1)!!}{2^k} \sqrt{\pi} \Phi\Big(- \frac{A}{\sqrt{\Delta}} \Big) + e^{-\frac{A^2}{2\Delta}} \sum_{l=1}^{k+1} \frac{(2k+1)!!}{2^{k+\frac{1}{2}} (2l-1)!!} \Big(\frac{A^{2l-1}}{\Delta^{l-\frac{1}{2}}} \Big) \Big\} \\ \text{if } r = 2k+2. \end{cases}$$

and

$$\int_A^\infty \omega^r (1 - \Phi(\frac{\omega}{\sqrt{\Delta}})) d\omega = -(1 - \Phi(\frac{A}{\sqrt{\Delta}})) \frac{A^{r+1}}{r+1} + \frac{1}{\sqrt{\Delta}} \int_A^\infty \frac{\omega^{r+1}}{r+1} \phi(\frac{\omega}{\sqrt{\Delta}}) d\omega$$

Finally, we denote $D^{(0)} = 0$ and consider

$$(-\frac{\partial}{\partial \Delta} + \mathcal{A} - r(x)) \Big((1 - \Phi(\frac{C^{(-1)}(x)}{\sqrt{\Delta}})) \sum_{k=0}^{\infty} D^{(k)}(x) \Delta^k \Big)$$

Notice that we may reuse formulae (3.38) by replacing $C^{(-1)}$ and $B^{(k)}$ there with $-C^{(-1)}$ and $D^{(k)}$. The rest derivation is exactly the same, hence is omitted.

The new restrictions on $B^{(k)}(x), C^{(k)}(x)$ and $D^{(k)}(x)$ given below can be derived by matching the coefficients before $\Phi(\cdot)$, $\phi(\cdot)$ and Δ^k. For any $k \geq -1$, we have

$$\begin{aligned} 0 =& B^{(0)} - f(x) \\ 0 =& (k+2) B^{(k+2)} + (r(x) + \lambda(x)) B^{(k+1)} - \mathcal{L} B^{(k+1)} - \lambda(x) h_k(x, C^{(-1)}) \\ 0 =& - C^{(0)} \Big(1 + C^{(-1)} \mathcal{L} C^{(-1)} \Big) - \sigma(x) C^{(-1)} \frac{dC^{(0)}(x)}{dx} + B^{(0)} \mathcal{L} C^{(-1)} + \sigma(x) \frac{dB^{(0)}(x)}{dx} \\ 0 =& - C^{(k+2)} \Big(k + 3 + C^{(-1)} \mathcal{L} C^{(-1)} \Big) - \sigma(x) C^{(-1)} \frac{dC^{(k+2)}}{dx} \\ & + \lambda(x)(H_0(k+1, x) - J_0(k+1, x) + G_0(k, x)) + (B^{(k+2)} - D^{(k+2)}) \mathcal{L} C^{(-1)} \\ & + \sigma(x)(\frac{dB^{(k+2)}}{dx} - \frac{dD^{(k+2)}}{dx}) + (\mathcal{L} - r(x) - \lambda(x)) C^{(k+1)} \\ 0 =& - (k+2) D^{(k+2)} + (\mathcal{L} - r(x) - \lambda(x)) D^{(k+1)} \\ & + \lambda(x) \Big(H_1(k, x) - J_1(k, x) + G_1(k, x) + h_{k+1}(x, C^{-1}) - H_2(k+1, x) \\ & + J_2(k+1, x) \Big) \end{aligned}$$

where

$$J_0(k,x) = \sum_{i=0}^{k}\sum_{l=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^l(C^{(-1)}(x))^{l+s}}{(s+2i+1)!!(s)!!(l)!}\frac{\partial^{s+2i+l}j_{k-i}}{\partial\omega^{s+2i+l}}|_{\omega=C^{(-1)}(x)}$$

$$J_1(k,x) = \sum_{i=0}^{k}\sum_{r=0}^{\infty}\frac{(-1)^r(C^{(-1)}(x))^{r}}{(r)!(2i+2)!!}\frac{\partial^{r+2i+1}j_{k-i}}{\partial\omega^{r+2i+1}}|_{\omega=C^{(-1)}(x)}$$

$$J_2(k,x) = \sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\frac{(-1)^s(C^{(-1)}(x))^{r+1+s}}{(r+1)!s!}\frac{\partial^{r+s}j_{k}}{\partial\omega^{r+s}}|_{\omega=C^{(-1)}(x)}$$

and

$$j_k(x,\omega) = -D^{(k)}((C^{(-1)})^{inv}(\omega))\nu((C^{(-1)})^{inv}(\omega)-x)\Big(|\frac{dC^{(-1)}}{dx}|_{(C^{(-1)})^{inv}(\omega)}\Big)^{-1}$$

BIBLIOGRAPHY

Aït-Sahalia, Y. (1999), "Transition Densities for Interest Rate and Other Nonlinear Diffusions," *Journal of Finance*, 54, 1361–1395.

— (2002), "Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223–262.

— (2008), "Closed-Form Likelihood Expansions for Multivariate Diffusions," *Annals of Statistics*, 36, 906–937.

Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. J. (2010a), "Modelling Financial Contagion Using Mutually Exciting Jump Processes," Tech. rep., Princeton University.

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010b), "High-Frequency Covariance Estimates with Noisy and Asynchronous Data," *Journal of American Statistical Association*, 105, 1504–1517.

Aït-Sahalia, Y. and Kimmel, R. (2007), "Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Financial Economics*, 83, 413–452.

— (2009), "Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions," *Journal of Financial Economics, forthcoming*.

Aït-Sahalia, Y. and Mykland, P. A. (2003), "The Effects of Random and Discrete Sampling When Estimating Continuous-Time Diffusions," *Econometrica*, 71, 483–549.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005), "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," *Review of Financial Studies*, 18, 351–416.

— (2009), "Ultra High Frequency Volatility Estimation with Dependent Microstructure Noise," *Journal of Econometrics, forthcoming*.

Aït-Sahalia, Y. and Yu, J. (2009), "High Frequency Market Microstructure Noise Estimates and Liquidity Measures," *Annals of Applied Statistics*, 3, 422–457.

Amemiya, T. (1973), "Regression analysis when the dependent variable is truncated normal," *Econometrica*, 41, 997–1016.

Amengual, D. (2008), "The Term Structure of Variance Risk Premia," Tech. rep., Princeton University.

Andersen, P. and Gill, R. (1982), "Cox's regression model for counting processes: a large sample study," *Annals of Statistics*, 19.

Andersen, T., Bollerslev, T., and Meddahi, N. (2009), "Realized volatility forecasting and market microstructure noise," *Journal of Econometrics, forthcoming*.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Vega, C. (2003), "Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange," *American Economic Review*, 93, 38–62.

Bakshi, G. S. and Ju, N. (2005), "A Refinement to Aït-Sahalia's (2002) "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach"," *Journal of Business*, 78, 2037–2052.

Bandi, F. M. and Russell, J. R. (2003), "Microstructure Noise, Realized Volatility and Optimal Sampling," Tech. rep., University of Chicago Graduate School of Business.

— (2008), "Microstructure Noise, Realized Volatility and Optimal Sampling," *Review of Economic Studies*, 75, 339–369.

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

— (2008b), "Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading," Tech. rep., Department of Mathematical Sciences, University of Aarhus.

Barndorff-Nielsen, O. E. and Shephard, N. (2002), "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, B*, 64, 253–280.

Bates, C. and White, H. (1985), "A unified theory of consistent estimation for parametric models," *Econometric Theory*, 1, 151–178.

Bates, D. S. (1996), "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Review of Financial Studies*, 9, 69–107.

Black, F. and Scholes, M. (1973), "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637–654.

Brunnermeier, M. K., Nagel, S., and Pedersen, L. H. (2009), "Carry Trades and Currency Crashes," *NBER Macroeconomics Annual 2008*, 23, 313–347.

Campbell, J. Y., de Medeiros, K. S., and Viceira, L. M. (2009), "Global Currency Hedging," *Journal of Finance, forthcoming*, –, –.

Carr, P. and Madan, D. B. (1998), "Option Valuation Using the Fast Fourier Transform," *Journal of Computational Finance*, 2, 61–73.

Cox, J. C., , and Ross, S. A. (1976), "The Valuation of Options for Alternative Stochastic Processes," *Journal of Financial Economics*, 3, 145–166.

Cox, J. C. (1975), "The Constant Elasticity of Variance Option Pricing Model," *The Journal of Portfolio Management (1996 special Issue)*, 22, 15–17.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–408.

deB. Harris, F. H., McInish, T. H., Shoesmith, G. L., and Wood, R. A. (1995), "Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets," *Journal of Financial and Quantitative Analysis*, 30, 563–581.

Delattre, S. and Jacod, J. (1997), "A Central Limit Theorem for Normalized Functions of the Increments of a Diffusion Process, in the Presence of Round-Off Errors," *Bernoulli*, 3, 1–28.

Derman, E. and Kani, I. (1994), "Riding on the Smile," *RISK*, 7, 32–39.

Domowitz, I. and White, H. (1982), "Misspecified models with dependent observations," *Journal of Econometrics*, 20, 35–58.

Duffie, D., Pan, J., and Singleton, K. J. (2000), "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.

Dumas, B., Fleming, J., and Whaley, R. E. (1998), "Implied Volatility Functions: Empirical Tests," *Journal of Finance*, 53, 2059–2106.

Dupire, B. (1994), "Pricing with a Smile," *RISK*, 7, 18–20.

Egloff, D., Leippold, M., , and Wu, L. (2010), "The Term Structure of Variance Swap Rates and Optimal Variance Swap Investments," *Journal of Financial and Quantitative Analysis*, 45, 1279–1310.

Egorov, A. V., Li, H., and Xu, Y. (2003), "Maximum Likelihood Estimation of Time Inhomogeneous Diffusions," *Journal of Econometrics*, 114, 107–139.

Epps, T. W. (1979), "Comovements in Stock Prices in the Very Short Run," *Journal of the American Statistical Association*, 74, 291–296.

Fan, J., Li, Y., and Yu, K. (2010), "Vast Volatility Matrix Estimation using High Frequency Data for Portfolio SelectionSelection," Tech. rep., Princeton University.

Fan, J. and Wang, Y. (2007), "Multi-Scale Jump and Volatility Analysis for High-Frequency Financial Data," *Journal of the American Statistical Association*, 102, 1349–1362.

Gatheral, J. and Oomen, R. C. (2009), "Zero-intelligence realized variance estimation," *Finance and Stochastics, forthcoming*, –, –.

Gloter, A. and Jacod, J. (2000), "Diffusions with Measurement Errors: I - Local Asymptotic Normality and II - Optimal Estimators," Tech. rep., Université de Paris-6.

Goldenberg, D. H. (1991), "A unified method for pricing options on diffusion processes," *Journal of Financial Economics*, 29, 3 – 34.

Griffin, J. and Oomen, R. (2008), "Sampling returns for realized variance calculations: Tick time or transaction time?" *Econometric Reviews*, 27, 230–253.

Guillaume, D. M., Dacorogna, M. M., Dave, R. R., Muller, U. A., Olsen, R. B., and Pictet, O. V. (1997), "From the Bird's Eye to the Microscope: A Survey of New Stylized Facts of the Intradaily Foreign Exchange Markets," *Finance and Stochastics*, 1, 95–129.

Hagan, P., Lesniewski, A., and Woodward, D. (2005), "Probability Distribution in the SABR Model of Stochastic Volatility," Tech. rep., Bloomberg LP.

Hansen, P. R., Large, J., and Lunde, A. (2008), "Moving average-based estimators of integrated variance," *Econometric Reviews*, 27, 79–111.

Hansen, P. R. and Lunde, A. (2006), "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127–161.

Hawkes, A. G. (1971), "Point spectra of some mutually exciting point processes," *Journal of the Royal Statistical Society, Series B*, 33, 438–443.

Hayashi, T. and Yoshida, N. (2005), "On Covariance Estimation of Non-synchronously Observed Diffusion Processes," *Bernoulli*, 11, 359–379.

Henry-Labordere, P. (2005), "A General Asymptotic Implied Volatility for Stochastic Volatility Models," Tech. rep., Barclays Capital.

Heston, S. (1993), "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options," *Review of Financial Studies*, 6, 327–343.

Ho, T. S. Y. and Lee, S.-B. (1986), "Term Structure Movements and Pricing Interest Rate Contingent Claims," *Journal of Finance*, 41, 1011–1029.

Hurn, A., Jeisman, J., and Lindsay, K. (2007), "Seeing the Wood for the Trees: A Critical Evaluation of Methods to Estimate the Parameters of Stochastic Differential Equations," *Journal of Financial Econometrics*, 5, 390–455.

Jacod, J. (1994), "Limit of Random Measures Associated with the Increments of a Brownian Semimartingale," Tech. rep., Université de Paris-6.

— (2007), "Statistics and high frequency data: SEMSTAT Seminar," Tech. rep., Université de Paris-6.

— (2008), "Asymptotic properties of realized power variations and related functionals of semimartingales," *Stochastic Processes and their Applications*, 118, 517–559.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009a), "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and their Applications*, 119, 2249–2276.

— (2009b), "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Their Applications*, 119, 2249–2276.

Jacod, J. and Shiryaev, A. N. (2003), *Limit Theorems for Stochastic Processes*, New York: Springer-Verlag, 2nd ed.

Jensen, B. and Poulsen, R. (2002), "Transition Densities of Diffusion Processes: Numerical Comparison of Approximation Techniques," *Journal of Derivatives*, 9, 1–15.

Kalnina, I. and Linton, O. (2008), "Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error," *Journal of Econometrics*, 147, 47–59.

Karatzas, I. and Shreve, S. E. (1991), *Brownian Motion and Stochastic Calculus*, New York: Springer-Verlag.

Kimmel, R. (2009), "Complex Times: Asset Pricing and Conditional Moments under Non-Affine Diffusions," Tech. rep., The Ohio State University.

Kinnebrock, S. and Podolskij, M. (2008), "Estimation of the Quadratic Covariation Matrix in Noisy Diffusion Models," Tech. rep., Oxford-Man Institute, University of Oxford.

Kou, S. (2002), "A jump-diffusion model for option pricing," *Management Science*, 48, 1086–1101.

Kristensen, D. and Mele, A. (2010), "Adding and Subtracting Black-Scholes: A New Approach to Approximating Derivative Prices in Continuous-Time Models," Tech. rep., Columbia University.

Large, J. (2007), "Accounting for the Epps Effect: Realized Covariation, Cointegration and Common Factors," Tech. rep., Oxford-Man Institute, University of Oxford.

Li, Y., Mykland, P., Renault, E., Zhang, L., and Zheng, X. (2010), "Realized Volatility When Sampling Times are Possibly Endogenous," Tech. rep., Hong Kong University of Science and Technology.

Li, Y. and Mykland, P. A. (2007), "Are Volatility Estimators Robust with Respect to Modeling Assumptions?" *Bernoulli*, 13, 601–622.

McCullagh, P. (1987), *Tensor Methods in Statistics*, London, U.K.: Chapman and Hall.

Mencia, J. and Sentana, E. (2009), "Valuation of VIX Derivatives," Tech. Rep. 0913, CEMFI.

Merton, R. C. (1976), "Option Pricing when Underlying Stock Returns are Discontinuous," *Journal of Financial Economics*, 3, 125–144.

Muthuswamy, J., Sarkar, S., Low, A., and Terry, E. (2001), "Time Variation in the Correlation Structure of Exchange Rates: High Frequency Analysis," *Journal of Futures Markets*, 21, 127–144.

Mykland, P. A. and Zhang, L. (2009), "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1445.

Newey, W. K. and McFadden, D. L. (1994), *Large sample estimation and hypothesis testing. In: Handbook of Econometrics*, vol. IV, Elsevier.

Pan, J. (2002), "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3–50.

Rockafellar, R. (1970), *Convex Analysis*, Princeton University Press.

Rubinstein, M. (1995), "As Simple as One, Two, Three," *RISK*, 8, 44–47.

Schroder, M. (1989), "Computing the Constant Elasticity of Variance Option Pricing Formula," *The Journal of Finance*, 44, 211–219.

Scott, L. O. (1997), "Pricing Stock Options in a Jump Diffusion Model with Stochastic Volatility and Interest Rates: Applications of Fourier Inversion Methods," *Mathematical Finance*, 7, 413–426.

Stone, M. H. (1927), "Developments in Hermite Polynomials," *Annals of Mathematics*, 29, 1–13.

Sun, Y. (2006), "Best quadratic unbiased estimators of integrated variance in the presence of market microstructure noise," Tech. rep., University of California at San Diego.

Tang, K. and Xiong, W. (2010), "Index Investment and Financialization of Commodities," Tech. rep., Princeton University.

Voev, V. and Lunde, A. (2007), "Integrated Covariance Estimation Using High-Frequency Data in the Presence of Noise," *Journal of Financial Econometrics*, 5, 68–104.

White, H. (1980), "Nonlinear regression on cross-section data," *Econometrica*, 48, 721–746.

— (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

Xiu, D. (2010), "Quasi-Maximum Likelihood Estimation of Volatility with High Frequency Data," *Journal of Econometrics*, 159, 235–250.

Yu, J. (2007), "Closed-Form Likelihood Estimation of Jump-Diffusions with an Application to the Realignment Risk Premium of the Chinese Yuan," *Journal of Econometrics*, 141, 1245–1280.

Zhang, L. (2006), "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12, 1019–1043.

— (2009), "Estimating Covariation: Epps Effect and Microstructure Noise," *Journal of Econometrics, forthcoming*.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

— (2009), "Edgeworth Expansions for Realized Rolatility and Related Estimators," *Journal of Econometrics, forthcoming*.

Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business & Economic Statistics*, 14, 45–52.