ABSTRACT

CUI, WENHAO. Essays on Financial and Time Series Econometrics. (Under the direction of Denis Pelletier).

The three essays of this dissertation consider the volatility estimation using high frequency financial data and sign restricted SVARs model.

In Chapter 2, we study a class of nonparametric volatility estimators based on the Laplace transform, which are robust to the presence of both microstructure noise and the endogeneity of observation times. Asymptotic properties and feasible central limit theorems are established. In the presence of time endogeneity, our bias-corrected Laplace estimator takes advantage of the informational content of time endogeneity, which leads to an efficiency gain and narrower confidence bounds. The finite sample properties of the estimator are studied through Monte Carlo simulations. The performance of the Laplace estimator is compared with other commonly used estimators through forecasting exercises by employing high frequency data. We conclude that the naive and bias-corrected Laplace estimator perform better than most estimators in terms of forecasting equity return volatility.

Chapter 3 studies the efficient estimation problem of the volatility of asset returns in the presence of both market microstructure noise and time endogeneity using the realized kernel. Central limit theorem of the realized kernel estimator are established under a general nonparametric setting. We further show that the realized kernel estimator can be made to converge at the fastest possible rate, robust to different specifications of the noise and robust to dynamics in the noise process. The finite sample properties of the estimator are studied through Monte Carlo simulations and feasible confidence bounds of the proposed estimator are constructed, while empirical work illustrates its use in practice.

In Chapter 4, the main contribution is the study of four different parameterizations of SVARs and making probability statements across different parameterizations. We also address the problem that in a sign restricted SVAR model, the data provide no information to distinguish between any two pairs of structural parameters in the admissible set. We propose two ways to solve this problem: combining sign restrictions with exclusive zero restrictions, and constructing a joint credible set that is robust to different prior specifications based on our main result. We then demonstrate our approach with an empirical example where we identify monetary policy shocks.

Essays on Financial and Time Series Econometrics

by
Wenhao Cui

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2019

APPROVED BY:

Ilze Kalnina	Giuseppe Fiori
Harrison Fell	Denis Pelletier Chair of Advisory Committee

DEDICATION

To my parents Long Cui and Guiyun Zhang.

BIOGRAPHY

Wenhao Cui was born and raised in Urumqi, Xinjiang, China. In 2014, he received his bachelor's degree in International Finance from Fudan University, Shanghai, China. In the same year, he joined North Carolina State University to pursue his Ph.D. degree in Economics under the direction of Professor Denis Pelletier. His research interests include Financial Econometrics and Time Series Econometrics.

ACKNOWLEDGEMENTS

I would like to thank my advisor Professor Denis Pelletier for his insightful guidance in my research and also teaching during my Ph.D. study. He is a perfect example for me as a rigorous and wise scholar. I also want to thank Professor Boke Jiang, who is my undergraduate advisor and helped me a lot during my undergraduate study.

I am very grateful to my committee members Professor Ilze Kalnina, Professor Giuseppe Fiori and Professor Harrison Fell for their time and valuable comments. I also want to thank our current program director Professor Xiaoyong Zheng for his effort to provide good research environment and his great support in my job search process. I also benefited a lot from discussions with Professor James Nason on Chapter 4. I would like to thank Professor Kelly Lynn Mulvey for agreeing to be my graduate school representatives. Many thanks to the faculties and staff in Department of Economics and Department of Agricultural and Resource Economics at NC State University.

Lastly but most importantly, I want to thank my family and friends. I want to thank my parents for their unconditional support and love. I would like to thank my girlfriend, Shuyang Wang for her support, encouragement, company and for every moments we shared together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Over the last several decades especially the last 20 years, the domain of econometric methods for high frequency financial data and time series data have been experiencing an exponential growth. This is due to the increasing availability of high frequency financial data and the development of new mathematical tools to analyze these data. The motivation of this work is to develop new tools and contribute to the existing literature.

In Chapter 2, we consider the problem of estimating volatility under the presence of time endogeneity. Volatility is a key parameter in financial economics and the recent financial econometrics literature has devoted much attention to its estimation. As a matter of fact, volatility estimation is a crucial component in the evaluation of financial risk management, in hedging strategies, in forecasting and for portfolio selection. In Chapter 2, we study a class of kernel estimators based on the Laplace transform. Such a class of volatility estimators are motivated by two challenges. The first is the presence of microstructure noise in the high frequency data. It has been widely recognized that using very high frequency data requires taking into account the effect of market microstructure noise. Another challenge in the high frequency data that we address is the time endogeneity of observation times (nonzero covariance between price process and time). It is often assumed in the literature that the observation times are independent of the prices process. In this case, the error of the volatility estimator usually follows a mixed normal distribution. However, in practice, the independence assumption is rarely satisfied and there is endogeneity in the price sampling times. The time between two observations might be informative about the price process itself, with trades occurring more or less frequently depending upon the information flow, which itself will generally influence the efficient price process. According to our result, after taking time endogeneity into account, the limit distribution of the normalized error contains a random "bias" term that can be consistently estimated. Given that we can consistently estimate the asymptotic "bias" term, we propose a new bias-corrected Laplace estimator by subtracting the estimated bias term. Consistently estimating the bias allows taking advantage of the informational content of time endogeneity, which leads to an efficiency gain and narrower confidence bounds. The finite sample properties of the Laplace estimator are analyzed through Monte Carlo studies. We first apply our result to equidistant data and study the optimal choice of the convolution frequency and leverage effect in a model without microstructure noise and conclude that the Laplace estimator could be adopted in a model with leverage effect. We then study the performance of the Laplace estimator in a model with noise components and conclude that by choosing the convolution frequency carefully, the Laplace estimator has the ability to filter out high frequency noise components. Next, we evaluate the performance of confidence intervals of the bias-corrected Laplace estimator and confirm the efficiency gain when observation times are possibly endogenous. Last, we apply our methodology to the millisecond time stamped trade data in different forecasting exercises. We compare the performance of the Laplace estimator of forecasting equity return volatility with other commonly used estimators, such as the realized volatility, bi-power variation, subsampled version of RV and two-scale RV. We conclude that in the case of equidistant observations, the Laplace estimator performs just as well as (if not superior to) all the other estimators. When it comes to the case that the observation times being possibly endogenous, our results show that the bias-corrected Laplace estimator gives the most accurate forecasts.

Chapter 2 is organized as follows. Section 2.2 introduces the Laplace estimator, the asymptotic properties are analyzed in Section 2.3. Section 2.4 derives feasible CLTs to construct confidence bounds or to compute p-value. Simulation studies are carried out in Section 2.5, and forecasting exercises are presented in Section 2.6. We conclude in Section 2.7.

In Chapter 3, we follow the same line as in Chapter 2 by considering the problem of estimating volatility under the presence of both time endogeneity and microstructure noise. In Chapter 3 we will focus on the realized kernel estimator of the Integrated Volatility, which is equivalent to the frequency domain estimator. The aim of this paper is to study the robustness of the realized kernel estimator of the Integrated Volatility under time endogeneity and the presence of microstructure noise. As a starting point, we assume the noise component is represented by independent identically distributed random variables. We derive the explicit analytical expression of the bias of the realized kernel estimator and show that the resulting estimator after correcting the bias is asymptotically mixed Gaussian with a fastest convergence rate at $N^{1/5}$. According to our result, the effect of time endogeneity will not deteriorate the convergence rate but will affect the asymptotic variance of the limiting process. We also consider the performance of the flat-top realized kernel estimator as an extension under time endogeneity. Our finding is that the convergence rate of the flat-top kernel estimator is unaffected but depends heavily on the white noise assumption of the noise component. Next, we relax the white noise assumption to allow the microstructure noise to be serially dependent, so that the realized kernel estimator could be applied to the ultra high frequency data that are sampled at the highest possible frequency.

We conclude that under the case of the dependent noise, only the asymptotic variance of the limiting process will be affected and the fastest convergence rate is $N^{1/5}$. To dealt with the case that the microstructure noise interacts with the sampling frequency and the empirical evidence that microstructure noise is small relative to the Integrated Volatility, we also consider the case of a shrinking noise. Under the case of a shrinking noise, we can achieve a faster convergence rate than $N^{1/5}$. The fastest convergence rate depends on the rate of the shrinkage of the noise. We also conduct simulation study to examine the finite sample property of the realized kernel estimator under time endogeneity and the presence of the microstructure noise. Our results confirm the finding that the realized kernel estimator could be adopted in a model with both time endogeneity and dependent noise. Another surprising finding is that, under the case of a small noise-to-signal ratio, the realized kernel estimator based on the shrinking noise performs better even if the data generating process admits a non-shrinking noise in finite sample. We also make the central limit results feasible and construct confidence bounds of the proposed estimator. Last, we apply our methodology to the millisecond time stamped trade data and illustrate that the realized kernel estimator could be adopted in a model that are sampled at the highest possible frequency.

Chapter 3 is organized as follows. Section 3.2 introduces our version of the realized kernel estimator and discusses the connection between it and the frequency domain estimator. We also study its asymptotic properties under the presence of time endogeneity and the absence of the microstructure noise. Section 3.3 studies the asymptotic property of the realized kernel estimator under time endogeneity and the presence of microstructure noise. In Section 3.4, we discuss three different extensions: the flat-top realized kernel estimator, the case of the dependent noise and the case of the shrinking noise. Simulation studies are carried out in Section 3.5, and an empirical analysis is presented in Section 3.6. We conclude in Section 3.7.

In Chapter 4, we study a sign restricted SVAR model. The structural vector autoregressive models (SVAR) has been the workhorse for analyzing the dynamics caused by macroeconomic shocks. More often than not, a vector autoregressive model (VAR) is first fitted to the data, and then structural equation parameters are recovered. However, a reduced-form vector autoregressive model can not pin down a unique structural vector autoregressive model. Identification restrictions are needed so that the parameters in the structural equations can be recovered. It was proposed that inference using the structural vector autoregressive model might be based on prior beliefs about the signs of the impacts of certain shocks. This method is commonly referred to as sign restrictions and it has become part of the mainstream of empirical macroeconomics. We consider four different parameterizations of an SVAR: reduced-form parameterization, structural parameterization, reduced-form MA parameterization and IRFs (structural MA) parameterization. We focus on making probability statement about all four different parameterizations using the change of variables formula based on the theory of exterior product. The main contribution is the calculation of the determinants of the Jacobian matrices between different parameterizations. We also address the problem that a sign restricted structural vector autoregressive model is generally set identified, the data provide no information to distinguish between any two pairs of structural parameters in the admissible set. One possible way to tackle this problem is to shrink the admissible set by imposing more restrictions, such as combining sign restrictions with zero restrictions. We suggest treating zero restrictions as soft restrictions in the form of prior information and developing an efficient importance sampler to sample from the posterior distribution. Another way to tackle the problem is to construct a joint credible set which is robust to different choices of priors. We construct a robust joint credible set based on the posterior density of impulse responses.

The structure of Chapter 4 is as follows. In Section 4.2, we describe four parameterizations of SVARs and the identification. In Section 4.3, we provide a detailed definition of the Haar measure on the orthogonal group, i.e., a flat prior and then present our main result. In Section 4.4, we illustrate two applications of our main result and focus on solving the problem associated with sign restrictions. In Section 4.5, we demonstrate our method with an empirical example. In Section 4.6, we conclude.

Chapter 2

Laplace Estimator of Integrated Volatility When Sampling Times Are Endogenous

We study a class of nonparametric volatility estimators based on the Laplace transform, which are robust to the presence of both microstructure noise and the endogeneity of observation times. Asymptotic properties and feasible central limit theorems are established. In the presence of time endogeneity, our bias-corrected Laplace estimator takes advantage of the informational content of time endogeneity, which leads to an efficiency gain and narrower confidence bounds. The finite sample properties of the estimator are studied through Monte Carlo simulations. The performance of the Laplace estimator is compared with other commonly used estimators through forecasting exercises by employing high frequency data. We conclude that the naive and bias-corrected Laplace estimator perform better than most estimators in terms of forecasting equity return volatility.

2.1 Introduction

Volatility is a key parameter in financial economics and the recent financial econometrics literature has devoted much attention to its estimation. As a matter of fact, volatility estimation is a crucial component in the evaluation of financial risk management, in hedging strategies, in forecasting and for portfolio selection.

Volatility can be estimated with parametric or nonparametric methods. When using parametric methods, the volatility is usually modeled as a functional form of market or latent variables. On the other hand, nonparametric methods address the estimation of the historical volatility without explicitly assuming a functional form of the volatility. The estimation of volatility usually focus on a small time window (e.g., a day) employing high frequency data.

In this paper, we study a class of kernel estimators based on the Laplace transform. Such a class of volatility estimators are motivated by two challenges. The first is the presence of microstructure noise in the high frequency data. It has been widely recognized that using very high frequency data requires taking into account the effect of market microstructure noise. A large amount of literature uses realized volatility as a nonparametric measure of volatility.[1] In the absence of market microstructure noise, it is a consistent estimator of the quadratic variation as the time between observations goes to zero. However, as sampling frequency increases, the cumulative effect of noise components starts dominating the efficient price process, and leads to the divergence of the realized volatility. A large number of alternative estimators have been proposed in the literature to tackle the microstructure noise, such as the sub-sampling method ([77]), the pre-averaging method ([50]), the realized kernel methods ([16]) and the multi-scale estimators ([78]).

Another challenge in the high frequency data that we address is the time endogeneity of observation times (nonzero covariance between price process and time). It is often assumed in the literature that the observation times are independent of the prices process. In this case, the error of the volatility estimator usually follows a mixed normal distribution. However, in practice, the independence assumption is rarely satisfied and there is endogeneity in the price sampling times. The time between two observations might be informative about the price process itself, with trades occurring more or less frequently depending upon the information flow, which itself will generally influence the efficient price process. In [58], they extend realized volatility to a framework with time endogeneity and establish a central limit theorem. Later, [69] introduce a general endogenous bivariate nonparametric framework (price process is two-dimensional) and establish a central limit theorem for the Hayashi–Yoshida estimator.[2] In this paper, we extend the case of time endogeneity to the class of kernel estimators based on the Laplace transform and establish feasible central limit theorems.

According to the fundamental theorem of asset pricing (see [31]), the price process should follow a semi-martingale. Suppose the log-price of some asset prices whose evolution is given by an Itô process

$$dX_t = \mu_t \, dt + \sigma_t \, dW_t, \tag{2.1.1}$$

where W_t is a Wiener process. In this framework a natural measure of volatility over the time interval $[0, T]$ is the integrated volatility $\int_0^T \sigma_t^2 \, dt$. Under this typical setting, we derive the Laplace estimator of integrated volatility given discrete observations where observation times could be possibly endogenous. We show that the Laplace estimator has the ability to filter out

[1] Also called realized variance.

[2] The Hayashi–Yoshida estimator is free of any "synchronization" processing of the original data, hence free of bias or other problems caused by it. See [44].

the microstructure noise using Monte Carlo simulations. Robustness of the Laplace estimator arises due to the presence of the realized kernel. We prove the convergence and the asymptotic normality of the estimator when observation times are possibly endogenous.

In the paper, we consider two cases: when observation times are irregularly spaced (Theorem 2.3.2) and endogenous (Theorem 2.3.4). In case of irregularly spaced observation time, we show that the asymptotic limit of $(1/HN)$ calibrates the estimator between the Laplace estimator and the realized volatility.[3] Next, we consider the more general case of observation times that are possibly endogenous. According to our result, after taking time endogeneity into account, the limit distribution of the normalized error contains a random "bias" term that can be consistently estimated. Given that we can consistently estimate the asymptotic "bias" term, we propose a new bias-corrected Laplace estimator by subtracting the estimated bias term. Consistently estimating the bias allows taking advantage of the informational content of time endogeneity, which leads to an efficiency gain and narrow confidence bounds. We also consider a special case that the observations are evenly spaced, which is the most commonly used sampling scheme in the literature. We show that the naive Laplace estimator achieves the same efficiency level as the realized volatility when the convolution frequency is appropriately selected, which indicates the presence of the rescaled Dirichlet kernel associate with the naive Laplace estimator is not detrimental to the asymptotic variance and convergence rate.

The finite sample properties of the Laplace estimator are analyzed through Monte Carlo studies. We first apply our result to equidistant data and study the optimal choice of the convolution frequency and leverage effect in a model without microstructure noise and conclude that the Laplace estimator could be adopted in a model with leverage effect. We then study the performance of the Laplace estimator in a model with noise components and conclude that by choosing the convolution frequency carefully, the Laplace estimator has the ability to filter out high frequency noise components. Next, we evaluate the performance of confidence intervals of the bias-corrected Laplace estimator and confirm the efficiency gain when observation times are possibly endogenous.

Last, we apply our methodology to the millisecond time stamped trade data in different forecasting exercises. We compare the performance of the Laplace estimator of forecasting equity return volatility with other commonly used estimators, such as the realized volatility, bi-power variation, subsampled version of RV and two-scale RV. We conclude that in the case of equidistant observations, the Laplace estimator performs just as well as (if not superior to) all the other estimators. When it comes to the case that the observation times being possibly endogenous, our results show that the bias-corrected Laplace estimator gives the most accurate forecasts.

The paper is organized as follows. Section 2.2 introduces the Laplace estimator, the asymptotic

[3] H denotes the inverse of convolution frequency and N the number of observations.

properties are analyzed in Section 2.3. Section 2.4 derives feasible CLTs to construct confidence bounds or to compute p-value. Simulation studies are carried out in Section 2.5, and forecasting exercises are presented in Section 2.6. We conclude in Section 2.7.

2.2 Laplace Estimator of Integrated Volatility

The aim of this section is to motivate and introduce a class of kernel estimators based on the Laplace transform. We compute the Laplace transform of the latent volatility process given a continuous trajectory of the price process. Given a discrete, possibly unevenly spaced or endogenous sampling of the price process, we study the asymptotic properties of the Laplace estimator of integrated volatility in Section 2.3.

We first describe the setting underlying most of the volatility measures. The evolution of the log-price is described by a standard Itô process,

$$dX_t = \mu_t \, dt + \sigma_t \, dW_t, \tag{2.2.1}$$

where W_t is a Brownian motion on a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. It is further assumed that:

Assumption H2.1. *The process μ_t and σ_t are almost surely continuous in $[0,\infty)$ and adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$.*

Remark 2.2.1. Assumption H2.1 is only required to construct feasible confidence intervals in Section 2.4. For a minimal condition that guarantees the existence of the Laplace estimator, we can relax Assumption H2.1 to the following condition.

Assumption H2.1′. *The process μ_t and σ_t are adapted stochastic processes such that*

$$E\left[\int_0^T \sigma_t^4 \, dt\right] < \infty, \quad E\left[\int_0^T \mu_t^2 \, dt\right] < \infty.$$

The target of inference is the integrated volatility

$$\langle X, X \rangle_t = \int_0^t \sigma_u^2 \, du. \tag{2.2.2}$$

One of the most commonly used estimators of integrated volatility is the realized volatility developed by [14].[4] The realized volatility and its variations, e.g., subsampled version of RV and two-scale RV have been used and studied extensively in the literature. [64] extend the

[4]The asymptotic property of the realized volatility was first studied in [14]. The realized volatility has been developed earlier using the theory of quadratic variation. See [53] and [4].

realized volatility to handle irregularly spaced data. Noise robust volatility measures have also been developed, see [77], [16], [61] and [1] for different extensions. [58] study realized volatility in a model with time endogeneity.[5] Among the aforementioned estimators, the Fourier based estimator and the realized kernel estimators are especially attractive for their natural robustness to the microstructure noise. Motivated by this, we study a class of kernel estimators based on the Laplace transform to overcome the presence of time endogeneity and microstructure noise.

We now introduce and explain the Laplace estimator of integrated volatility. The Laplace transform of the price increment is defined by[6]

$$\mathcal{L}(dX)(a+is) = \int_{-\infty}^{0} exp\left((a+is)t\right)\, dX_t, \tag{2.2.3}$$

where $a \geq 0$ and $i = \sqrt{-1}$. The Bohr convolution product between two complex functions Φ and Ψ is defined as

$$(\Phi *_B \Psi)(v) = \lim_{R\to+\infty} \frac{1}{2R}\int_{-R}^{R} \Phi(a+is)\bar{\Psi}\left(a+i(s+v)\right)\, ds, \tag{2.2.4}$$

where $\bar{\Psi}$ denotes the conjugate of the complex function Ψ. By Theorem 2.1 in [30], as $R \to +\infty$, we obtain the following convergence in probability,

$$\left(\mathcal{L}(dX)(a+is) *_B \mathcal{L}(dX)(a+is)\right)(v) \xrightarrow{p} \int_{-\infty}^{0} exp\left((2a-iv)t\right)\sigma_t^2\, dt. \tag{2.2.5}$$

In this paper, we will consider the special case where $a = 0$, and $v = 0$, so that the right-hand-side of Eq. (2.2.5) equals to the integrated volatility as in Eq. (2.2.2).

Let $(t_{n,i})_{i=0,\ldots,N}$ be observations times in the interval $[0,T]$, where n characterizes the observation frequency and N stands for the actual number of observations before time T, for simplicity we assume $t_{n,N} = T$ and for notation clarity we always suppress n. Let $\Delta X_{t_i} = X_{t_{i+1}} - X_{t_i}$ and $\Delta t_i = t_{i+1} - t_i$. Thus the sample analog of the left-hand-side of Eq. (2.2.5) is given by

$$\hat{C}_T = \frac{1}{2R}\int_{-R}^{R}\sum_{t_{i+1}\leq T}\sum_{t_{j+1}\leq T} exp(ist_i - ist_j)\Delta X_{t_i}\Delta X_{t_j}\, ds. \tag{2.2.6}$$

Remark 2.2.2. Applying the Euler formula and integrating over s, we obtain another representation of the naive Laplace estimator in Eq. (2.2.6),

$$\hat{C}_T = \sum_{t_{i+1}\leq T}\sum_{t_{j+1}\leq T}\frac{sin\left(R(t_i-t_j)\right)}{R(t_i-t_j)}\Delta X_{t_i}\Delta X_{t_j}, \tag{2.2.7}$$

[5] They call time endogeneity a situation in which observation times implies a nonzero limit for tricity, see [58].
[6] Here we follow [30] and use a unilateral Laplace transform.

which is generally referred as the rescaled Dirichlet kernel estimator (see, e.g., [16]).

We consider a class of kernel estimators based on the Laplace transform. More specifically, we consider a class of spectral window kernel $K_R(\cdot)$ being added to the naive Laplace estimator in Eq. (2.2.6), with $K_R(s) = K\left(\frac{s}{R}\right)$.[7] We define the spectral window kernel $K(\cdot)$ as a Laplace transform of the lag window (realized kernel). After taking the Bohr convolution product of $K(\cdot)$, we have a class of kernel estimators based on the Laplace transform, which takes the following form,

$$\begin{aligned}\bar{C}_T &= \frac{1}{2R}\int_{-R}^{R}\sum_{t_{i+1}\leq T}\sum_{t_{j+1}\leq T} K_R(s)exp(ist_i - ist_j)\Delta X_{t_i}\Delta X_{t_j}\,ds \\ &= \sum_{t_{i+1}\leq T}\sum_{t_{j+1}\leq T} k\left(\frac{t_i - t_j}{H}\right)\Delta X_{t_i}\Delta X_{t_j}, \end{aligned} \tag{2.2.8}$$

where $H = 1/R$, $k\left(\frac{t}{H}\right) = \frac{1}{2R}\int_{-R}^{R} K_R(s)exp(ist)\,ds$ and $\frac{1}{2R}\int_{-R}^{R} K_R(s)\,ds = 1$. Here $k(\cdot)$ is the lag window kernel, also referred as the realized kernel with infinite lags.[8]

2.3 Asymptotic Properties

In this section we prove the weak convergence of the Laplace estimator of integrated volatility defined in Eq. (2.2.8) and the asymptotic normality for the error of the estimator. Throughout the section, we consider two different cases: when observation times are irregularly spaced but exogenous (Theorem 2.3.2) and the case of endogenous (Theorem 2.3.4). The main concept is stable convergence.[9]

Definition 2.3.1. Suppose that all relevant processes (X_t, μ_t and σ_t) are adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$. Let Z_n be a sequence of $\mathcal{F}_T$ -measurable random variables. We say that Z_n converges stably in law to Z ($Z_n \xrightarrow{\mathcal{L}-s} Z$) as $n \to \infty$ if Z is measurable with respect to an extension of $\mathcal{F}_T$ so that for all $A \in \mathcal{F}_T$ and for all bounded continuous g, $E\left(\mathbb{1}_A g(Z_n)\right) = E\left(\mathbb{1}_A g(Z)\right)$ as $n \to \infty$, where $\mathbb{1}_A$ is the indicator function of A. The same definition applies to triangular arrays.

Stable convergence in law is stronger than the usual convergence in law (convergence in distribution), and is crucial for our analysis in Section 2.4 for the feasible central limit theorems.

---

[7]See [67] for an application of the spectral window generator in the context of Fourier based estimators.

[8]Note that the realized kernel defined here is slightly different the realized kernel in [16] and [18]. They define the kernel to be $k\left(\frac{i-j}{\bar{H}}\right)$, which is the case of equidistant observations here. Our results are comparable by noticing that $\bar{H} = O_p(HN)$.

[9]Here we follow the same definition as [64] and [63]. See also [3].

2.3.1 When sampling times are irregularly spaced

In this section, we take the time points to be nonrandom, but possibly irregularly spaced, this also covers the case where the t_i are random but independent of the process X_t. In this case, we need the following assumptions on the variation of time to analyze the asymptotic property of the Laplace estimator.

Assumption H2.2. $\max_i(\Delta t_i) \to 0$ *as* $N \to \infty$, *and* $\frac{\max_i(\Delta t_i)}{T/N} = O(1)$.

Assumption H2.3. *As* $\frac{1}{HN} \to c$ *and* $H \to 0$, $N \to \infty$, *where* c *is a nonnegative constant and let* $t^*(\tau)$ *be the largest time* t_i *smaller than or equal to* τ, *we have the convergence of the sequence of the functions,*

$$H^{-1} \int_0^t \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right)^2 du\, dv \to \int_0^t \Psi(v)\, dv, \tag{2.3.1}$$

for all $t \in [0, T]$ *and* Ψ *is some integrable function.*

Assumption H2.2 and H2.3 are the standard assumptions when dealing with irregularly spaced data. Assumption H2.2 is commonly assumed in the literature, however it may be weakened to allow Poisson type sampling schemes (see [58] for the case of time endogeneity and [18] in the presence of micro-structure noise).

Theorem 2.3.2. *Under the Assumptions H2.1 – H2.3, and that* $\frac{1}{HN} \to c$ *and* $H \to 0$, $N \to \infty$, *where* c *is a nonnegative constant. Then we have*

$$H^{-1/2}(\bar{C}_T - \langle X, X\rangle_T) \xrightarrow{\mathcal{L}-s} 2\int_0^T \sqrt{\Psi(v)}\sigma_v^2\, dB_v, \tag{2.3.2}$$

where the convergence is stable in law, B_t *is a Brownian motion independent of the underling* σ*-field.*

Remark 2.3.3. In Theorem 2.3.2, the fastest convergence rate of the Laplace estimator is of $N^{1/2}$. It is also worth noting that, we only consider the case that $\frac{1}{HN} \to c$ and $H \to 0$, $N \to \infty$. If we have the trivial case that $\frac{1}{HN} \to \infty$, the Laplace estimator is asymptotically equivalent to the realized volatility, due to $\lim_{H\to 0} k\left(\frac{t_i - t_j}{H}\right) = 0$, for $i \neq j$. As a conclusion, the value of c enables us to calibrate the estimator between the Laplace based estimator and the usual realized volatility. On the other hand, the case $c = 0$ is an important case that is covered in Theorem 2.3.2. Under the presence of microstructure noise, the choice of $c = 0$, enables one to filter out the highest frequency signals (microstructure noise).

2.3.2 When sampling times are endogenous

In this section, we consider the most general case where the sampling times are possibly endogenous. This is a more general setting that includes the case where there is no time endogeneity. In this paper, we call time endogeneity a situation in which the covariance between the kernel weighted return process ($\sum_{t_{i+1}\leq T} k\left(\frac{t_i-t_j}{H}\right)\Delta X_{t_i}$) and the observation frequency (Δt_j) is non-zero. We require the following assumptions.

Assumption H2.4. *As $H \to 0$ and $N \to \infty$ and that $\frac{1}{HN} \to c$, where c is a nonnegative constant, then the following convergence in probability hold*

$$H^{-1/2} \sum_{t_{j+1}\leq T} \left(\sum_{t_{i+1}\leq T} k\left(\frac{t_i - t_j}{H}\right) \Delta X_{t_i} \right) \Delta t_j \xrightarrow{p} \int_0^T a_v \sigma_v \, dv, \tag{2.3.3}$$

$$H^{-1} \sum_{t_{j+1}\leq T} \left(\sum_{t_{i+1}\leq T} k\left(\frac{t_i - t_j}{H}\right) \Delta X_{t_i} \right)^2 \Delta t_j \xrightarrow{p} \int_0^T b_v \sigma_v^2 \, dv, \tag{2.3.4}$$

for some processes a_v and b_v.

Assumption H2.4 could be viewed as our definition of time endogeneity. The convergence of this covariance in Eq. (2.3.3) is crucial for the consistency of the Laplace estimator, Eq. (2.3.4) is also needed to derive the limiting distribution of the Laplace estimator. Assumption H2.4 is satisfied under the case of no time endogeneity with $a_v = 0$ and $b_v = 2\Psi(v)$. Generally, Eq. (2.3.3) and (2.3.4) could be checked easily under the specific endogenous sampling scheme using Chebyshev's inequality under the maintained Assumptions H2.1 – H2.3.

Theorem 2.3.4. *Under the Assumptions H2.1 – H2.4, and that $\frac{1}{HN} \to c$ and $H \to 0$, $N \to \infty$, where c is a nonnegative constant. Then we have*

$$H^{-1/2}(\bar{C}_T - \langle X, X\rangle_T) \xrightarrow{\mathcal{L}-s} 2\int_0^T a_v\sigma_v \, dX_v + 2\int_0^T \sqrt{b_v - \Psi(v) - a_v^2\sigma_v^2} \, dB_v, \tag{2.3.5}$$

where the convergence is stable in law, B_v is a Brownian motion independent of the underling σ-field.

Remark 2.3.5. The fastest rate of convergence is still $N^{1/2}$, but now an asymptotic bias term shows up due to the time endogeneity (a nonzero probability limit of the covariance), which suggests an improvement of the Laplace estimator, by correcting the bias,

$$\bar{C}_T^{BC} = \bar{C}_T - 2H^{1/2} \int_0^T \widehat{a_v\sigma_v} \, dX_v, \tag{2.3.6}$$

where $\int_0^T \widehat{a_v\sigma_v} \, dX_v$ is a consistent estimator of $\int_0^T a_v\sigma_v \, dX_v$.

Remark 2.3.6. Theorem 2.3.4 does not change the integrated variance of the limiting process comparing to the non-endogenous case. As shown in the Appendix A.1, in the absence of time endogeneity, we have $b_v = 2\Psi(v)$, and so that the variance of the limiting process in both Theorem 2.3.2 and 2.3.4 are equal to $4\int_0^T \left(b_v - \Psi(v)\right)\sigma_v^4\, dv$. However, in Theorem 2.3.4 it reapportions $2\int_0^T a_v\sigma_v\, dX_v$ from asymptotic variance to asymptotic mean. This property is first discussed in [58] in the context of realized volatility. Here, we refer it as the variance reduction property of $\bar{C}_T^{BC}$, and we provide Monte Carlo evidence in Section 2.5.4. The intuition behind the variance reduction is that by consistently estimating the bias, it allows us to take advantage of the informational content of time endogeneity, which leads to an efficiency gain with respect to the naive Laplace estimator.

2.3.3 Efficiency of the Laplace estimator

In this section, we revisit the simplest and the most commonly adopted sampling scheme in practice, sampling on an equidistant grid. The aim of this section is to show that the naive Laplace estimator $\hat{C}_T$ defined in Eq. (2.2.6) can achieve the same level of efficiency as the realized volatility, which is known as one of the most efficient nonparametric estimator.

Theorem 2.3.7. *Under the Assumption= H2.1, and that $\frac{R}{N} \to c$ as $R,\, N \to \infty$, where c is a positive constant. Then we have*

$$\sqrt{\frac{N}{T}}(\hat{C}_T - \langle X, X\rangle_T) \xrightarrow{\mathcal{L}-s} \int_0^T \sqrt{2(1+2\eta(c))}\sigma_t^2\, dB_t, \tag{2.3.7}$$

where the convergence is stable in law, B_t is a Brownian motion independent of the underling σ-field and the constant $\eta(c)$ is equal to

$$\eta(c) = \frac{1}{2\tilde{c}^2} r(\tilde{c})\left(1 - r(\tilde{c})\right), \tag{2.3.8}$$

where $\tilde{c} = \frac{cT}{\pi}$ and $r(x) = x - [x]$, with $[x]$ the integer part of x.

Remark 2.3.8. When $c = (\pi/T)k$, $k = 1, 2, \ldots$, we attain the smallest asymptotic variance for the Laplace estimator which is equal to $\int_0^T 2\sigma_t^4\, dt$, and this is exactly the same asymptotic variance as the realized volatility established in [14],

$$RV = \sum_{t_{i+1}\le T} \Delta X_{t_i}^2. \tag{2.3.9}$$

The rescaled Dirichlet kernel appears as a consequence of the convolution product, and is exclusive to harmonic analysis methods such as the Fourier based estimator (see [59], [60], [61] and [26]). Thus the Laplace estimator has similar features to the Fourier estimator, i.e., robustness to irregular-asynchronous observations and the ability to filter out the microstructure noise by a suitable choice of R (recall that $H = R^{-1}$). The first is due to the fact that the

Laplace estimator incorporates all data through integration. And we will provide Monte Carlo evidence in Section 2.5.2 regarding microstructure noise.

2.4 Feasible Estimation and Confidence Bounds

Knowing that $\bar{C}_T$ and $\bar{C}_T^{BC}$ are an approximation of integrated volatility $\langle X, X\rangle_T$ with an "error" of order of magnitude $H^{1/2}$ is good, but not sufficient for concrete applications, such as constructing a confidence interval for $\bar{C}_T$ and $\bar{C}_T^{BC}$ at a given level of significance. In this section, we focus on a precise evaluation of the asymptotic variance in order to make the previous CLTs (Theorem 2.3.2, 2.3.4 and 2.3.7) feasible.

To establish a feasible CLT (a consistent estimator of the asymptotic variance and bias) we use the blocking method as in Section 4.1 of [58].

Proposition 2.4.1. *Assume that* $1/min_i(\Delta t_i) = o_p(N^{1+\epsilon})$ *and* $\max_i(\Delta t_i) = o_p(N^{-(1-\epsilon)})$ *for some* $0 < \epsilon < 1/2$ *and* $H = cN^{-\alpha}$*. Then for any* $1 > \beta \geq (1+\epsilon) - \alpha/2$*, if we let* $M_N = [N^\beta]$*, and define* $\widehat{\sigma}_t^2$ *as*

$$\widehat{\sigma}_t^2 = \frac{\bar{C}_{t_{iM_N}} - \bar{C}_{t_{(i-1)M_N}}}{t_{iM_N} - t_{(i-1)M_N}}, \quad \text{if } t \in [t_{iM_N}, t_{(i+1)M_N}) \tag{2.4.1}$$

and $\widehat{\sigma}_t^2 = 0$ *for* $t \in [0, t_{M_N})$.[10] *Then under the Assumption H2.1 that* σ_t *is continuous, we have*

$$\widehat{\sigma}_t^2 \xrightarrow{p} \sigma_t^2. \tag{2.4.2}$$

Proposition 2.4.1 enables us to consistently estimate spot volatility σ_t. Next, we focus on estimating $\Psi(t)$, a_t and b_t. We set

$$\begin{cases} \int_0^t \widehat{\Psi(s)}\, ds & = H^{-1} \sum_{t_{i+1} \leq t} \frac{\Delta t_i^2}{2} + H^{-1} \sum_{t_{i+1}\leq t} \sum_{j \leq i-1} k\left(\frac{t_i - t_j}{H}\right)^2 \Delta t_i \Delta t_j \\ \int_0^t \widehat{a_s\sigma_s}\, ds & = H^{-1/2} \sum_{t_{j+1}\leq t} \sum_{t_{i+1}\leq t} k\left(\frac{t_i - t_j}{H}\right) \Delta X_{t_i} \Delta t_j \\ \int_0^t \widehat{b_s\sigma_s^2}\, ds & = H^{-1} \sum_{t_{j+1}\leq t} \left(\sum_{t_{i+1}\leq t} k\left(\frac{t_i-t_j}{H}\right)\Delta X_{t_i}\right)^2 \Delta t_j \end{cases}$$

[10] We intentionally shift the time when we define $\widehat{\sigma_t^2}$, so that it is adapted to the filtration $(\mathcal{F}_t)$.

Then naturally, the derivative of above equations leads to

$$\begin{cases} \widehat{\Psi}(t) & = \frac{\int_0^{t_{iM_N}} \widehat{\Psi(s)\,ds} - \int_0^{t_{(i-1)M_N}} \widehat{\Psi(s)\,ds}}{t_{iM_N} - t_{(i-1)M_N}} \\ \widehat{a}_t & = \frac{\int_0^{t_{iM_N}} \widehat{a_s\sigma_s\,ds} - \int_0^{t_{(i-1)M_N}} \widehat{a_s\sigma_s\,ds}}{(t_{iM_N} - t_{(i-1)M_N})\sqrt{\widehat{\sigma_t^2}}} \\ \widehat{b}_t & = \frac{\int_0^{t_{iM_N}} \widehat{b_s\sigma_s^2\,ds} - \int_0^{t_{(i-1)M_N}} \widehat{b_s\sigma_s^2\,ds}}{(t_{iM_N} - t_{(i-1)M_N})\widehat{\sigma_t^2}} \end{cases} \tag{2.4.3}$$

Proposition 2.4.2. *Under the Assumptions of Proposition 2.4.1 and Theorem 2.3.4, then*

1. $\sum_i \widehat{\Psi}(t_{(i-1)M_N})\widehat{\sigma}^4_{t_{(i-1)M_N}}(t_{iM_N} - t_{(i-1)M_N}) \xrightarrow{p} \int_0^T \Psi(t)\sigma_t^4\,dt,$
2. $\sum_i \left(\widehat{b}_{t_{(i-1)M_N}} - \widehat{\Psi}(t_{(i-1)M_N}) - \widehat{a}^2_{t_{(i-1)M_N}}\right)\widehat{\sigma}^4_{t_{(i-1)M_N}}(t_{iM_N} - t_{(i-1)M_N})$
$\xrightarrow{p} \int_0^T (b_t - \Psi(t) - a_t^2)\sigma_t^4\,dt,$
3. *the stochastic integral* $(\hat{a}\cdot\hat{\sigma})\cdot X = \int_0^T \hat{a}_t\hat{\sigma}_t\,dX_t \xrightarrow{p} \int_0^T a_t\sigma_t\,dX_t.$

Combining Proposition 2.4.2 with the stable convergence in law results in Theorem 2.3.2 and 2.3.4, we obtain a corollary.

Corollary 2.4.3. *Under the Assumptions of Proposition 2.4.2, we have the following convergence in law results,*

1. *Under the Assumptions of Theorem 2.3.2, we have*

$$H^{-1/2}\frac{\bar{C}_T - \langle X, X\rangle_T}{4\sum_i \widehat{\Psi}(t_{(i-1)M_n})(\widehat{\sigma}_{t_{(i-1)M_n}})^4(t_{iM_n} - t_{(i-1)M_n})} \xrightarrow{\mathcal{L}} N(0,1). \tag{2.4.4}$$

2. *Under the Assumptions of Theorem 2.3.4, we have*

$$H^{-1/2}\frac{(\bar{C}_T - \langle X, X\rangle_T) - 2(\hat{a}\cdot\hat{\sigma})\cdot X}{4\left(\widehat{b}_{t_{(i-1)M_n}} - \widehat{\Psi}(t_{(i-1)M_n}) - \widehat{a}^2_{t_{(i-1)M_n}}\right)(\widehat{\sigma}_{t_{(i-1)M_n}})^4(t_{iM_n} - t_{(i-1)M_n})} \xrightarrow{\mathcal{L}} N(0,1). \tag{2.4.5}$$

Corollary 2.4.3 enables us to compute p-value as well as constructing confidence intervals for the Laplace based estimator of integrated volatility $\bar{C}_T$ as well as the bias-corrected Laplace estimator $\bar{C}_T^{BC}$, which will be used in the Monte Carlo study in Sections 2.5.3 and 2.5.4.

2.5 Simulation Study

In this section, we study the finite sample properties of the Laplace estimator of integrated volatility. The data generating process is describe as

$$\begin{cases} dX_t = \sigma_t dW_t, \\ d\sigma_t^2 = \gamma(\theta - \sigma_t^2)dt + \alpha\sigma_t dZ_t, \end{cases} \tag{2.5.1}$$

where $\langle dW, dZ\rangle_t = \lambda dt$ and W_t and Z_t are standard Brownian motions. We follow the choice of [11] and [30] and set the parameters as $\gamma = 0.01$. $\theta = 1$, $\alpha = 0.05$, $\sigma_0^2 = 1$, and $X_0 = log(100)$. Second-by-second return and volatility paths over a daily trading period of 6 hour for a total of 1000 trading days are computed using the Euler discretization scheme.

Throughout Section 2.5 we consider the performance of the naive Laplace estimator defined in Eq. (2.2.6) and its bias-corrected version. In Section 2.5.1, we study the proper choice of R and compare the efficiency of the Laplace estimator with realized volatility. We also examine the performance of the Laplace estimator when leverage effect is present. In Section 2.5.2, we consider the case where the efficient price process is contaminated by microstructure noise and conclude that the Laplace estimator is robust to noise and vastly superior to the naive realized volatility. In Sections 2.5.3 and 2.5.4, we construct confidence intervals based on Corollary 2.4.3 when sampling times are possibly endogenous. In Section 2.5.5, we consider the case where the efficient price process is contaminated by microstructure noise under the case of endogenous sampling, and conclude that the Laplace estimator is robust to the microstructure noise and the time endogeneity at the same time.

2.5.1 Choice of c and leverage effect

In this section, we first investigate the choice of c regarding Theorem 2.3.7 and we consider the case that there is no leverage effect, i.e., $\lambda = 0$. To attain the smallest asymptotic variance for the estimator, we need to set $c = (\pi/T)k$, $k = 1, 2, \ldots$. It is also worth noting that in related literature regarding the Fourier estimator, the choice $k = 1$ corresponds to the natural choice of the Nyquist frequency for the Fourier estimator.[11] This suggest a natural choice of $c = \pi/T$ in the case of finite horizon.

The simulation results shown in Table 2.1 confirm previous analysis. According to Theorem 2.3.7, the exact MSE of the Laplace estimator is going to be 2 and 4 times larger than that of realized volatility, when $k = 1/2$ or $1/4$ (corresponds to $\tilde{c} = 1/2$ and $1/4$). In case of 1 minute

[11] In order to recover all Fourier components of a periodic waveform, it is necessary to use a sampling rate at least twice the highest waveform frequency. The Nyquist frequency, also called the Nyquist limit, is the highest frequency that can be coded at a given sampling rate in order to be able to fully reconstruct the signal.

Table 2.1 Laplace estimator with different choices of $\tilde{c}$.

sampling rate	1 min			5 min		
$\tilde{c}$	MSE	RMSE	Bias	MSE	RMSE	Bias
1	0.0003	1.0009	0.0006	0.0017	1.0126	-0.0015
4	0.0003	0.9986	0.0005	0.0017	1.0025	-0.0012
1/2	0.0007	1.9626	0.0015	0.0033	1.9692	0.0007
1/4	0.0013	3.8137	0.0005	0.0067	4.0434	0.0025

The mean square error of an estimator $\hat{IV}$ is computed as $MSE(\hat{IV}) = E(\hat{IV} - IV)^2$, the bias is computed as $Bias(\hat{IV}) = E(\hat{IV} - IV)$. RMSE here denote the relative MSE of the Laplace estimator with respect to the MSE of the realized volatility.

Table 2.2 Laplace estimator with leverage effect.

sampling rate	1 min		5 min	
λ	MSE	RMSE	MSE	RMSE
-1	0.0004	1.0006	0.0017	1.0034
- 1/2	0.0003	1.0009	0.0017	1.0126
0	0.0003	1.0009	0.0017	1.0126
1/2	0.0003	1.0009	0.0017	1.0127
1	0.0004	1.0006	0.0017	1.0027

sampling rate, RMSE is 1.9626 and 3.8137. When sampling rate became even lower to 5 min, RMSE is 1.9692 and 4.0434. Similarly, Theorem 2.3.7 predicts that the MSE of the Laplace estimator should equal to the MSE of realized volatility when $k = 1$ or 4 (corresponds to $\tilde{c} = 1$ and 4), as can see from Table 2.1, the RMSE is very close to 1 in both cases of sampling rate.

The second experiment studies the Laplace estimator's behavior with respect to the leverage effect λ,[12] for this study we set $c = \pi/T$.

Table 2.2 shows the results of this experiment in the same format as Table 2.1. Even the leverage effect parameter $\lambda \neq 0$, the MSE of the Laplace estimator does not deteriorate comparing to realized volatility, RMSE in all cases are sufficiently close to 1, indicating that the Laplace estimator could be adopted in a model with leverage effect.

2.5.2 Microstructure noise

In the previous section it is assumed that the observed prices are the true prices and are observed without error. However, a distinct characteristic of high-frequency financial data is that they are observed with noise (a measurement error) due to the imperfections of the trading process (e.g.,

[12]The terminology originates in the fact that an asset's volatility tends to be negatively correlated with the asset's returns.

Table 2.3 MSE of the Laplace estimator with microstructure noise.

	$\zeta = 1/8$	$\zeta = 1/4$	$\zeta = 1/3$	$\zeta = 1/2$	$\zeta = 2/3$	$\zeta = 1$
$\xi = 0$	0.0370	0.0211	0.0143	0.0062	0.0023	**0.0003**
$\xi = 0.4$	0.0369	0.0211	0.0144	0.0063	**0.0024**	0.0074
$\xi = 0.8$	0.0370	0.0212	0.0145	0.0065	**0.0026**	0.1079
$\xi = 1.6$	0.0373	0.0215	0.0149	0.0071	**0.0053**	1.6951
$\xi = 3.2$	0.0393	0.0233	0.0167	**0.0107**	0.0414	27.0083

see [22]). Market microstructure noise summarizes the discrepancy between the efficient log-price process and the observed log-price process, as generated by the mechanics of the trading process. Such as the presence of a bid-ask spread, differences in trade sizes, informational asymmetries of traders, discreteness of price changes in markets that are subject to a tick size.

In this section, we assume the efficient log-price follows Eq. (2.5.1), and the observed price $\widetilde{X}$ process is described as

$$\widetilde{X}_t = X_t + \epsilon_t, \tag{2.5.2}$$

where ϵ is the microstructure noise. We consider the simplest case such that ϵ_t's are assumed to be i.i.d. following a Gaussian distribution $N(0, \eta^2)$. We also assume that the efficient price process X is independent of the microstructure noise. We follow [30] and [17] parametrize $\eta = \xi std(r)$, where $std(r)$ is the standard deviation of the 1-min returns. The quantity ξ represents noise-to-signal ratio. In this study, we focus on the performance of the 1-min Laplace estimator, and we parameterized $R = (\pi/T)n^{\zeta}$,[13] we also set $\lambda = 0$ so that there is no leverage effect.

Table 2.3 shows the performance of the Laplace estimator when used to estimate the volatility with the high frequency data possibly contaminated by the microstructure noise in terms of MSE. The value in bold represents the smallest value for each row. The first line of Table 2.3 shows the case that the noise-to-signal ratio is 0 (corresponds to the case that there is no microstructure noise), the optimal value of $\zeta = 1$. In the presence of noise, the optimal value of ζ goes from 1 to 2/3. This is not surprising due to the similarity between the Laplace estimator and the Fourier estimator. As in the discussion for the Fourier estimator in [61], cutting the highest frequencies in the Fourier estimator permits the high frequency noise components to be filtered out.

As a reference, Table 2.4 provides the performance of realized volatility using sparse sampling. As one would expect, the Laplace estimator (when choosing $\zeta = 2/3$) is vastly superior comparing to the naive RV estimator when noise presents.

[13] This parameterization is inspired by the Fourier based estimator, see [61]. Whenever $\zeta < 1$, it corresponds to the special case of $c = 0$.

Table 2.4 MSE of realized volatility with microstructure noise.

	1 min RV	5 min RV	10 min RV
$\xi = 0$	0.0003	0.0017	0.0034
$\xi = 0.4$	0.0074	0.0021	0.0037
$\xi = 0.8$	0.1073	0.0066	0.0055
$\xi = 1.6$	1.6856	0.0729	0.0251
$\xi = 3.2$	26.8589	1.1082	0.3057

Table 2.5 Summary statistics of confidence intervals based on the simulated data.

	Average Width	Coverage Frequency
RV	0.0211	94.30%
5-sec Laplace estimator	0.0211	94.40%
Bias-corrected Laplace estimator	0.0211	93.50%

2.5.3 Confidence intervals in the equidistant case

In this section, we focus on the performance of the bias-corrected Laplace estimator $\hat{C}_T^{BC}$ in Remark 2.3.5. We consider the data generating process as Eq. (2.5.1), and we assume there is no leverage effect ($\lambda = 0$) and no microstructure noise. We set $c = \pi/T$ which is a natural choice in light of Section 2.5.1 and 2.5.2. The aim is to show that under the case of no time endogeneity, the bias-corrected Laplace estimator is as efficient as the Laplace estimator and the realized volatility.

To have a feasible CLT we use the blocking method as in Section 2.4. Specially, we choose $\beta = 3/4$ and $M_N = [N^\beta]$ and consistent estimators can be obtained as in Proposition 2.4.1 and 2.4.2, confidence intervals are constructed using Corollary 2.4.3. In this section, we consider the performance of the 5-sec Laplace estimator,[14] so that $N = 4320$, and the size of each block to be $M_N = 540$.

In Figure 2.1, we plot the 95% confidence interval using three different methods. The orange lines represent confidence intervals for the realized volatility constructed using the CLT based on quarticity; Blue lines represent confidence intervals constructed using the 5-sec Laplace estimator $\hat{C}_T$ as in Theorem 2.3.2; Red lines represent confidence intervals for the bias-corrected Laplace estimator $\hat{C}_T^{BC}$ as in Remark 2.3.5. From Figure 2.1, in case of non-endogenous data, the bias-corrected Laplace estimator $\hat{C}_T^{BC}$ performs just as well as the realized volatility and $\hat{C}_T$.

Table 2.5 reports summary statistics calculated from the whole sample path (1000 days) and it confirms the eyeball result from Figure 2.1, as the coverage frequency of all three candidates

[14] We choose such a high frequency in order to obtain accurate spot estimates.

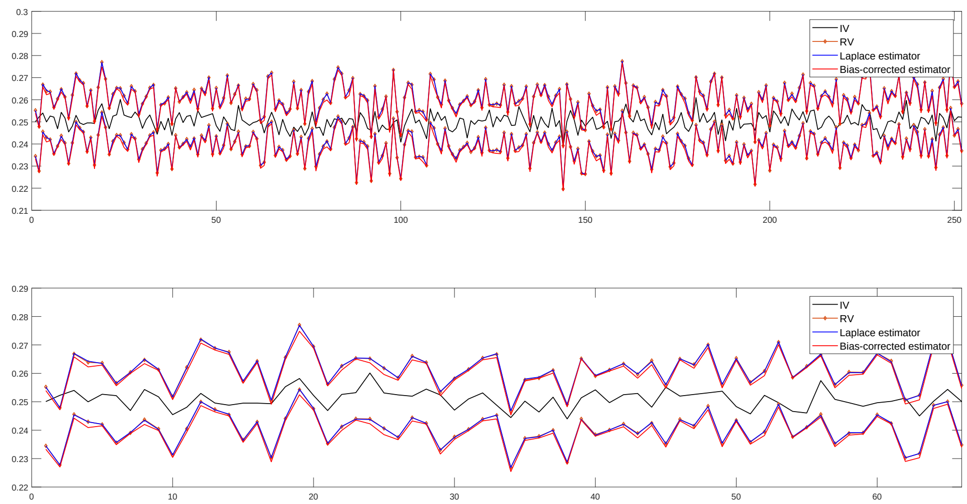

Figure 2.1 Confidence intervals computed base on three methods: realized volatility, Theorem 2.3.2 and Remark 2.3.5 for non-endogenous data sampling every 5 seconds (Black: Integrated volatility; Orange: Realized volatility; Blue: Laplace estimator using Theorem 2.3.2; Red: Bias-corrected Laplace estimator using Remark 2.3.5). Upper panel: for 252 days; Lower panel: for 66 days.

are close to the targeted 95%, and have the same average width.

2.5.4 Confidence intervals in the endogenous case

In this section, we study the performance of the bias-corrected Laplace estimator $\hat{C}_T^{BC}$ in Remark 2.3.5 when sampling times are endogenous. For this study, the data generating process follows Eq. (2.5.1), but we assume the following sampling scheme:[15]

The times $t_{n,i}$ are defined recursively: $t_{n,0} = 0$, and $t_{n,i+1}$ is the first time $t \geq t_{n,i}$ so that $X_t - X_{t_{n,i}} =$ either $n^{-1/2}a$ or $-n^{-1/2}b$, where a, $b > 0$. Let N be such that $t_{n,N} \leq 0.25 < t_{n,N+1}$, where 0.25 represent 6/24 of a day.

Under this sampling scheme, we have

$$X_{t_{n,i+1}} - X_{t_{n,i}} = n^{-1/2} Z_i,$$

where Z_i's are i.i.d. with mean zero (by X_t being a standard Itô process) and $P(Z = a) = b/(a+b)$, $P(Z = -b) = a/(a+b)$. For this study, we set $n = 3600$, with $a = 0.4$, $b = 0.2$, i.e., we sample each time when there is an increase of 0.67% or a decrease of 0.33% in the price.

The simulation here focus on two points. First, the performance of the Laplace estimator in

[15]This sampling scheme is also used in [58], see example 4 in Section 3.

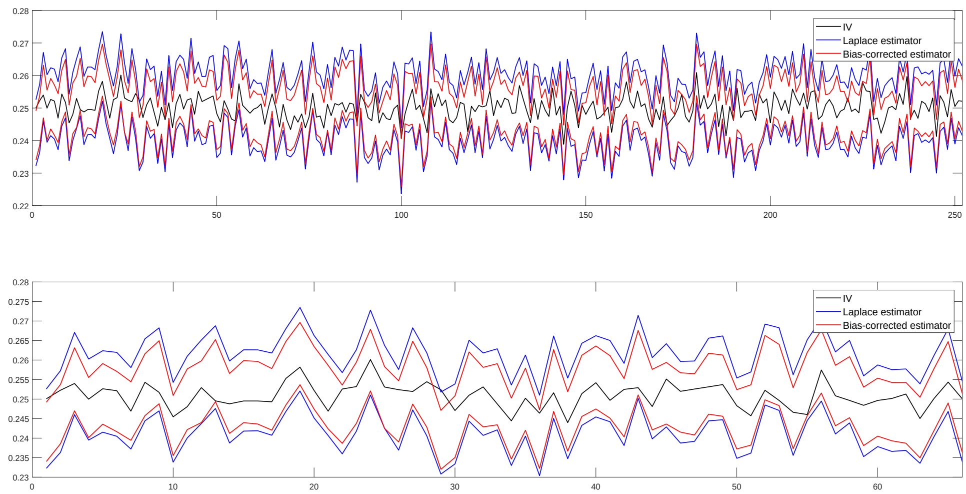

Figure 2.2 Confidence intervals computed base on two methods: Theorem 2.3.2 and Remark 2.3.5 for endogenous data (Black: Integrated volatility; Blue: Laplace estimator using Theorem 2.3.2; Red: Bias-corrected Laplace estimator using Remark 2.3.5). Upper panel: for 252 days; Lower panel: for 66 days.

Table 2.6 Summary statistics of confidence intervals based on the simulated data.

	Average Width	% Width Reduced	Coverage Frequency
Laplace estimator	0.0209	–	98.80%
Bias-corrected Laplace estimator	0.0156	25.36%	94.50%

Theorem 2.3.2 and Remark 2.3.5 when sampling times are endogenous. Second, the variance reduction of the bias-corrected Laplace estimator $\hat{C}_T^{BC}$, as in Theorem 2.3.4, after correcting the bias, the asymptotic variance becomes $4\int_0^T \left(b_t - \Psi(t) - a_t^2\right) \sigma_t^4 \, dt$, which is smaller than the asymptotic variance of $\hat{C}_T$ in Theorem 2.3.2, which is equal to $4\int_0^T \left(b_t - \Psi(t)\right) \sigma_t^4 \, dt$. The total reduction of the variance is hence equal to $4\int_0^T a_t^2 \sigma_t^4 \, dt$. In the non-endogenous case, since there is no correlation between X and t, we have $a_t = 0$, variance reduction hence is 0, as in Table 2.5. If we have a non-zero value of a_t, as is in the most cases of endogenous sampling schemes, this additional information allows us an efficiency gain with respect to the naive estimator.

Figure 2.2 shows the 95% confidence intervals using two different methods. Blue lines represent confidence intervals constructed using the Laplace estimator $\hat{C}_T$ as in Theorem 2.3.2; Red lines represent confidence intervals for the bias-corrected Laplace estimator $\hat{C}_T^{BC}$ as in Remark 2.3.5.

Table 2.6 reports summary statistics calculated from the whole sample path (1000 days). From Table 2.6, the width of the confidence intervals constructed using Remark 2.3.5 are much

Table 2.7 MSE of the estimators with microstructure noise under endogenous sampling.

	$\xi = 0.4$	$\xi = 0.8$	$\xi = 1.6$	$\xi = 3.2$
$\hat{C}_T$	0.0003	0.0003	0.0004	0.0017
$\hat{C}_T^{BC}$	0.0004	0.0004	0.0003	0.0010
RV	0.0064	0.1029	1.6395	26.2893

narrower than the confidence intervals constructed using Theorem 2.3.2. On the other hand, the coverage frequency of $\hat{C}_T$ is higher than $\hat{C}_T^{BC}$ due to the wider confidence intervals despite the bias.[16] The coverage frequency of the bias-corrected Laplace estimator $\hat{C}_T^{BC}$ is very close to expected 95%, and features a much narrower confidence interval.

2.5.5 Microstructure noise with time endogeneity

In this section, we assume the efficient log-price follows Eq. (2.5.1), and the microstructure noise component follows from Section 2.5.2 under the endogenous sampling scheme as in Section 2.5.4.[17] In light of Section 2.5.2, we parameterized $R = (\pi/T)N^{\zeta}$, with $\zeta = 2/3$ to cut the highest frequencies in the Laplace estimator and the bias-corrected Laplace estimator to filter out the high frequency noise components. We also report the MSE of the realized volatility (RV) for comparison.

Table 2.7 shows the performance of the Laplace estimator, the bias-corrected Laplace estimator and the realized volatility (RV) when used to estimate the volatility with the high frequency data affected by noise under endogenous sampling scheme in terms of MSE. The performance of the Laplace estimator and the bias-corrected Laplace estimator are robust to different values of the noise-to-signal ratio ξ, whereas the performance of the realized volatility (RV) could be heavily affected. We conclude that the Laplace estimator and the bias-corrected Laplace estimator are robust to the presence of the microstructure noise and the time endogeneity.

2.6 Empirical Application: Modeling and Forecasting Equity Return Volatility

In this section, we apply the proposed Laplace estimator of integrated volatility ($\hat{C}_T$ and $\hat{C}_T^{BC}$) to the millisecond time stamped trade data and compare its performance of forecasting equity return volatility with other commonly used estimators, e.g., realized volatility (Eq. (2.6.1)),

[16]Notice that the bias is an asymptotic bias, so both estimators are consistent.

[17]Here we parametrize $\eta = \xi std(r')$, where $std(r')$ is the standard deviation of the returns under endogenous sampling scheme.

bi-power variation (Eq. (2.6.2)), subsampled version of RV (Eq. (2.6.7)) and two-scale RV (Eq. (2.6.8)).

Volatility forecasting is crucial in risk management and asset pricing. And having access to accurate volatility forecasts is of the utmost importance for participants and regulators in financial market. One of the characteristic of financial data is that the autocorrelations of the volatility show strong persistence that last for long time periods, which motivates a long list of competing GARCH and fractional integration models.

We analyze a large liquid stock, International Business Machines Corporation (ticker: IBM). The sample period is from January 3, 2011 to December 30, 2011 (252 trading days). Following the data cleaning procedure in [17], we first remove all trading records outside the 9:30 AM – 4:00 PM EST window when the exchange is open. Second, we delete entries with corrected trades, whose correction indicator is different from zero, as well as those with an abnormal sale condition (where COND has a letter code other than 'E' and 'F'). Third, if multiple transactions have the same time stamp, we only keep one trade record with the volume-weighted price.

After the standard data cleaning procedure, we then proceed and calculate the 5-min realized volatility (RV), bi-power variation (BPV) and the Laplace estimator ($\hat{C}_T$) as,[18]

$$RV = \sum_{t_{i+1} \leq T} \Delta X_{t_i}^2, \tag{2.6.1}$$

$$BPV = \frac{\pi}{2} \sum_{t_{i+1} \leq T} |\Delta X_{t_i}||\Delta X_{t_{i+1}}|, \tag{2.6.2}$$

$$\hat{C}_T = \sum_{t_{i+1} \leq T} \sum_{t_{j+1} \leq T} \frac{sin(R(t_i - t_j))}{R(t_i - t_j)} \Delta X_{t_i} \Delta X_{t_j}. \tag{2.6.3}$$

In the following, irrespective of their actual frequencies, all return and volatility quantities are intended to be annualized to facilitate comparison among different frequencies. Table 2.8 provides a standard set of summary statistics for the different estimators considered in this section.[19]

2.6.1 Motivations and modeling

A long list of variations of GARCH and stochastic volatility type models have been proposed for estimating as well as forecasting financial market volatility. However, due to the latent nature of volatility, most specific structure models proposed in the literature do not perform well when estimated directly with the intraday high frequency data. Standard GARCH and stochastic volatility models are not able to reproduce all the features presented in financial data (e.g.,

[18]For 5-min $\hat{C}_T$, we set $c = \pi/T$ which is a natural choice in light of the study in Section 2.5.1.
[19]See Section 2.6.4 for a description of SSRV, TSRV, $\hat{C}_{robust}$, $\hat{C}_T^E$ and $\hat{C}_T^{BC}$.

Table 2.8 Summary statistics.

	Min	Mean	Median	Max	AR
RV	1.4232	11.4338	7.3654	110.2034	0.7631
BPV	1.2725	10.7529	6.6831	105.6589	0.7713
$\hat{C}_T$	1.4110	11.4792	7.2767	112.1905	0.7500
SSRV	1.4768	11.7210	7.4391	137.8429	0.7387
TSRV	1.0805	9.1153	5.7454	107.0082	0.7221
$\hat{C}_{robust}$	1.4524	11.8354	7.4120	114.7184	0.7550
$\hat{C}_T^E$	1.3637	12.2925	7.6062	137.4267	0.7761
$\hat{C}_T^{BC}$	1.1411	10.9247	6.6975	126.9677	0.7922

The table reports summary statistics for the different volatility quantities considered in Section 2.6. The column labeled AR reports the standard first order autocorrelation coefficients.

autocorrelations of volatility, return distributions).

[6] suggested the use of reduced-form time series forecasting models for the realized volatility. Later, the simple Heterogeneous AR (HAR) model proposed by [28] has become the preferred model for forecasting. Empirically, the volatility forecasts from the HAR model or other similar reduced-form time series models generally provide better forecasts than the traditional parametric stochastic volatility and GARCH models.

HAR model focus on the heterogeneity that arises from the difference in the time horizons. A financial market is composed of participants with different trading frequencies. Dealers, market makers and speculators have a very high intraday trading frequency. Whereas institutional investors trade much less frequently. This leads to the idea that agents with different time horizons lead to different types of volatility components: the short-term traders with daily or higher trading frequency, the medium-term investors who re-balance weekly and the long-term agents with a characteristic time of one or more months. Thus one can obtain a very simple time series representation of the cascade model (HAR),

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t, \tag{2.6.4}$$

where $RV_{t-j|t-h} = \frac{1}{h+1-j}\sum_{i=j}^{h} RV_{t-i}$, with $j \leq h$. This choice of a daily, weekly and monthly lag conveniently captures the long-memory dynamic dependencies.

Many different extensions of the standard HAR model have been proposed, we therefore also consider the forecasts from the HAR-with-Jumps (HAR-J) in our forecast comparisons (See [7]). HAR-J model relies on a decomposition of the total variation into a continuous and a discontinuous (jump) part. One of the most commonly used jump robust volatility measure is bi-power variation (BPV) of [15], which consistently estimates the continuous variation in the presence of jumps. The HAR-J model includes a measure of the jump variation as an extra explanatory variable on the basis of the standard HAR model,

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + \beta_J J_{t-1} + u_t, \tag{2.6.5}$$

where $J_t = \max(RV_t - BPV_t, 0)$, and BPV measure is defined as in Eq. (2.6.2).

[24] argues that realized volatility are subject to measurement error. As such, realized volatility will be equal to the sum of the true integrated volatility (IV) and a measurement error. Therefore, the dynamic modeling of realized volatility of forecasting the true latent IV suffers from an errors-in-variables problem and will likely lead to attenuation bias. They propose a simple approach for improving volatility forecasting by exploiting the heteroskedasticity in the errors. Intuitively, on days when the variance of the measurement error is small, the daily RV provides a stronger signal for next day's volatility than on days when the variance is large. They adapted this idea to the popular HAR model, which they dub the HARQ model,

$$RV_t = \beta_0 + (\beta_1 + \beta_{1Q} RQ_{t-1}^{1/2}) RV_{t-1} + \beta_2 RV_{t-1|t-5} + \beta_3 RV_{t-1|t-22} + u_t, \tag{2.6.6}$$

where $RQ = \frac{N}{3} \sum_{t_{n,i+1} \leq T} \Delta X_{t_{n,i}}^4$ and is a consistent estimator of integrated quarticity (IQ).[20] For ease of interpretation, we demean $RQ^{1/2}$ so that the estimate of β_1 corresponds to the average autoregressive coefficient. Following the previous idea that uninformative days with large measurement errors will have smaller impact on the forecasts than days where RV is estimated precisely, which suggests $\beta_{1Q} < 0$.

Throughout Section 2.6, we will compare the performance of the Laplace estimator based on the above three models: HAR model (Eq. (2.6.4)), HAR-J model (Eq. (2.6.5)) and HARQ model (Eq. (2.6.6)). To allow for a direct comparison across the different volatility measures and models, we always take daily 5-minute RV as the forecast target. That is, for the study in Section 2.6.2, 2.6.3 and 2.6.4, we always use 5-minute RV as the dependent variable on the left-hand-side of the regression, we use different volatility measures as explanatory variables.

2.6.2 In-sample estimation results

We begin by considering the full in-sample results. All of the models are estimated by OLS based on 252 daily observations. Standard OLS regression estimators are consistent and normally distributed. In order to account for the possible presence of heteroskedasticity and serial correlation in the data, the [65] covariance correction is used. Following [24], we rely on a

[20] Within each trading day, $IQ_t = \int_0^t \sigma_u^4 \, du$, and it is on the same scale as the asymptotic variance of the estimator (same scale as the measurement error), see [14].

Table 2.9 In-sample estimation results.

	RV			BPV		$\hat{C}_T$		
	HAR	HARQ	HAR-J	HAR	HARQ	HAR	HARQ	HAR-J
β_0	2.1013	1.2437	2.1233	2.2323	1.6143	2.1609	1.3049	2.2104
	(0.7106)	(0.6747)	(0.7247)	(0.7486)	(0.7401)	(0.7231)	(0.6802)	(0.7258)
β_1	0.7584	0.9697	0.7755	0.8152	0.9800	0.7298	0.9405	0.8142
	(0.0863)	(0.2586)	(0.1369)	(0.1022)	(0.2825)	(0.0777)	(0.2536)	(0.1510)
β_2	-0.0638	-0.0812	-0.0717	-0.1025	-0.1207	-0.0451	-0.0620	-0.0839
	(0.0980)	(0.1033)	(0.1079)	(0.1200)	(0.1312)	(0.0991)	(0.1018)	(0.1120)
β_3	0.1345	0.0581	0.1367	0.1577	0.1014	0.1362	0.0599	0.1405
	(0.1158)	(0.1607)	(0.1187)	(0.1304)	(0.1697)	(0.1159)	(0.1625)	(0.1183)
β_{1Q}	–	-0.0198	–	–	-0.0155	–	-0.0197	–
	–	(0.0149)	–	–	(0.0157)	–	(0.0151)	–
β_J	–	–	-0.1713	–	–	–	–	-0.6552
	–	–	(1.3104)	–	–	–	–	(0.9982)
R^2	0.5788	0.5879	0.5791	0.5667	0.5718	0.5730	0.5822	0.5770
MSE	85.6973	83.8518	85.6518	88.1624	87.1299	86.8875	85.0157	86.0653
QLIKE	0.1354	0.1387	0.1357	0.1396	0.1427	0.1368	0.1394	0.1377

The table provides in-sample parameter estimates and measures of fit for the various specifications. The table reports the actual parameter estimates for the IBM with robust standard errors in parentheses, together with the R^2s, MSE and QLIKE losses from the regressions.

standard MSE measure to access the in-sample fits.

$$MSE(RV_t, X_t) = (RV_t - X_t)^2,$$

where X_t refers to the fit from any one of the different models or explanatory variables. We also consider the QLIKE loss,

$$QLIKE(RV_t, F_t) = \frac{RV_t}{F_t} - log(\frac{RV_t}{F_t}) - 1,$$

where F_t refers to the one-day-ahead forecasts from the different models or volatility measures.

The top panel in Table 2.9 reports the parameter estimates for IBM, with robust standard errors in parentheses, for combinations of three different volatility measures (RV, BPV and $\hat{C}_T$) and three different model specifications (HAR, HARQ and HAR-J).[21] The second panel report the R^2, MSE and QLIKE for the IBM.

As expected, in all cases, all of the β_{1Q} are negative, which is consistent with the simple

[21]Notice that in the case of HAR-J model with BPV as explanatory variable, the jump component is 0. And it is equivalent to HAR model with BPV as explanatory variable. For this reason, there is a total of 8 different specifications.

Table 2.10 Out-of-sample forecast losses.

		RV			BPV		$\hat{C}_T$		
		HAR	HARQ	HAR-J	HAR	HARQ	HAR	HARQ	HAR-J
MSE	RW	1.0000	0.9777	1.0596	1.0046	1.0527	1.0028	1.0330	1.0109
	IW	1.0000	1.0262	1.0454	1.0194	0.9926	1.0170	1.0276	0.9022
QLIKE	RW	1.0000	1.0369	0.9930	1.0005	1.0841	1.0223	1.0363	1.0076
	IW	1.0000	0.8751	1.0384	1.0716	1.1434	1.0919	0.8980	0.9347

The table reports the ratio of the losses for the different specifications relative to the losses of the HAR model with RV as explanatory variable.

intuition of exploiting the errors. Another noticeable difference is that all HARQ models lead to larger estimate of β_1, which confirm the existence of attenuation bias. By failing to take into account the time-varying nature of the informativeness of the volatility measures, the estimated AR coefficients are sharply attenuated. There is no clear difference between the R^2, MSE and QLIKE for the same model specifications but across different volatility measures. On the other hand, given a specific volatility measure, in terms of R^2 and MSE, HARQ model shows better in-sample fit results.

2.6.3 Out-of-sample forecast results

In this section, we focus our discussion on the one-day-ahead forecasts for the IBM. The forecasts are based on re-estimating the parameters of the different specifications each day with a fixed length rolling window (RW) comprised of the previous 100 days, as well as increasing window (IW) using all of the up-to-date available observations. The sample sizes for the increasing window for the IBM thus range from 100 to 252 days. We follow the procedure in [24] and apply an "insanity filter" to the forecasts, see [74].[22]

The average MSE and QLIKE for IBM are reported in Table 2.10. The results across different models are somewhat mixed. For RV, HAR and HARQ provide better forecasts comparing to HAR-J model; for BPV, HAR provide better forecasts comparing to HARQ model; whereas for $\hat{C}_T$, HAR and HAR-J model outshine the HARQ model in most cases. When comparing the performance across different volatility measures, RV and $\hat{C}_T$ provide the most accurate forecasts and in most cases perform better than BPV. We thus conclude that for IBM, the forecasts based on $\hat{C}_T$ are as good as the forecasts based on RV, and are better than the forecasts based on BPV.

[22]The integrated quarticity is notoriously difficult to estimate in the literature. Due to the estimation errors in RQ, the HARQ models may on occasions produce implausibly large or small forecasts. We apply the following filter to the forecasts. If a forecast is outside the range of values of the target variable observed in the estimation period, the forecast is replaced by the unconditional mean over that period. This same filter is applied to all of the specifications.

Table 2.11 Out-of-sample forecast losses based on noise-robust estimators.

		SSRV			TSRV			$\hat{C}_{robust}$		
		HAR	HARQ	HAR-J	HAR	HARQ	HAR-J	HAR	HARQ	HAR-J
MSE	RW	0.9573	1.0543	1.0176	0.7021	1.1146	0.9961	0.8394	1.0477	1.1420
	IW	0.9802	1.0935	0.9263	1.0138	0.9630	1.0877	0.9631	1.0860	0.9856
QLIKE	RW	0.9682	1.1488	0.9923	0.7504	1.0831	1.0170	0.7009	1.1199	1.0177
	IW	1.0056	1.0069	1.1026	1.0225	1.0334	1.1207	1.0062	1.0235	1.0045

The table reports the ratio of the losses for the different specifications relative to the losses of the HAR model with RV as explanatory variable.

2.6.4 Noise components and endogenous sampling

The most commonly used 5-minute RV, BPV and $\hat{C}_T$ estimator provides a simple way of mitigating the market microstructure noise arising from high frequency data (sparse sampling). A large number of alternative estimators that allow for the use of ultra high sampling frequencies have been proposed in the literature. In this section we consider some of the most commonly used noise robust volatility measures: subsampled version of RV and two-scale RV, and compare the forecasts based on the Laplace estimator. Our implementation of all the noise robust volatility measures will be based on 1-minute returns.

The subsampled version of RV (SSRV) was first introduced by [77]. Specifically, by computing the 5-minute RV on time grids with 5-minute intervals starting at 9:30 (RV_1), 9:31 (RV_2), 9:32 (RV_3), 9:33 (RV_4) and 9:34 (RV_5). The SSRV estimator is obtained as the average of these five different RV estimators,

$$SSRV = \frac{1}{5}(RV_1 + RV_2 + RV_3 + RV_4 + RV_5). \tag{2.6.7}$$

The two-scale RV (TSRV) in [77] bias-corrects the SSRV estimator,

$$TSRV = SSRV - \frac{1}{5}RV_{\text{1-min}}. \tag{2.6.8}$$

The Laplace estimator will be based on Eq. (2.6.3), but in light of Section 2.5.2, we parameterized $R = (\pi/T)n^{\zeta}$, with $\zeta = 2/3$ to cut the highest frequencies in the Laplace estimator to filter out the high frequency noise components. We denote the noise robust Laplace estimator as $\hat{C}_{robust}$.

Table 2.11 compares the out-of-sample performance of different specifications with the standard HAR based on 5-min RV for IBM. It is important to keep in mind that all of the 1-min noise robust volatility measures are still subject to measurement errors. In terms of HAR model, all three noise robust volatility measures lead to huge improvement comparing to the 5-min RV forecasts. In terms of HARQ and HAR-J models, the results are somewhat mixed and

Table 2.12 Out-of-sample forecast losses based on the Laplace estimators.

		$\hat{C}_T^E$			$\hat{C}_T^{BC}$		
		HAR	HARQ	HAR-J	HAR	HARQ	HAR-J
MSE	RW	0.9846	0.9278	0.9812	0.9412	0.9230	0.9623
	IW	0.8542	0.9074	0.8438	0.8197	0.8563	0.8286
QLIKE	RW	0.9186	0.9827	1.0046	0.9099	0.8787	1.0008
	IW	0.9876	0.9051	1.0764	1.0089	0.8522	1.0376

The table reports the ratio of the losses for the different specifications relative to the losses of the HAR model with RV as explanatory variable.

no uniform conclusion could be made. On the other hand, among three noise robust volatility measures, $\hat{C}_{robust}$ generally outperform the SSRV, and has the same level of performance as TSRV.

We also consider a different sampling scheme other than evenly spaced time grids. For the next study, we sampling every fifty trades, and this sampling scheme could possibly lead to time endogeneity.[23] For this reason, we consider the naive Laplace estimator $\hat{C}_T$ (here denoted by $\hat{C}_T^E$ which should not be confused with the 5-min $\hat{C}_T$ in Sections 2.6.2 and 2.6.3), which is still consistent based on Theorem 2.3.2 and the bias-corrected Laplace estimator $\hat{C}_T^{BC}$ based on Remark 2.3.5. After sampling every fifty trades, for each trading day, we have 685 observations on average, corresponds to an average sampling rate of 30 seconds. In light of Section 2.5.2, we parameterized $R = (\pi/T)n^{\zeta}$, with $\zeta = 2/3$ to filter out the high frequency noise signals.

Table 2.12 compares the out-of-sample performance of different specifications with the standard HAR based on 5-min RV for IBM. Both $\hat{C}_T^E$ and $\hat{C}_T^{BC}$ offer huge improvement across all models. Overall $\hat{C}_T^{BC}$ provide better forecasts than $\hat{C}_T^E$ due to the sampling scheme here could possibly leads to time endogeneity. The bias-corrected Laplace estimator $\hat{C}_T^{BC}$ takes advantage of this additional information and allows an efficiency gain and thus provide better forecasts.

2.7 Conclusions

This paper studies a class of volatility measures based on the Laplace transform, which is robust to the presence of microstructure noise and the time endogeneity. Asymptotic properties and feasible central limit theorems are established. Given the presence of time endogeneity, our bias-corrected version Laplace estimator allows taking advantage of the informational content of time endogeneity, which leads to an efficiency gain and narrower confidence bounds. Through

[23] As previously argued, the time between observations may be informative about the price process, with trades occurring more or less frequently depending upon the news flow, which itself will generally influence the price process.

Monte Carlo study, we study the optimal choice of the convolution frequency in the presence of microstructure noise. We also compare the performance of the Laplace estimator with other popular estimators through forecasting exercises by employing high frequency data. We conclude that the naive and the bias-corrected Laplace estimator perform better than most estimators in terms of forecasting equity return volatility.

A question that is yet to be addressed rigorously is a data-driven bandwidth choice in the presence of microstructure noise. A natural candidate is suggested in the Monte Carlo study in Section 2.5.2.

Another very promising extension that will be considered in the future paper is inferencing for a multivariate parameter. The challenge arises due to the additional complications such as the fact that different stocks do not trade at the same time or so-called asynchronous trading.

The proposed Laplace estimator of integrated volatility could also be extended to an estimator of spot volatility, and naturally such a spot volatility estimator should have the same advantages, robust to noise components and time endogeneity. Intuitively, by applying the Fourier inversion formula to Eq. (2.2.8), we could easily obtain a consistent estimator. However, the difficulty arises from deriving a stable convergence in law results.

Chapter 3

Robust Volatility Estimation Under Time Endogeneity

This chapter studies the efficient estimation problem of the volatility of asset returns in the presence of both market microstructure noise and time endogeneity using the realized kernel. Central limit theorem of the realized kernel estimator are established under a general nonparametric setting. We further show that the realized kernel estimator can be made to converge at the fastest possible rate, robust to different specifications of the noise and robust to dynamics in the noise process. The finite sample properties of the estimator are studied through Monte Carlo simulations and feasible confidence bounds of the proposed estimator are constructed, while empirical work illustrates its use in practice.

3.1 Introduction

Volatility is an important parameter for asset pricing theory as well as practical financial management. This has led to a rich literature on its estimation, mostly based on time series observations of asset returns. Over the last two decades tick-by-tick data for many liquid financial markets have become available. There have been dramatic development in the theory and application of the model free measurement of volatility. The econometric literature first focused on spot volatility, see [35] and [27]. The first generation of the literature in econometrics and statistics regarding the Integrated Volatility including but not limited to: [5], [13], [14], [6] and [64].

The efficiency of methodologies proposed often rely on the assumption of no microstructure noise, i.e., the observability of the true price process. Whereas in real life, observed prices are contaminated by market microstructure noise, especially as the sampling frequency increases. Observations are often subject to a substantial amount of noise, such as price discreteness, bid-ask spread bounces, informational asymmetries of traders, data errors, etc. Microstructure noise may also interacts with the sampling frequency and distinguishes itself from the classical measurement error problem in statistics, see [70], [39], [42], [66] and [43]. The above methodologies have recently been extended to account for the microstructure noise, for example, two-scale Realized Volatility ([77]), multi-scale Realized Volatility ([78]), flat-top realized kernel estimator ([16]), pre-averaging method ([50]), Fourier estimator ([61] and [67]) and the QMLE estimator ([76]). Related works include but not limited to: [10], [11], [41], [33] and [52].

Another common trait of the prevailing approaches is to directly measure the size of the observed returns over an exogenously given time grid, even in the form of equidistant sampling. In real life, a distinctive characteristic of the high frequency financial data is that they are observed at times that are random and not evenly spaced. In fact, there is time endogeneity in the price sampling times, as the time between observations could be informative about the price process. This is due to the frequency of trades more or less depends upon the information flow, which will influence the price process. This issue has only been brought to econometricians' attention recently. Earlier literature often considers the case where the sampling times are irregular but independent of the return, see [2], [37] and [45] for example. [58] first study the effect of time endogeneity on the estimation of Integrated Volatility in a nonparametric setting. They establish a central limit theorem for the Realized Volatility in a general endogenous time setting and propose a test and document that time endogeneity is present in the high frequency financial data. [57] extend the result in [58] and consider the case of both market microstructure noise and time endogeneity. They adopt a local/moving average approach and consider the time endogeneity on the sub-grid level. Related works also includes but not limited to [21], [46], [56] and [69].

In this paper we will focus on the realized kernel estimator of the Integrated Volatility, which is equivalent to the frequency domain estimator as illustrated in Section 3.2.1, see [67] also. The aim of this paper is to study the robustness of the realized kernel estimator of the Integrated Volatility towards time endogeneity and the presence of microstructure noise. More formally, according to the fundamental theorem of asset pricing, see [32], we start with the evolution of the log-price of some asset prices follows an Itô process

$$dX_t = \mu_t \, dt + \sigma_t \, dW_t, \tag{3.1.1}$$

where W_t is a Wiener process and the Integrated Volatility over the time period $[0, T]$ is given by $\int_0^T \sigma_t^2 \, dt$. We first consider the naive case that the efficient price process X is observable and derive the limiting distribution of the realized kernel estimator under time endogeneity. The fastest convergence rate is $N^{1/2}$ with N denotes the number of observations. Our result is in line with [58]. In the sense that the quadratic variation of the limiting process is unchanged, whereas it merely reapportion some of this quadratic variation to asymptotic mean. Our result also covers the situation where the observations are equidistant or irregularly spaced but independent of the return as special cases.

Next, we move to case of the presence of microstructure noise. We suppose the logarithm of the observed price process Y is the sum of the efficient price process X and a microstructure noise component U. As a starting point, we assume the noise component is represented by independent identically distributed random variables. We derive the explicit analytical expression of the bias of the realized kernel estimator and show that the resulting estimator after correcting the bias is asymptotically mixed Gaussian with a fastest convergence rate at $N^{1/5}$. According to our result, the effect of time endogeneity will not deteriorate the convergence rate but will affect the asymptotic variance of the limiting process.

Inspired by the work of [16], where they study the performance of the flat-top realized kernel estimator and conclude the fastest convergence rate can be improved to $N^{1/4}$. We also consider the performance of the flat-top realized kernel estimator as an extension under time endogeneity. Our finding is that the convergence rate of the flat-top kernel estimator is unaffected but depends heavily on the white noise assumption of the noise component. Our result is also robust to the case of the dependent noise. As discussed in [68] and [52], the white noise assumption may be unsatisfactory for empirical application. We relax the white noise assumption to allow the microstructure noise to be serially dependent, so that the realized kernel estimator could be applied to the ultra high frequency data that are sampled at the highest possible frequency. We conclude that under the case of the dependent noise, only the asymptotic variance of the limiting process will be affected and the fastest convergence rate is $N^{1/5}$. To dealt with the case that the microstructure noise interacts with the sampling frequency and the empirical evidence that microstructure noise is small relative to the Integrated Volatility, we also consider the case of a shrinking noise. Under the case of a shrinking noise, we can achieve a faster convergence rate than $N^{1/5}$. The fastest convergence rate depends on the rate of the shrinkage of the noise. The case of the shrinking noise turns out to be important empirically due to the finite sample size and a small noise-to-signal ratio.

We also conduct simulation studies to examine the finite sample property of the realized kernel estimator under time endogeneity and the presence of the microstructure noise. Our results confirm the finding that the realized kernel estimator could be adopted in a model with both time endogeneity and dependent noise. Another surprising finding is that, under the case of a small noise-to-signal ratio, the realized kernel estimator based on a shrinking noise performs better even if the data generating process admits a non-shrinking noise in finite sample. We also make the central limit results feasible and construct confidence bounds of the proposed estimator. Last, we apply our methodology to the millisecond time stamped trade data and illustrate that the realized kernel estimator could be adopted in a model that are sampled at the highest possible frequency.

The paper is organized as follows. Section 3.2 introduces our version of the realized kernel estimator and discusses the connection between it and the frequency domain estimator. We also study its asymptotic properties under the presence of time endogeneity and the absence of the microstructure noise. Section 3.3 studies the asymptotic property of the realized kernel estimator under time endogeneity and the presence of microstructure noise. In Section 3.4, we discuss three different extensions: the flat-top realized kernel estimator, the case of dependent noise and the case of shrinking noise. Simulation studies are carried out in Section 3.5, and an empirical analysis is presented in Section 3.6. We conclude in Section 3.7.

As regards the notations, $\Delta(\cdot)$ is the differencing operator, $\langle X, X\rangle_T$ denotes the Integrated Volatility $\int_0^T \sigma_u^2\, du$, $[X, X]_T$ denotes the quadratic variation of a stochastic process which is defined as $\lim_{N\to\infty} \sum_{i=0}^N \left[X_{t_{i+1}} - X_{t_i}\right]^2$, in Section 3.4.2 we use $a \simeq b$ to denote $a = b + o(b)$. We denoted by $\xrightarrow{p}$ convergence in probability and $\xrightarrow{\mathcal{L}-s}$ stable convergence in law. Most of the results rely on the concept of the stable convergence in law, see also [3].

3.2 Realized Kernel Estimators Under Time Endogeneity

The aim of this section is to motivate and introduce a class of realized kernel estimators. Given a possibly endogenous sampling of the price process, we study the asymptotic properties of the estimator of Integrated Volatility in Section 3.2.2.

3.2.1 Setting and methodology overview

The evolution of the log-price is described by a standard Itô process,

$$dX_t = \mu_t\, dt + \sigma_t\, dW_t, \tag{3.2.1}$$

where W_t is a Brownian motion on a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, P)$ satisfying the usual conditions. It is further assumed that:

Assumption H3.1. *The process μ_t and σ_t are adapted stochastic processes such that*

$$E\left[\int_0^T \sigma_t^4\, dt\right] < \infty, \quad E\left[\int_0^T \mu_t^2\, dt\right] < \infty,$$

and $ess\, sup||\sigma^2||_\infty < \infty$, where $sup||\sigma^2||_\infty = sup_{0\leq t\leq T}|\sigma_t^2|$.

The target of inference is the Integrated Volatility

$$\langle X, X\rangle_T = \int_0^T \sigma_u^2\, du. \tag{3.2.2}$$

Let $(t_{n,i})_{i=0,\dots,N}$ be observations times in the interval $[0,T]$, where n characterizes the observation frequency and N stands for the actual number of observations before time T, for simplicity we assume $t_{n,N}=T$ and for notation clarity we always suppress n. Let $\Delta X_{t_i} = X_{t_{i+1}} - X_{t_i}$ and $\Delta t_i = t_{i+1} - t_i$.

Normally, the Integrated Volatility is computed over discrete time intervals using the quadratic variation formula,

$$[X,X]_T = \lim_{N\to\infty} \sum_{i=0}^{N} \left[X_{t_{i+1}} - X_{t_i}\right]^2.$$

The limit is finite (also well defined) with probability one and almost surely equals to the Integrated Volatility as in Eq. (3.2.2). A natural estimator of the Integrated Volatility is the Realized Volatility by exploiting the quadratic variation formula ([4] and [14]). From a practical point of view, the performance of the Realized Volatility is unsatisfying due to two important issues. First, the observed price process is subject to a substantial amount of noise due to the imperfection of the system and thus deviates from Eq. (3.2.1). Second, the time between two observations are informative about the price process, with trades occurring more or less frequently depending upon the news flow, and price reflects publicly available information. The objective in this study is to overcome both problems.

In this paper, we focus on the realized kernel estimator due to its natural robustness to the presence of the microstructure noise. Also as discussed in [18], the realized kernel estimator is also related to the frequency domain estimator (Fourier and Laplace type estimators) which does not require data synchronization. The performance of realized kernel estimator under the presence of the microstructure noise has been examined in [16] and [18]. In this section, we study the asymptotic limit of the estimator under time endogeneity. In [18], they define the univariate realized kernel as,

$$k(X) = \sum_{t_{i+1}\le T} \sum_{t_{j+1}\le T} k\left(\frac{i-j}{H_1}\right) \Delta X_{t_i} \Delta X_{t_j}, \tag{3.2.3}$$

where $k(\cdot): \mathbb{R} \to \mathbb{R}$ is a non-stochastic weight (even) function with $k(0)=1$. In this study, we define the realized kernel estimator as,

$$k(X) = \sum_{t_{i+1}\le T} \sum_{t_{j+1}\le T} k\left(\frac{t_i - t_j}{H}\right) \Delta X_{t_i} \Delta X_{t_j}. \tag{3.2.4}$$

Our definition of the realized kernel is motivated by the Fourier type estimators (based on the Fourier transformation, see [61]) and Laplace type estimators (based on the Laplace transfer, see [30]), where t_i and t_j shows up naturally after taking the Bohr convolution product. Our definition is comparable with the definition in [18] under the case of equidistant observations, with $H_1 = HN/T$.

Realized kernel estimator also belongs to the class of Fourier (Laplace) type estimator.[1] As illustrated in [67], one can define the spectral window generator $K(\cdot)$ as a Fourier transform of the lag window (realized kernel) $k(\cdot)$:

$$K(\lambda) = \frac{1}{2\pi}\int_{-\infty}^{\infty} k(t)e^{-i\lambda t}\,dt \qquad k(t) = \int_{-\pi}^{\pi} K(\lambda)e^{i\lambda t}\,d\lambda$$

where λ denotes the angular frequency. Then the frequency domain (Fourier and Laplace type) estimator with $K(\cdot)$ is equivalent to the time domain (realized kernel) estimator with $k(\cdot)$.

3.2.2 Asymptotic properties under time endogeneity

In this section, we derive the asymptotic distribution of the realized kernel estimator under the case of time endogeneity and in the absence of the microstructure noise.

Assumption H3.2. $\max_i(\Delta t_i) \to 0$ *as* $N \to \infty$, *and* $\frac{\max_i(\Delta t_i)}{T/N} = O(1)$.

Assumption H3.3. *As* $\frac{1}{HN} \to c$ *and* $H \to 0$, $N \to \infty$, *where c is a nonnegative constant and let* $t^*(\tau)$ *be the largest time* t_i *smaller than or equal to* τ, *we have the convergence of the sequence of the functions,*

$$H^{-1}\int_0^t\int_0^v k\left(\frac{t^*(u)-t^*(v)}{H}\right)^2 du\,dv \to \int_0^t \Psi(v)\,dv, \tag{3.2.5}$$

for all $t \in [0,T]$ *and* Ψ *is some integrable function.*

Assumption H3.2 and H3.3 are the standard assumption when dealing with irregularly spaced data. Assumption H3.2 is commonly assumed in the literature, it could be weakened to allow Poisson type sampling schemes. See [58] for the case of time endogeneity and [18] in the presence of the microstructure noise. Assumption H3.3 is necessary for a central limit type result.

Assumption H3.4. *As* $\frac{1}{HN} \to c$ *and* $H \to 0$, $N \to \infty$, *where c is a nonnegative constant, the following convergence in probability hold*

$$H^{-1/2}\sum_{t_{j+1}\le T}\left(\sum_{t_{i+1}\le T} k\left(\frac{t_i - t_j}{H}\right)\Delta X_{t_i}\right)\Delta t_j \xrightarrow{p} \int_0^T a_v\sigma_v\,dv, \tag{3.2.6}$$

$$H^{-1}\sum_{t_{j+1}\le T}\left(\sum_{t_{i+1}\le T} k\left(\frac{t_i - t_j}{H}\right)\Delta X_{t_i}\right)^2\Delta t_j \xrightarrow{p} \int_0^T b_v\sigma_v^2\,dv, \tag{3.2.7}$$

[1] Laplace type estimator are based on the Laplace transform instead of Fourier transform, and it avoids the artificial "periodization" of the asset price process subjacent to Fourier series methodology, see [30]. Other than that, the difference between them are trivial and they have the same limiting distribution.

for some (well defined) process a_v and b_v.

Assumption H3.4 could be viewed as our definition of time endogeneity, the covariance between the kernel weighted return process ($\sum_{t_{i+1}\leq T} k\left(\frac{t_i-t_j}{H}\right)\Delta X_{t_i}$) and the observation frequency (Δt_j) is non-zero. The convergence of this covariance in Eq. (3.2.6) is crucial for the consistency of the estimator, Eq. (3.2.7) is needed to derive the limiting distribution of the estimator. As mentioned [58], there is a tight connection between $\left[\sum_{i=0}^{N}(X_{t_{i+1}}-X_{t_i})(t_{i+1}-t_i)\right]$ and a nonzero probability limit of tricity $\sum_{i=0}^{N}\left[X_{t_{i+1}}-X_{t_i}\right]^3$ in the context of the Realized Volatility. They further established a central limit theorem for the tricity under the null hypothesis that the observation times are independent of the efficient price process and conducted empirical work that provides evidence that there is time endogeneity in the high frequency data.

Theorem 3.2.1. *Suppose Assumptions H3.1 – H3.4 hold, we have*

$$H^{-1/2}(k(X)-\langle X,X\rangle_T) \xrightarrow{\mathcal{L}-s} 2\int_0^T a_v\sigma_v\,dX_v + 2\int_0^T \sqrt{b_v-\Psi(v)-a_v^2}\sigma_v^2\,dB_v, \qquad (3.2.8)$$

where the convergence is stable in law, B_v is a Brownian motion independent of the underling σ-field.

The fastest rate of convergence rate is by taking $H^{-1}=cN$ for some positive c, and the convergence rate is $N^{1/2}$. The case of $H^{-1}=cN$ with $c>0$ is useful under the case of sparse sampling where the effect of the microstructure noise is considered to be small and thus negligible. On the other hand, the case of $c=0$ is more attractive empirically, as will be shown in Section 3.3, due to the presence of the microstructure noise. Setting $c=0$ and exploiting the relationship between the realized kernel estimator and the frequency domain estimator discussed in Section 3.2.1, H^{-1} can be interpreted as the bandwidth of the convolution (the Bohr convolution product) frequency. By letting $\frac{1}{HN}\to 0$ ($H^{-1}=o(N)$), one essentially cuts the highest frequencies in the frequency domain estimator and permits the microstructure noise to be filtered out, providing very accurate volatility estimate over the interval.

Theorem 3.2.1 also covers the case that observation times are irregularly spaced but without time endogeneity. Under the case of no time endogeneity, we have $a_v=0$ (the covariance is zero) and $b_v=2\Psi(v)$, (see [29]). The main result in Section 3.3 could be accommodated analogously. Furthermore, if the only case of interest is $c=0$, then $\Psi(v)=\int_0^\infty k(t)^2\,dt$ by the use of Riemann sum. Substituting $a_v=0$ and $b_v=2\Psi(v)=2\int_0^\infty k(t)^2\,dt$ into Theorem 3.4.1, we obtain the same result as in [16].

Another special case that has been well studied in the literature is under the case of equidistant sampling scheme. It is well known that setting $k(t)=\frac{sin\,t}{t}$ or equivalently $K(\lambda)=1$ corresponds to the Laplace estimator. By choosinig c to be a multiple of π/T, the Laplace

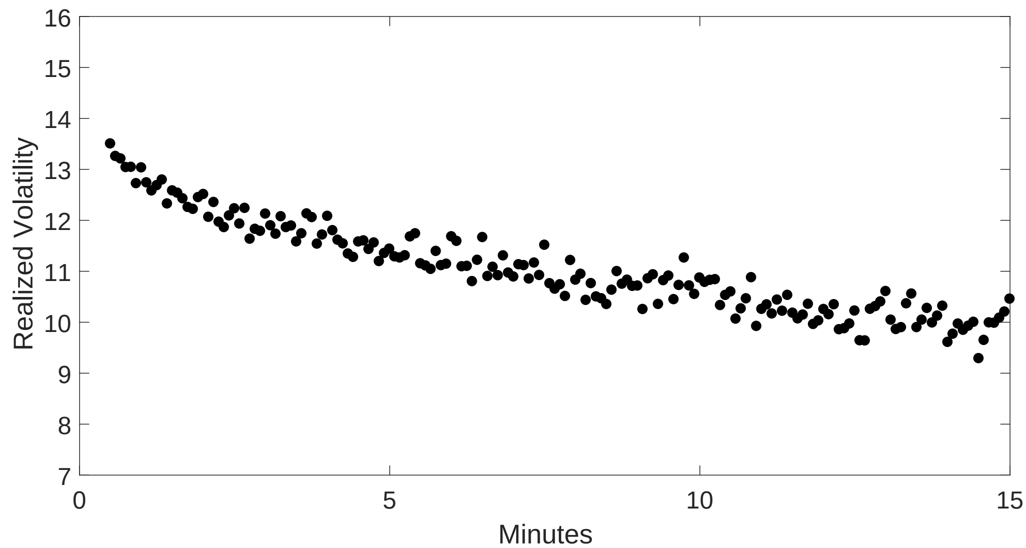

Figure 3.1 Volatility signature plot of the annualized Realized Volatility for IBM, 2011. For a description of the data, see Section 3.6

estimator achieve the same efficiency level as the Realized Volatility and suggests that the presence of autocovariance is not detrimental to the asymptotic variance and convergence rate, see [29].

Last but not least, we do not consider the case that $\frac{1}{HN} \to \infty$ in Theorem 3.2.1. Because under Assumption H3.6 in Section 3.3.2, we have $\lim_{t\to\infty} k(t) = 0$, the realized kernel estimator is asymptotically equivalent to the Realized Volatility, which has been well studied in the literature. For an extension of the Realized Volatility under the case of time endogeneity, see [58]. [57] studies the performance of the Realized Volatility based on a local/moving averaging approach under the case of time endogeneity and microstructure noise.

3.3 Main Results

In this section, we derive the asymptotic distribution of the realized kernel estimator under the presence of both time endogeneity and the microstructure noise. A good starting point, although unsatisfactory empirically, is to assume the additive noises are independent and follow a white noise process.

3.3.1 Assumption on the microstructure noise

Evidence from the volatility signature plot suggests that the behavior of the observed price deviates from Eq. (3.2.1). As can be seen from Figure 3.1, which uses the IBM transaction data in 2011. In fact, various studies document that the high frequency financial data often has infinite quadratic variation. To account for this phenomena, we assume the observed price process is the sum of two process, the efficient price process X and a microstructure noise U due to market frictions.

$$Y_{t_i} = X_{t_i} + U_{t_i}$$

Assumption H3.5. *U is a white noise process and is independent of the efficient price process X, with $E(U) = 0$, $Var(U) = \epsilon^2$ and $Var(U^2) < \infty$.*

We can think of X as the log-price in equilibrium and is the price that would prevail in the absence of the microstructure noise. The additive microstructure noise, where the efficient price is a continuous process and the noise is observed at exactly the times where there are trades, has been well studied in the literature, and various estimators have been proposed. See, for example, two-scale Realized Volatility ([77]), multi-scale Realized Volatility ([78]), flat-top realized kernel estimator ([16]), pre-averaging method ([50]), Fourier domain based estimator ([61] and [67]) and the QMLE estimator ([76]).

Here, in this section we focus on the case of Assumption H3.5, where the microstructure noise ΔU displays an $MA(1)$ structure with a negative first order autocorrelation ($-\epsilon^2$). Assumption H3.5 is known to be a realistic approximation in decentralized markets, for example the foreign exchange market. The assumption that U and X are independent is discussed in [41]. Their result suggests that the independence assumption is not too damaging statistically.

Assumption H3.5 is fairly restrictive and may be insufficient for empirical application, as it excludes the case of dependent noise and endogenous noise, see [52]. However, Assumption H3.5 is a useful starting point, as the case of dependent noise has been well studied in multiple literature (see [61], [16] and [67]), and the case of endogenous noise can be accommodated using theorem 1 in [18]. The discussion of different extensions of the structure of the microstructure noise are in Section 3.4.2. Assumption H3.5 also rules out the case where the noise may depend on the sampling frequency (shrinking noise). Section 3.4.3 provides an asymptotic analysis under the case of the shrinking noise.

In the following analysis as well as the simulation studies in Section 3.5, we assume $U_{t_0} = U_{t_N} = 0$, so that the border effects could be ignored. Equivalently the border effects could be ignored by using jittering to redefine the observed prices at two end points t_0 and t_N to be respectively the average of m number distinct observations to eliminate end-effects, see [18]. The econometrician does not observe the efficient price process X. As a result, the realized kernel estimator should be constructed using the contaminated returns,

$$\begin{aligned} k(Y) &= \sum_{t_{i+1}\leq T}\sum_{t_{j+1}\leq T} k\left(\frac{t_i-t_j}{H}\right)\Delta Y_{t_i}\Delta Y_{t_j} \\ &= \underbrace{\sum_{t_{i+1}\leq T}\sum_{t_{j+1}\leq T} k\left(\frac{t_i-t_j}{H}\right)\Delta X_{t_i}\Delta X_{t_j}}_{A} + 2\times \underbrace{\sum_{t_{i+1}\leq T}\sum_{t_{j+1}\leq T} k\left(\frac{t_i-t_j}{H}\right)\Delta X_{t_i}\Delta U_{t_j}}_{B} \\ &\quad + \underbrace{\sum_{t_{i+1}\leq T}\sum_{t_{j+1}\leq T} k\left(\frac{t_i-t_j}{H}\right)\Delta U_{t_i}\Delta U_{t_j}}_{C}. \end{aligned} \tag{3.3.1}$$

3.3.2 Asymptotic properties

In this section, we aim to derive the asymptotic distributions of A, B and C in Eq. (3.3.1). The asymptotic behavior of A follows from Theorem 3.2.1, Proposition 3.3.1 gives the asymptotic result for B, Proposition 3.3.2 gives the asymptotic result for C and at the end, the limiting distribution of the realized kernel estimator Eq. (3.3.1) will be derived under the presence of time endogeneity and the microstructure noise. All proofs are contained in Appendix B.1.

The realized kernel $k(\cdot)$ is assumed to satisfy the following assumption (see Assumption K in [18] and Assumption 3 and $3'$ in [67]).

Assumption H3.6. *The lag window $k(\cdot)$ satisfies (i) $k(0) = 1$, $k'(0) = 0$; (ii) k is even and twice continuously differentiable; (iii) $||k'||^2 = \int_0^\infty k'(x)^2\,dx < \infty$ and $||k''||^2 = \int_0^\infty k''(x)^2\,dx < \infty$. The prime notation here is taken to mean differentiation with respect to the argument.*

We normalize the lag window $k(\cdot)$ by imposing $k(0) = 0$, which assigns unit weight to the variance terms $\Delta X_{t_i}^2$. $k'(0) = 0$ indicates that the kernel gives close unit weight to the first order autocovariance terms $\Delta X_{t_i}\Delta X_{t_{i+1}}$. $k'(0) = 0$ is also important under the presence of time endogeneity, as shown in Appendix B.1, it assigns zero weight to the terms affected by the time endogeneity.

Proposition 3.3.1. *Under the Assumptions H3.1 – H3.6, and that $\frac{1}{HN} \to 0$ as $H \to 0$, $N \to \infty$, then we have*

$$\sqrt{HN}B \xrightarrow{\mathcal{L}-s} N\left(0, 2\epsilon^2 T\langle X, X\rangle_T \left[\int_0^\infty k'(t)^2\,dH(t)\right]\right), \tag{3.3.2}$$

where $H(t)$ is the quadratic variation of time, see Assumption A in [64].

A striking result about Proposition 3.3.1 is that the asymptotic variance is unaffected by the presence of time endogeneity, as shown in the Appendix B.1, this is essentially due to $k'(0) = 0$, so that it assign zero weight to the covariance between observation frequency and the return, we can safely treat it as if there is no time endogeneity. The quadratic variation of time $H(t)$ is defined as,

$$H_{(N)}(t) = \frac{\sum_{t_{i+1} \le t} \Delta t_i^2}{(T/N)} \to H(t) \text{ as } N \to \infty. \tag{3.3.3}$$

Here it shows up naturally due to the observation times are irregularly spaced. It is also worth noting that, given equidistant observations we have $H(t) = t$. Proposition 3.3.1 indicates that the asymptotic variance of the limiting process is $2\epsilon^2 T \langle X, X \rangle_T \left[\int_0^\infty k'(t)^2\, dt\right]$, the same as in [16]. Next, we derive the asymptotic result for C.

Proposition 3.3.2. *Under the Assumptions H3.1 – H3.6, and that $\frac{1}{HN} \to 0$ as $H \to 0$, $N \to \infty$, then we have*

$$H^2 NC \xrightarrow{p} \epsilon^2 |k''(0)| T H(T). \tag{3.3.4}$$

Proposition 3.3.2 implies that the leading bias term C is of order $O_p(H^{-2}N^{-1})$. In order for it to be asymptotically negligible requires that $1/(H^2 N) \to 0$. Another finding is that through a direct comparison between Proposition 3.3.1 and 3.3.2 indicates that, comparing to C, we can ignore the effect of B asymptotically. Combining Theorem 3.2.1 and Proposition 3.3.2, by setting $H^{-1/2} = O_p\left(H^2 N\right)$, and we can achieve the fastest convergence rate at $N^{1/5}$.

Theorem 3.3.3. *Under the Assumptions H3.1 – H3.6, and that $\frac{1}{H^5 N^2} \to c$ as $H \to 0$ and $N \to \infty$, where c is a nonnegative constant. Then we have*

$$H^{-1/2}(k(Y) - \langle X, X \rangle_T) \xrightarrow{\mathcal{L}-s} \sqrt{c}\epsilon^2 |k''(0)| T H(T) + 2\int_0^T a_v \sigma_v dX_v + 2\int_0^T \sqrt{b_v - \Psi(v) - a_v^2}\sigma_v^2 dB_v \tag{3.3.5}$$

where the convergence is stable in law, B_v is a Brownian motion independent of the underling σ-field.

The fastest convergence rate according to Theorem 3.3.3 is $N^{1/5}$, which is inferior to the flat-top realized kernel discussed in [16] even under the naive case of white noise. On the other hand, it has the same convergence rate as the realized kernel estimators discussed in [18] and the Fourier realized kernel estimators discussed in [67]. The slower convergence rate is mainly due to the asymptotic bias in C, as shown in Proposition 3.3.2. Using the flat-top realized kernel the asymptotic bias in C becomes 0 , due to the first order autocovariance balance the bias. This suggest an improvement to the estimator and will be studied in Section 3.4.1.

3.4 Extensions

In this section, we consider three extensions of the realized kernel estimator discussed in Section 3.3. The first extension is to improve the convergence rate using the flat-top realized kernel discussed in [16]. It is well known that under the white noise assumption, the flat-top kernel estimator under equidistant observations converges at a $N^{1/4}$ rate and is close to the parametric efficiency bound. In Section 3.4.1, we show that such an improvement is still possible under the presence of time endogeneity. In Section 3.4.2, we discuss the case of dependent noise. In Section 3.4.3, we consider the case of the shrinking noise and shows that the fastest convergence rate is no longer $N^{1/5}$. The case of the shrinking noise is of crucial importance empirically, as shown in the simulation studies in Section 3.5. Given a small noise-to-signal ratio and not too large number of observations, treating the small microstructure noise as if the case of the shrinking noise leads to more accurate estimate in terms of both MSE and bias.

3.4.1 Improved efficiency

The result in Theorem 3.3.3 looks weak compared to the result for the flat-top realized kernel estimator introduced by [16] under the same white noise assumption. Their flat-top realized kernel estimator allows a faster convergence rate at $N^{1/4}$. More specifically, they consider a special class of flat-top realized kernel,

$$k^F(Y) = \sum_{t_{i+1}\leq T} \sum_{t_{j+1}\leq T} k^F(\frac{i-j}{H_1})\Delta Y_{t_i}\Delta Y_{t_j}, \tag{3.4.1}$$

with

$$\begin{cases} k^F(\frac{i-j}{H_1}) = k(\frac{i-j}{H_1}), & \text{when } |i-j| > 1, \\ k^F(\frac{i-j}{H_1}) = 1, & \text{when } |i-j| \leq 1. \end{cases} \tag{3.4.2}$$

As mentioned in Section 3.3.2, the flat-top realized kernel corrects the asymptotic bias in $C\left(\sqrt{c}\epsilon^2|k''(0)|TH(T)\right)$ and allows an improvement of Proposition 3.3.2 to a central limit type result, which leads to a faster convergence rate under the case of the white noise error. The variance of ΔU is exactly balanced by both of the first order autocovariance of ΔU. Motivated by this, in this section, we consider the following improved realized kernel estimator,

$$
\begin{aligned}
k^F(Y) &= \sum_{t_{i+1}\le T}\sum_{t_{j+1}\le T} k^F\left(\frac{t_i-t_j}{H}\right)\Delta Y_{t_i}\Delta Y_{t_j} \\
&= \underbrace{\sum_{t_{i+1}\le T}\sum_{t_{j+1}\le T} k^F\left(\frac{t_i-t_j}{H}\right)\Delta X_{t_i}\Delta X_{t_j}}_{A_F} + 2\times \underbrace{\sum_{t_{i+1}\le T}\sum_{t_{j+1}\le T} k^F\left(\frac{t_i-t_j}{H}\right)\Delta X_{t_i}\Delta U_{t_j}}_{B_F} \\
&\quad + \underbrace{\sum_{t_{i+1}\le T}\sum_{t_{j+1}\le T} k^F\left(\frac{t_i-t_j}{H}\right)\Delta U_{t_i}\Delta U_{t_j}}_{C_F}.
\end{aligned}
\tag{3.4.3}
$$

where $k^F(\cdot)$ is the flat-top realized kernel,

$$
\begin{cases}
k^F\left(\frac{t_i-t_j}{H}\right) = k\left(\frac{t_i-t_j}{H}\right), & \text{when } |i-j|>1, \\
k^F\left(\frac{t_i-t_j}{H}\right) = 1, & \text{when } |i-j|\le 1.
\end{cases}
\tag{3.4.4}
$$

Here, the improved realized kernel estimator is the same as the flat-top realized kernel estimator with infinite lag based on irregularly spaced observations. The differences between ours and [16] are twofolds. First our estimator is irregularly spaced, and the realized kernel is based on t_i and t_j instead of i and j as in [16] and [18], which leads to different asymptotic variance. The second difference is the possible presence of time endogeneity. In [16], times are allowed to be irregularly spaced if $t_i = F(i/N)$, where F is allowed to be a smooth random process. However, as stated in Example 2 in Section 3 of [58], this accommodates some endogeneity of time, but with $a_v = 0$.

Theorem 3.4.1 below shows that the improved estimator achieve the fastest convergence rate at $N^{1/4}$, the same rate as the flat-top realized kernel estimators in [16].

Theorem 3.4.1. *Under the Assumptions H3.1 – H3.6, and that* $\frac{1}{H^2N} \to c$ *as* $H \to 0$ *and* $N \to \infty$*, where* c *is a nonnegative constant. Then we have*

$$
H^{-1/2}(k^F(Y) - \langle X, X\rangle_T) \xrightarrow{\mathcal{L}-s} 2\int_0^T a_v\sigma_v\, dX_v + MN\left(0, V_1+V_2+V_3\right), \tag{3.4.5}
$$

where MN denotes a mixed normal distribution, the convergence is stable in law and

$$
\begin{aligned}
V_1 &= 4\int_0^T (b_v - \Psi(v) - a_v^2)\sigma_v^4\, dv, \\
V_2 &= 8\epsilon^2 cT\left(\int_0^\infty k'(t)^2\, dH(t)\right)\langle X, X\rangle_T, \\
V_3 &= 4T^2c^2\epsilon^4 H(T)\left(\int_0^\infty k''(t)^2\, dH(t)\right).
\end{aligned}
$$

The fastest conference rate can be improved to $N^{1/4}$, but our result also covers the case that $\frac{1}{H^2N} \to 0$, so that the convergence rate is slower than $N^{1/4}$, under this case $c = 0$, V_2 and V_3 vanishes. A special case is that given equidistant observations, then $H(t) = t$, $a_t = 0$, $b_t = 2\Psi(t)$ with $\Psi(t) = \int_0^\infty k(t)^2\, dt$ and Theorem 3.4.1 indicates that the asymptotic variance of the limiting process is the same as in Theorem 4 of [16] by noticing that $H_1 = HN/T$.

Theoretically, the improvement of the realized kernel estimator discussed in this section depends heavily on Assumption H3.5. Empirically, the white noise assumption is often too restrictive, under such circumstance, an improvement seems impractical without knowledge of the temporal correlations of the microstructure noise. However, simulation results in Section 3.5.2 indicates that deviation from Assumption H3.5 is not too damaging statistically given a small noise-to-signal ratio. Our explanation of this phenomenon is that the case of a small noise-to-signal ratio and moderate small sample size resembles the case of the shrinking noise. Under the case of the shrinking noise, as will be shown in Theorem 3.4.2, we obtain a faster convergence rate than $N^{1/5}$. On the other hand, given the temporal correlations of the noise, one can construct a realized kernel accordingly to eliminate the asymptotic bias in C and achieve a faster convergence rate.

3.4.2 Dependent noise

As previously noted in Section 3.3.1, Assumption H3.5 may be unsatisfactory empirically (see [68] and [52]). Now we relax Assumption H3.5 to allow the microstructure noise to be serial dependent, so that $k(Y)$ could be applied to the ultra high frequency data that are sampled at the highest possible frequency.

An important observation is that the non-zero covariance between the return and observation frequency will not affect the pure noise terms C, so that time endogeneity plays no role in deriving its asymptotic result. Thus the result in the literature can be readily applied with minor modification. We provide simulation evidence that the realized kernel could be applied in a model with dependent microstructure noise in Section 3.5.3.

The case of dependent noise is well studied in the literature for both Fourier based estimator

([67]) and realized kernel estimator ([16] and [18]). In [16], they allow U to be serial dependent,

$$\sum_i a_i U_{t_i} = O_p(1) \quad \text{for any} \quad \sum_i a_i^2 = O(1).$$

Their finding is that $H^2NC = O_p(1)$, in which case the noise will not have any impact if we choose $c = 0$ in Theorem 3.3.3. On the other hand, the fastest possible rate of convergence rate is unchanged at $N^{1/5}$.

In [67], they assume a standard absolute summability condition for the autocovariance function on the microstructure noise.

$$\frac{1}{N}\sum_i \sum_j \gamma\left(t_i, t_j\right) \to \Gamma_U,$$

where $\gamma\left(t_i, t_j\right) = E\left(U_{t_i} U_{t_j}\right)$. Their finding is in line with the result in [16], the fastest rate of convergence rate is $N^{1/5}$. Following the same step as in Appendix B.1, if we assume the same absolute summability condition for the autocovariance function on the microstructure noise, the convergence rate in Theorem 3.3.3 is unchanged whereas the asymptotic bias is given by,[2]

$$H^2NC \simeq \sqrt{c}|k''(0)|TH(T)\Gamma_U^2.$$

3.4.3 Shrinking microstructure noise

Empirical evidence suggests that ϵ^2 is small relative to the Integrated Volatility, it becomes interesting to consider the case that $\epsilon^2 \to 0$ locally. In this section, we consider another possible specification of the additive noise process.

$$Y_{t_i} = X_{t_i} + U_{t_i}^N$$

Assumption H3.7. *$U^N = \rho_N \times U$ with U satisfies Assumption H3.5 and $\rho_N \to 0$.*

Here we assume U satisfies Assumption H3.5 without loss of generality, as the case of dependent noise could be treated accordingly. The case of $\rho_N \to 0$ corresponds to the case of the shrinking noise. It is important for the empirical application of the realized kernel estimator due to two different reasons: the finite sample size and the empirical evidence that the noise-to-signal ratio is small.

Substituting $Var(U^N) = \rho_N^2 \epsilon^2$ into Proposition 3.3.1 and 3.3.2, we have, as $\frac{1}{HN} \to 0$, $H \to$

[2]The proof is almost identical to the Proof of Proposition 3.3.2 given in Appendix B.1, see [67] also.

0, $N \to \infty$ and $\rho_N \to 0$,

$$\rho_N^{-1}\sqrt{HN}B \quad \xrightarrow{\mathcal{L}-s} \quad N\left(0, 2\epsilon^2 T\langle X, X\rangle_T\left[\int_0^\infty k'(t)^2\,dH(t)\right]\right), \tag{3.4.6}$$

$$\rho_N^{-2}H^2NC \quad \xrightarrow{p} \quad \epsilon^2|k''(0)|TH(T). \tag{3.4.7}$$

Following a similar argument, B is of smaller order compared to C and thus is asymptotically negligible. Next, combine Theorem 3.2.1 and Eq. (3.4.7), by setting $H^{-1/2} = O_p(\rho_N^{-2}H^2N)$, we have the fastest convergence rate at $\rho_N^{-2/5}N^{1/5}$.

Theorem 3.4.2. *Under the Assumptions H3.1 – H3.7, and that* $\frac{\rho_N^4}{H^5N^2} \to c$ *as* $H \to 0$, $N \to \infty$, $\frac{1}{HN} \to 0$ *and* $\rho_N \to 0$, *where* c *is a nonnegative constant. Then we have*

$$H^{-1/2}(k(Y) - \langle X, X\rangle_T) \xrightarrow{\mathcal{L}-s} \sqrt{c}\epsilon^2|k''(0)|TH(T) + 2\int_0^T a_v\sigma_v dX_v + 2\int_0^T \sqrt{b_v - \Psi(v) - a_v^2\sigma_v^2}dB_v \tag{3.4.8}$$

where the convergence is stable in law, B_v *is a Brownian motion independent of the underling* σ*-field*

Theorem 3.4.2 indicates that we can achieve a faster convergence rate than $N^{1/5}$ under the case of the shrinking noise. The convergence rate also depends on the speed $\rho_N \to 0$. If we parametrize $\rho_N = N^{-\psi}$. When $0 < \psi < \frac{3}{4}$, then $H = O_p(N^{-\frac{2+4\psi}{5}})$ and the fastest rate is $N^{\frac{1+2\psi}{5}}$. When $\psi \geq \frac{3}{4}$, we have $c = 0$ and it is asymptotically equivalent to the case of no microstructure noise.

The case of the shrinking noise can be extended to the flat-top realized kernel as well, see [16] for the case of the equidistant observations. Although the discussion in this section focuses on the case of the white noise, it could be trivially extended to cover the case of the dependent noise, as only the asymptotic variance will be affected and the result regarding the convergence rate stays intact, see Section 3.4.2. The case of the shrinking noise turns out to be important empirically due to the finite sample size. We provide Monte Carlo evidence in Section 3.5 that by choosing H based on Theorem 3.4.2 instead of Theorem 3.3.3 in finite sample, one can obtain a more accurate estimate (in terms of MSE and bias).

3.5 Simulation Study

The analysis in previous sections has been asymptotic as $H \to 0$ and $N \to \infty$. Here we study the finite sample property of the proposed estimator through a simulation. The goal of Section 3.5.2 and 3.5.3 is to assess the accuracy of the infeasible central limit results in Section 3.3.2 and 3.4.1. A surprising finding is that, under the case of a small noise-to-signal ratio, the realized kernel estimator based on Theorem 3.4.2 outperforms the estimator based on Theorem 3.3.3 and 3.4.1. In Section 3.5.4, we aim to make them feasible, and construct confidence bounds.

3.5.1 Simulation design

The following stochastic volatility model is used,

$$\begin{cases} dX_t = \sigma_t dW_t, \\ d\sigma_t^2 = \gamma(\theta - \sigma_t^2)dt + \alpha\sigma_t dZ_t, \end{cases} \tag{3.5.1}$$

where $\langle dW, dZ\rangle_t = \lambda dt$ and W_t and Z_t are standard Brownian motions. We set the parameters as $\gamma = 0.01$, $\theta = 1$, $\alpha = 0.05$, $\sigma_0^2 = 1$, $\lambda = -0.5$ and $X_0 = log(100)$. Here, we follow the choice of [11] and [30] and set $\lambda \neq 0$ for the presence of the leverage effect. Second-by-second return and volatility paths over a daily trading period of 6 hour for a total of 1000 trading days are computed using the Euler discretization scheme ($T = 0.25$).

Throughout Section 3.5, we choose $k(t) = \frac{sin\, t}{t}$ due to this corresponds to the frequency domain estimator with $K(\lambda) = 1$, the naive Laplace estimator. We consider four different versions of the estimator,

$$k(Y) = \sum_{t_{i+1}\le T}\sum_{t_{j+1}\le T} k\left(\frac{t_i - t_j}{H}\right)\Delta Y_{t_i}\Delta Y_{t_j}, \tag{3.5.2}$$

$$k^F(Y) = \sum_{t_{i+1}\le T}\sum_{t_{j+1}\le T} k^F\left(\frac{t_i - t_j}{H_F}\right)\Delta Y_{t_i}\Delta Y_{t_j}, \tag{3.5.3}$$

$$k^E(Y) = \sum_{t_{i+1}\le T}\sum_{t_{j+1}\le T} k\left(\frac{i - j}{H_1}\right)\Delta Y_{t_i}\Delta Y_{t_j}, \tag{3.5.4}$$

$$k^S(Y) = \sum_{t_{i+1}\le T}\sum_{t_{j+1}\le T} k\left(\frac{t_i - t_j}{H_S}\right)\Delta Y_{t_i}\Delta Y_{t_j}, \tag{3.5.5}$$

Eq. (3.5.4) corresponds to the realized kernel discussed in [16] and [18]. It is equivalent to our version of the realized kernel in Eq. (3.5.2) under the case of equidistant observations, with $H_1 = HN/T$. Eq. (3.5.3) is the flat-top version of the estimator. According to the Theorem 3.4.1, under the case of the white noise error, it has a faster convergence rate. Eq. (3.5.5) is the realized kernel using Theorem 3.4.2 under the case of the shrinking noise. In the following simulation we always set $H = N^{-2/5}$ as in Theorem 3.3.3, $H_F = N^{-1/2}$ as in Theorem 3.4.1, $H_1 = HN/T$ as discussed in Section 3.2.1 and we parametrize $H_S = N^{-\beta}$ with $\beta = [0.6, 0.7, 0.8, 0.9, 1, \tilde{\beta}]$. We set $\tilde{\beta} = \frac{2+4\bar{\psi}}{5}$ as in Theorem 3.4.2, and the choice of $\bar{\psi}$ will be discussed later in Section 3.5.2.

Throughout the Section 3.5, we adopt the following sampling scheme.[3] We always set $t_0 = 0$, and t_{i+1} is the first time $t \geq t_i$ so that $X_t - X_{t_i}$ = either $n^{-1/2}p$ or $-n^{-1/2}q$, where p, $q > 0$. We have

$$X_{t_{i+1}} - X_{t_i} = n^{-1/2} Z_i,$$

where Z_i's are i.i.d. with mean zero (by X_t is a standard Itô process) and $P(Z = p) = q/(p+q)$, $P(Z = -q) = p/(p+q)$. In this simulation study, we will set $n = 3600$ and consider two different values of $[p, q]$, with $[p_1, q_1] = [0.5, 0.35]$ and $[p_2, q_2] = [1, 1.2]$. Here, we choose the values of $[p, q]$ arbitrarily. The first set of values corresponds to an average sampling frequency of five seconds. The second set of values corresponds to an average sampling frequency of thirty seconds.

3.5.2 The case of the white noise

In this section, we study the finite sample performance of the estimators in Eq. (3.5.2) – (3.5.5) under the case of white noise error. Throughout Section 3.5, we only model the noise at exactly the sampling times. We assume the efficient log-price follows Eq. (3.5.1), and the observed price Y process is,

$$Y_t = X_t + U_t, \tag{3.5.6}$$

where U is the microstructure noise that follows Assumption H3.5. We assume U_t's follow a Gaussian distribution $N(0, \eta^2)$, where $\eta = \xi\sqrt{IQ}$, where IQ is the simulated Integrated Quarticity and $\xi = [0, 0.01, 0.1, 1]$ is the noise-to-signal ratio. Regarding the choice of $\bar{\psi}$, we treat the microstructure noise as if it is a shrinking process with $N^{-\bar{\psi}} = \xi\sqrt{IQ}$. Note that by setting $\tilde{\beta} = \frac{2+4\bar{\psi}}{5}$, where $N^{-\bar{\psi}} = \xi\sqrt{IQ}$, we are essentially treating the noise as if the case of the shrinking noise, with $\epsilon^2 = 1$ in Theorem 3.4.2.

Table 3.1 shows the simulation results. Panel A of the table reports the sampling scheme with $[p, q] = [0.5, 0.35]$, which corresponds an average sampling rate of five seconds. Panel B of the table reports the sampling scheme with $[p, q] = [1, 1.2]$, which corresponds an average sampling rate of thirty seconds. Values in bold in each column of the panel are the three smallest values in that column. Under the case of the white noise error, a direct comparison between $k(Y)$, $k^F(Y)$ and $k^E(Y)$ in terms of MSE confirms that $k^F(Y)$ converges at a faster rate regardless of the noise-to-signal ratio. The performances of $k(Y)$ and $k^E(Y)$ under all specifications seem on the same level, indicating they converge at the same rate. Another important observation is that the MSE of $k(Y)$, $k^F(Y)$ and $k^E(Y)$ are robust to changes in the noise-to-signal ratio, indicating that they can be applied in a model with small/large signal ratio and the convergence rate are unaffected.

Another important finding through Table 3.1 is that under the case of a small noise-to-signal

[3]This endogenous sampling scheme is also discussed in [58].

Table 3.1 Simulation results under the case of white noise

Panel A: $[p, q] = [0.5, 0.35]$								
	$\xi = 0$		$\xi = 0.01$		$\xi = 0.1$		$\xi = 1$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
$k(Y)$	0.0475	0.0079	0.0475	0.0079	0.0475	0.0080	0.0485	0.0104
$k^F(Y)$	0.0215	0.0020	0.0215	0.0020	0.0215	0.0020	**0.0231**	0.0027
$k^E(Y)$	0.0475	0.0080	0.0475	0.0080	0.0475	0.0081	0.0480	0.0095
$k^S(Y)$, $\beta = 0.6$	0.0101	0.0033	0.0101	0.0033	0.0102	0.0040	**0.0165**	0.0515
$k^S(Y)$, $\beta = 0.7$	0.0045	0.0035	0.0045	0.0036	**0.0046**	0.0056	0.0825	0.2528
$k^S(Y)$, $\beta = 0.8$	**0.0019**	0.0005	0.0019	0.0007	**0.0022**	0.0135	1.9278	1.3424
$k^S(Y)$, $\beta = 0.9$	**0.0008**	-0.0004	**0.0008**	0.0003	0.0067	0.0723	52.9429	7.1847
$k^S(Y)$, $\beta = 1$	**0.0004**	-0.0006	**0.0004**	0.0030	0.1447	0.3756	1436.9	37.674
$k^S(Y)$, $\beta = \bar{\beta}$	–	–	**0.0008**	0.0003	**0.0051**	0.0052	**0.0313**	0.0131
Panel B: $[p, q] = [1, 1.2]$								
	$\xi = 0$		$\xi = 0.01$		$\xi = 0.1$		$\xi = 1$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
$k(Y)$	0.0803	0.0055	0.0803	0.0055	0.0803	0.0056	0.0818	0.0079
$k^F(Y)$	0.0500	0.0078	0.0499	0.0078	0.0497	0.0077	**0.0545**	0.0088
$k^E(Y)$	0.0805	0.0056	0.0805	0.0056	0.0804	0.0056	0.0811	0.0066
$k^S(Y)$, $\beta = 0.6$	0.0281	0.0058	0.0281	0.0057	0.0280	0.0063	**0.0368**	0.0373
$k^S(Y)$, $\beta = 0.7$	0.0147	0.0031	0.0147	0.0031	0.0148	0.0040	**0.0522**	0.1245
$k^S(Y)$, $\beta = 0.8$	**0.0081**	0.0028	0.0081	0.0028	**0.0086**	0.0074	0.2823	0.4514
$k^S(Y)$, $\beta = 0.9$	**0.0043**	0.0038	**0.0043**	0.0040	**0.0054**	0.0206	3.0447	1.6337
$k^S(Y)$, $\beta = 1$	**0.0022**	0.0010	**0.0022**	0.0016	**0.0071**	0.0602	39.6448	6.0750
$k^S(Y)$, $\beta = \bar{\beta}$	–	–	**0.0016**	0.0019	0.0105	0.0063	0.0580	0.0152

ratio ($\xi = 0, 0.01, 0.1$), it is beneficial to treat the noise as if the case of the shrinking noise. We set $H_S = N^{-\beta}$ with $\beta = [0.6, 0.7, 0.8, 0.9, 1, \tilde{\beta}]$ to experiment the optimum choice of β. As shown in the Table 3.1, under the case of no microstructure noise $\xi = 0$, the optimal choice is to set $H = O(N^{-1})$ as suggested in Theorem 3.2.1. As the noise-to-signal ratio increases, the optimal choice of β decreases from 1 to $0.8 - 0.9$, and when $\xi = 1$, the optimal choice of $\beta = 0.6$. This suggests that by treating the noise as if the case of the shrinking noise in Theorem 3.4.2, the realized kernel estimator performs better than the case of Theorem 3.3.3 and 3.4.1. A rule of thumb given through this simulation is by setting $\beta = \tilde{\beta}$. As shown in Table 3.1, such a choice improves the estimate consistently (in terms of both MSE and bias).

3.5.3 The case of the dependent noise

In this section, we assume the efficient log-price follows Eq. (3.5.1), and the observed price Y process is,

$$Y_t = X_t + U_t, \tag{3.5.7}$$

where U is an $AR(1)$ noise process with $U_{t_i} = 0.8U_{t_{i-1}} + V_{t_i}$, and we assume V_t's follow a Gaussian distribution $N(0, 0.36\eta^2)$. Here, the $AR(1)$ coefficient is chosen randomly. In order to make the result comparable with Table 3.1, we set $Var(V)$ so that $Var(U) = \eta^2$. The choice of the remaining parameters remain unchanged.

Table 3.2 shows the simulation results. Panel A of the table reports the sampling scheme with $[p, q] = [0.5, 0.35]$, which corresponds an average sampling rate of five seconds. Panel B of the table reports the sampling scheme with $[p, q] = [1, 1.2]$, which corresponds an average sampling rate of thirty seconds. Values in bold in each column of the panel are the three smallest values in that column. Under the case of the dependent noise, the performance of $k(Y)$ and $k^E(Y)$ is comparable to the case of the white noise error regardless the value of the noise-to-signal ratio, which indicates that the realized kernel estimator could be adopted in a model with the dependent noise and time endogeneity. This confirms the discussion in Section 3.4.2. Surprisingly, the MSE of $k^F(Y)$ is smaller than both $k(Y)$ and $k^E(Y)$. This seems at odds with the discussion in Section 3.4.1, which depends heavily on the white noise assumption. On the other hand, we find that as the noise-to-signal ratio increases, the MSE of $k^F(Y)$ increases as well, especially when the noise-to-signal ratio is large $\xi = 1$. This indicates that the smaller MSE of $k^F(Y)$ stems from the finite sample size and a small noise-to-signal ratio. As we discussed previously, treating the noise as if the case of the shrinking noise as in Theorem 3.4.2 allows a faster convergence rate. This point, discussed earlier in Section 3.5.2, has also been confirmed by the performance of $k^S(Y)$. As the noise-to-signal ratio increases from 0 to 1, the optimal value of β decrease from 1 to 0.4 and 0.5. By treating the noise as if the case of the shrinking noise and set $\tilde{\beta} = \frac{2+4\bar{\psi}}{5}$ (where $N^{-\bar{\psi}} = \xi\sqrt{IQ}$), we consistently achieve an efficiency gain over $k(Y)$, $k^F(Y)$ and $k^E(Y)$.

Table 3.2 Simulation results under the case of dependent noise

Panel A: $[p, q] = [0.5, 0.35]$								
	$\xi = 0$		$\xi = 0.01$		$\xi = 0.1$		$\xi = 1$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
$k(Y)$	0.0475	0.0079	0.0475	0.0080	0.0476	0.0081	0.0497	0.0187
$k^F(Y)$	0.0215	0.0020	0.0215	0.0020	0.0215	0.0025	**0.0356**	0.0566
$k^E(Y)$	0.0475	0.0080	0.0476	0.0080	0.0476	0.0081	**0.0491**	0.0173
$k^S(Y)$, $\beta = 0.6$	0.0101	0.0033	0.0101	0.0033	0.0105	0.0063	0.1245	0.2730
$k^S(Y)$, $\beta = 0.7$	0.0045	0.0035	0.0045	0.0038	**0.0054**	0.0191	2.2650	1.4102
$k^S(Y)$, $\beta = 0.8$	**0.0019**	0.0005	0.0019	0.0015	**0.0090**	0.0748	58.8484	7.4604
$k^S(Y)$, $\beta = 0.9$	**0.0008**	-0.0004	**0.0009**	0.0034	0.1356	0.3591	1292.1	35.4963
$k^S(Y)$, $\beta = 1$	**0.0004**	-0.0006	**0.0005**	0.0118	1.5499	1.2373	15290	123.0187
$k^S(Y)$, $\beta = \bar{\beta}$	–	–	**0.0008**	0.0036	**0.0058**	0.0158	**0.0371**	0.0378
Panel B: $[p, q] = [1, 1.2]$								
	$\xi = 0$		$\xi = 0.01$		$\xi = 0.1$		$\xi = 1$	
	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias
$k(Y)$	0.0803	0.0055	0.0804	0.0055	0.0805	0.0056	0.0880	0.0208
$k^F(Y)$	0.0500	0.0078	0.0500	0.0078	0.0504	0.0087	**0.0720**	0.0595
$k^E(Y)$	0.0805	0.0056	0.0805	0.0056	0.0806	0.0056	**0.0876**	0.0204
$k^S(Y)$, $\beta = 0.6$	0.0281	0.0058	0.0280	0.0058	0.0284	0.0078	0.1243	0.1983
$k^S(Y)$, $\beta = 0.7$	0.0147	0.0031	0.0147	0.0032	0.0155	0.0093	0.7178	0.6939
$k^S(Y)$, $\beta = 0.8$	**0.0081**	0.0028	0.0081	0.0031	**0.0106**	0.0254	7.3215	2.4401
$k^S(Y)$, $\beta = 0.9$	**0.0043**	0.0038	**0.0043**	0.0046	**0.0140**	0.0803	66.1667	7.6740
$k^S(Y)$,$\beta = 1$	**0.0022**	0.0010	**0.0023**	0.0031	0.0476	0.2003	419.0546	19.8848
$k^S(Y)$, $\beta = \bar{\beta}$	–	–	**0.0016**	0.0033	**0.0120**	0.0182	**0.0725**	0.0498

3.5.4 Constructing the confidence bounds

In this section, we aim to make the previous CLTs feasible for concrete empirical applications, such as calculating the p-value or constructing confidence intervals for the estimator. The main strategy in this section is the blocking method. In order to use the blocking method, we need to further impose that σ_t is almost surely continuous. More specifically, since all the CLTs given have a convergence rate of $H^{-1/2}$. If we parameterize $H = cN^{-\omega}$, we can set the size of each block $M_N = [N^{\phi}]$ with $(1-\omega/2) < \phi < 1$ to obtain consistent estimate of σ_t, $\Psi(t)$, a_t, b_t and $h(t)$ as in Section 4 in [29]. Here $h(t)$ is the time derivative of $H(t)$ defined in [64].

$$\begin{cases} \widehat{\sigma}_t^2 = \frac{k(Y)_{t_{iM_N}} - k(Y)_{t_{(i-1)M_N}}}{t_{iM_N} - t_{(i-1)M_N}} \\ \int_0^t \widehat{\Psi(s)}\, ds = H^{-1} \sum_{t_{i+1}\le t} \frac{\Delta t_i^2}{2} + H^{-1} \sum_{t_{i+1}\le t} \sum_{j\le i-1} k\left(\frac{t_i - t_j}{H}\right)^2 \Delta t_i \Delta t_j \\ \widehat{H(t)} = \frac{N}{T} \sum_{t_{i+1}\le t} \Delta t_i^2 \\ \int_0^t \widehat{a_s\sigma_s}\, ds = H^{-1/2} \sum_{t_{j+1}\le t} \sum_{t_{i+1}\le t} k\left(\frac{t_i-t_j}{H}\right) \Delta Y_{t_i} \Delta t_j \\ \int_0^t \widehat{b_s\sigma_s^2}\, ds = H^{-1} \sum_{t_{j+1}\le t} \left(\sum_{t_{i+1}\le t} k\left(\frac{t_i-t_j}{H}\right) \Delta Y_{t_i} \right)^2 \Delta t_j \end{cases}$$

The spot volatility can be consistently estimated because regardless of the presence of time endogeneity and the microstructure noise, the realized kernel estimator is always consistent at $H^{-1/2}$. Parameters a_t and b_t can be consistently estimated because our definitions of time endogeneity in Eq. (3.2.6) and (3.2.7) are naturally robust to the presence of the microstructure noise. It is straightforward to check that substituting Y for X in Eq. (3.2.6) and (3.2.7) will not deteriorate the consistency. However, we do notice that under the presence of the microstructure noise, the above blocking strategy requires a huge amount of observations each day and is unlikely for most illiquid financial assets.

In this section, we construct the confidence intervals of $k(Y)$ based on Theorem 3.3.3 using the simulated data as in Eq. (3.5.1) under the case of no noise as an illustration. We set $[p,q] = [0.5, 0.35]$, $\xi = 0$, $H^{-1} = \pi N/T$ and $M_N = N^{3/4}$. We choose $H = O(N^{-1})$ to achieve the fastest convergence rate as in Theorem 3.2.1, and the choice of $H^{-1} = \pi N/T$ corresponds to the Nyquist frequency of the frequency domain estimators. The confidence intervals of $k(Y)$ corrects the asymptotic bias due to time endogeneity. More precisely, these are the confidence intervals of $k(Y) - 2H^{1/2} \int_0^T \widehat{a_v\sigma_v}\, dX_v$, where $\int_0^T \widehat{a_v\sigma_v}\, dX_v$ is a consistent estimator of $\int_0^T a_v\sigma_v\, dX_v$.

Figure 3.2 reports the 68% (red) and 95% (blue) confidence bounds of the realized kernel estimators, the black line corresponds to the true value (Integrated Volatility). The upper panel plot the confidence bounds for the first 50 days out of the simulated data, the lower panel plot the first 100 days. Based on a total of 1000 days simulated data, the actual coverage frequency of 68% and 95% confidence bounds are 67.5% and 94.5%, sufficiently close to the intended value.

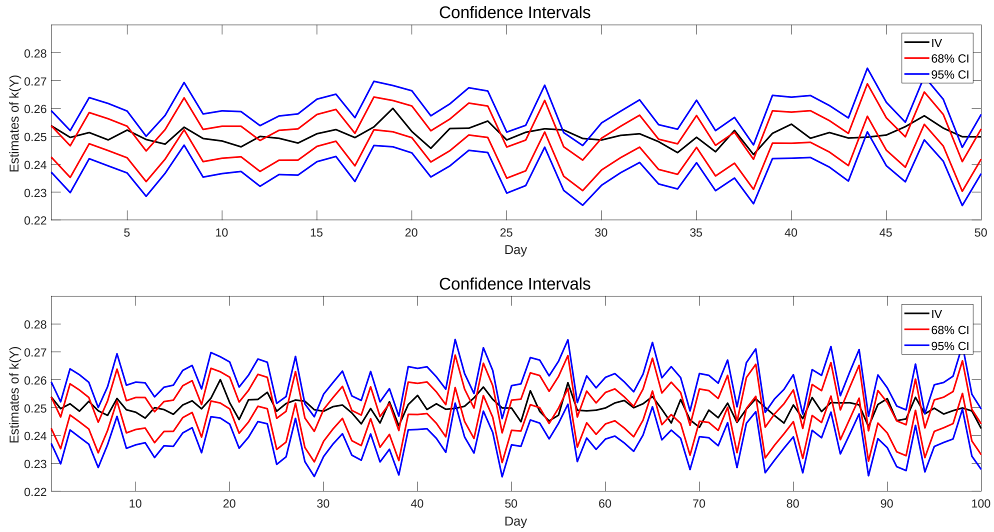

Figure 3.2 Confidence intervals. Upper panel: for 50 days; Lower panel: for 100 days.

Other than the blocking strategy, an alternative method to estimate σ_t consistently is by exploiting the link between the realized kernel estimator and the frequency domain estimators discussed in Section 3.2.1. In order to obtain a consistent estimate of the instantaneous (spot) volatility, we simply apply the Fourier inversion theorem to the frequency domain estimator. If one is willing to assume the modulus of continuity of σ_t, then (see [60]),

$$\lim_{N\to\infty} sup_t||\hat{\sigma}_t - \sigma_t||_2 = 0.$$

3.6 Empirical Analysis

In this section, we apply the realized kernel estimator with $k(t) = \frac{sin\, t}{t}$ to the high frequency trade data. We analyze a large liquid stock, International Business Machines Corporation (IBM) and follow the data cleaning procedure used in [17]. The period of analysis is from Jan 3rd, 2011 to Dec 30th, 2011. The total sample size is 6,123,671, corresponds to an average daily sample size of 24,300. We note that there is time endogeneity in the high frequency trading data, the time between observations may be informative about the price process, with trades occurring more or less frequently depending upon the news flow, which itself will generally influence the price process. In [58], they also provide empirical evidence through a test and document that time endogeneity is present in the financial data.

In the first case, we exhaust the available data and every transaction on each day will be used to construct volatility estimators. Since we use all the available data at ultra high frequency, the

Table 3.3 Summary statistics

	Min	Mean	Median	Max	AR	Variance
10-min RV	1.4184	10.8634	6.2698	90.5295	0.7050	162.3772
10-min $k(Y)$	1.4310	10.8825	6.2637	93.5002	0.6921	166.9817
RV full sample	5.3752	38.9611	20.2575	521.1126	0.8508	3531.8565
$k(Y)$ full sample	0.1153	10.7094	5.7604	141.6061	0.4576	223.2856

The table reports summary statistics for the different estimators. The column labeled AR reports the standard first order autocorrelation coefficients.

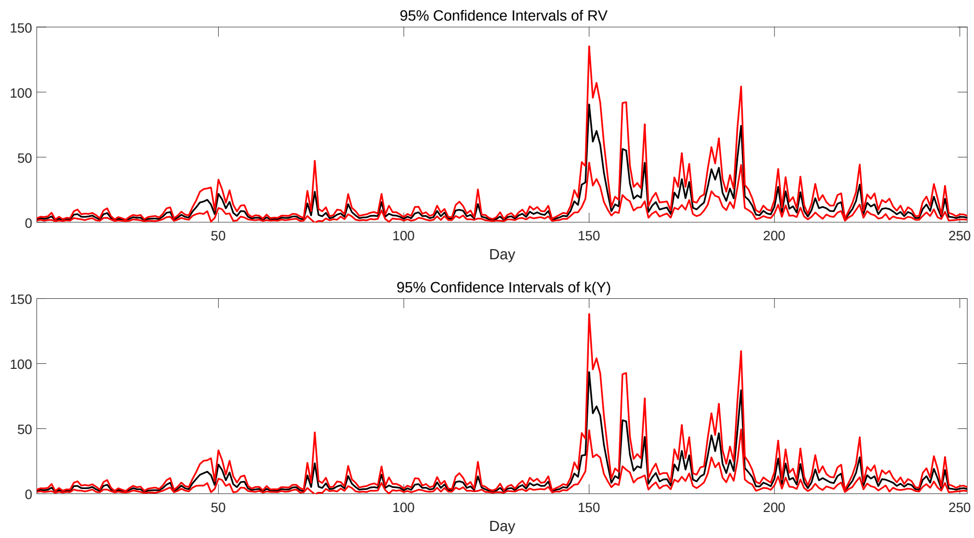

Figure 3.3 Confidence intervals. Upper panel: RV; Lower panel: $k(Y)$.

effect of the microstucture noise and time endogeneity can not be ignored. In the second case, we sample on an evenly spaced grid of 10 min. As suggested in Figure 3.1, under this sparse sampling scheme, the effect of the microstructure noise is small and could be safely ignored. We set $H^{-1} = N^{2/5}$ following Theorem 3.3.3 under the first case, and for the second case we set $H^{-1} = \pi N/T$ corresponds to the Nyquist frequency following Theorem 3.2.1.

Table 3.3 reports the summary statistics of the estimators considered: 10 min Realized Volatility, 10 min $k(Y)$ and RV, $k(Y)$ using the highest frequency trading data. When using the highest frequency trading data, we take the previously ignored border effects into consideration, so that $k(Y)$ reported here are unbiased. Under the case of 10 min data, as illustrated in Table 3.3, the performances of RV and $k(Y)$ are almost identical. This indicates that under the case of equidistant observations, $k(Y)$ could achieve the same efficiency level as RV. Figure 3.3 reports the confidences interval of RV and $k(Y)$ under this case. Under the case of equidistant sampling, the choice of the realized kernel $k(\cdot)$ is crucial and has been discussed extensively in [16]. When

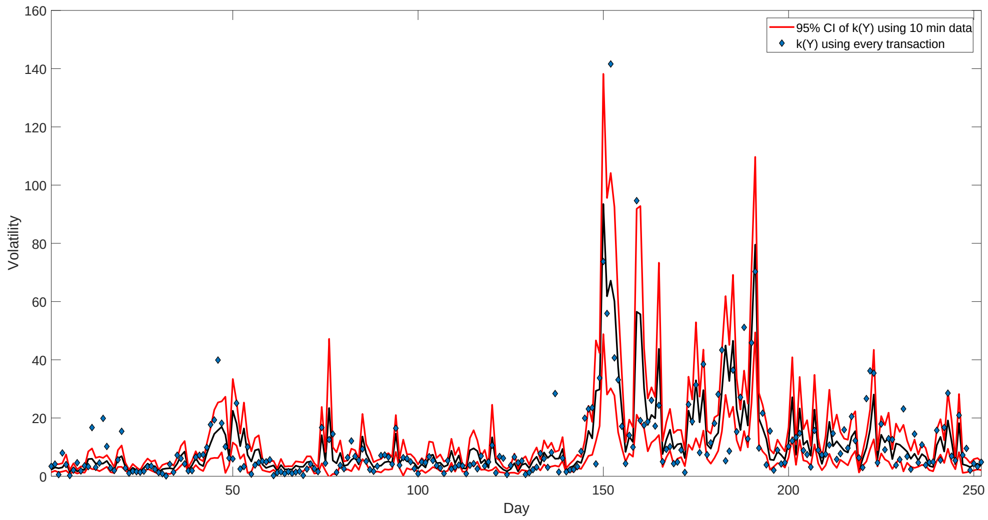

Figure 3.4 95% CI of $k(Y)$ based on 10 min data and $k(Y)$ using every transaction.

$k(t) = \frac{sin\,t}{t}$, the choice of H equals to the Nyquist frequency equates the asymptotic variance of $k(Y)$ to that of RV.

On the other hand, under the case of full sample size, the performances of RV and $k(Y)$ are no longer comparable. First of all, when using all the transaction data on each day, the results regarding RV and $k(Y)$ are more speculative than the case of the 10 min data inevitably. In [12], they suggests a consistent estimator of $\hat{\epsilon}^2 = RV/(\text{number of observations})$. As expected, $k(Y)$ performs very well when using all the available data. The increase in MSE is moderate and can be explained by the presence of the microstructure noise. As shown in Figure 3.4, for most of the days, the full sample $k(Y)$ are in the 95% confidence intervals constructed from the 10 min data. On a few days the estimates $k(Y)$ based on 10 min data and full sample are almost identical, indicating that the realized kernel estimator is robust to both time endogeneity and the presence of microstructure noise.

3.7 Conclusions

In this study, we have provided a detailed analysis of the performance of the realized kernel estimator of the Integrated Volatility under time endogeneity and the presence of market microstructure noise due to frictions. We derive central limit theorems for the realized kernel estimators in a nonparametric setting. Our framework contains various assumptions about the kernel weight and a general endogenous setting. We further show that the realized kernel estimator can be made to converge at the fastest possible rate, robust to different specifications of the noise and robust to dynamics in the noise process. Evidence from Monte Carlo experiments confirmed our theoretical results. We have also applied our methodology to the millisecond time stamped trade data and illustrated that the realized kernel estimator could be adopted in a model that are sampled at the highest possible frequency.

There are many possible extensions to this paper. A very promising extension that will be considered in the future paper is inferencing for multiple assets. There are two challenges. The first is how to model and capture time endogeneity in a multivariate setting, and the second challenge is the effect of the asynchronous trading. Another possible extension is a consistent estimate of instantaneous (spot) volatility. We gave two consistent estimators that are robust to microstructure noise and time endogeneity in Section 3.5.4. The difficulty arises from deriving a stable convergence in law results. The third possible extension is to study the most efficient endogenous sampling scheme. As the asymptotic variance of the limiting process depends on a_t, b_t, $\Psi(t)$ and $H(t)$, an interesting question is to design the most efficient sampling scheme and choose the realized kernel $k(\cdot)$ to minimize asymptotic variance.

Chapter 4

Sign Restrictions and Parameterizations of SVARs

Sign restriction has become part of the mainstream of empirical macroeconomics for analyzing the dynamics caused by macroeconomic shocks. The main contribution of this paper is the study of four different parameterizations of SVARs and making probability statements across different parameterizations. We also address the problem that in a sign restricted SVAR model, the data provide no information to distinguish between any two pairs of structural parameters in the admissible set. We propose two ways to solve this problem: combining sign restrictions with exclusive zero restrictions, and constructing a joint credible set that is robust to different prior specifications based on our main result. We then demonstrate our approach with an empirical example of identifying monetary policy shock.

4.1 Introduction

The structural vector autoregressive models (SVAR) has been the workhorse for analyzing the dynamics caused by macroeconomic shocks. More often than not, a vector autoregressive models (VAR) is first fitted to the data, and then structural equation parameters are recovered. However, a reduced-form vector autoregressive model can not pin down a unique structural vector autoregressive model. Identification restrictions are needed so that the parameters in the structural equations can be recovered.

Much of the literature focused on putting parametric restrictions on the short-run ([72] and [20]) and long-run ([23]) impulse responses or structural equation parameters to achieve identification. Later, it was proposed that inference using the structural vector autoregressive model might be based on prior beliefs about the signs of the impacts of certain shocks ([34], [25] and [75]). This method is commonly referred to as sign restrictions and it has become part of the mainstream of empirical macroeconomics. However, the prior beliefs about the signs of the impacts of certain shock are generally not sufficient to identify a structural vector autoregressive model, and this results in an admissible set of possible inferences. Even though there is a large literature that proposes frequentist methods for conducting inference in SVARs model, most applied macroeconomic studies work in the Bayesian paradigm. Different Bayesian methods of implementing sign restrictions have been proposed. The "agnostic" prior approach first proposed by [75] and [71] works with a reduced-form vector autoregressive model and then transforms the reduced-form parameters to structural parameters; [19] argue that one should work on the structural vector autoregressive model directly so that researchers can explicitly acknowledge and defend the role of prior beliefs in influencing structural conclusions.

In Section 4.2 and 4.3, we consider four different parameterizations of an SVAR: reduced-form parameterization, structural parameterization, reduced-form MA parameterization and IRFs (structural MA) parameterization. We focus on making probability statement about all four different parameterizations using the change of variables formula based on the theory of exterior product. We intend to answer the following questions: if we estimate a reduced-form sign restricted VAR model, what is the corresponding posterior distribution of the impulse response and the posterior distribution of the structural parameters. Instead of arguing which parameterization researchers should work with, we focus on building a bridge between different parameterizations so that researchers could work with the parameterization that is computationally efficient. The main contribution in Section 4.3 is the calculation of the determinants of the Jacobian matrices between different parameterizations. Our result is based on the theory of exterior product[1]. Specifically, we follow the analysis of [51] on the orthogonal group and use the invariant measure throughout the paper[2].

Since a sign restricted structural vector autoregressive model is generally set identified, the data provide no information to distinguish between any two pairs of structural parameters in the admissible set. Specifically, for a set identified SVAR the prior for the reduced-form parameters is updated, whereas the prior associated with the orthogonal matrix is never updated. [19] show that inference in a sign restricted SVAR model could be solely driven by the choice of priors. In Section 4.4, we address this problem. One possible way to tackle this problem is to shrink the admissible set by imposing more restrictions, such as combining sign restrictions with zero restrictions. We suggest treating zero restrictions as soft restrictions in the form of prior information and developing an efficient importance sampler to sample from the posterior distribution. The advantage of treating zero restrictions as prior information is twofold. Firstly,

[1] The determinant of the Jacobian matrix between the reduced-form parameterization and the structural parameterization has been derived in [9] based on the volume measure. Our derivation (in Appendix C.2) is of independent econometric interest. It is novel and simpler.

[2] This turns out to be favorable and leads to analytically tractable expressions. in [48], they work on a similar problem, but use the Lebesgue measure over the orthogonal group.

this is a strict generalization of the traditional hard zero restrictions. The variance of the prior distribution can be used to control the confidence one has of this zero restrictions. Secondly, given sign restrictions are non-degenerate[3], treating zero restrictions in the form of prior information will not shrink the admissible set to a set of measure zero. Due to the admissible set is of positive measure, most commonly used rejection sampler could be adopted.

Another way to tackle the problem is to construct a joint credible set which is robust to different choices of priors. More often than not, the point-wise quantiles of the impulse responses of the admissible set are reported. However, this proves to be problematic for two reasons: firstly, it is commonly known that quantiles of the impulse responses have no economic interpretation due to the fact that they are generally associated with different structural models; secondly, quantiles of the impulse responses might not be a valid measure of the central tendency[4]. In this paper, we construct a robust joint credible set based on the posterior density of impulse responses. According to our Theorem 4.3.4, given the prior distribution of the orthogonal matrix is uniformly distributed (the Haar measure), the joint posterior distribution of the impulse responses is independent of the orthogonal matrix and flat over the orthogonal group. Then the robust credible set could be constructed as the highest posterior region associated with a flat prior over the orthogonal group. The key difference between this paper and [19] is that instead of focusing on the marginal distribution of the structural parameters of particular interest, we focus on the joint distribution. Constructing the credible set based on the joint distribution takes the dependence of structural parameters across different horizons and across structural equations into consideration. We further demonstrate this using the model of US monetary policy proposed by [75].

The structure of this paper is as follows. In Section 4.2, we describe four parameterizations of SVARs and the identification. In Section 4.3, we provide a detailed definition of the Haar measure on the orthogonal group, i.e., a flat prior and then present our main result. In Section 4.4, we illustrate two applications of our main result and focus on solving the problem associated with sign restrictions. In Section 4.5, we demonstrate our method with an empirical example. In Section 4.6, we conclude.

[3]Non-degenerate here refers to the case that the set of parameters satisfying the sign restrictions are of positive Lebesgue measure.

[4]See [47] for more discussion.

4.2 The Framework

4.2.1 Four Parameterizations of an SVAR

Consider an SVAR(p) model

$$A_0 y_t = \sum_{j=1}^{p} A_j y_{t-j} + \epsilon_t \tag{4.2.1}$$

where y_t is an $n \times 1$ vector, ϵ_t is an $n \times 1$ vector white noise process, normally distributed with mean zero and variance I_n. In the structural model, $(A_0, A_1, \ldots, A_p)$ are the parameters, and A_0 is invertible.

Left-multiplying Eq. (4.2.1) by the inverse of A_0, we obtain the reduced-form VAR representation of the model,

$$y_t = \sum_{j=1}^{p} A_0^{-1} A_j y_{t-j} + A_0^{-1} \epsilon_t$$

or for simplicity, this can be written as

$$y_t = \sum_{j=1}^{p} B_j y_{t-j} + \eta_t \tag{4.2.2}$$

where $B_j = A_0^{-1} A_j$, $\eta_t = A_0^{-1} \epsilon_t$, $\Sigma = \mathrm{E}(\eta_t \eta_t') = A_0^{-1}(A_0^{-1})'$. Here Σ denotes the covariance matrix of the reduced-form shocks. The parameters of the reduced-form model are $(B_1, \ldots, B_p, \Sigma)$, where Σ is a symmetric and positive definite matrix. We restrict the domain $(B_1, \ldots, B_p, \Sigma)$ to the set such that the reduced-form VAR(p) model can be mapped into a VMA(∞) model.

$$y_t = \sum_{j=0}^{\infty} \Psi_j \eta_{t-j} \tag{4.2.3}$$

where $\Psi_s = B_1 \Psi_{s-1} + B_2 \Psi_{s-2} + \ldots + B_p \Psi_{s-p}$, with $\Psi_0 = I_n$ and $\Psi_s = \mathbf{0}$ for $s < 0$. The parameters of the VMA representation are $(\Psi_1, \ldots, \Psi_p, \Sigma)$.

The h-th horizon impulse response matrix is denoted by the n $\times$ n matrix IRF^h, $h = 0$, 1, 2, ... IRF^h_{ij} denotes the (i, j) element of IRF^h and represents the effect on the i-th variable in y_{t+h} of a unit shock to the j-th element of ϵ_t. Since $\eta_t = A_0^{-1} \epsilon_t$, we can rewrite the VMA(∞) representation as

$$y_t = \sum_{j=0}^{\infty} \Psi_j A_0^{-1} \epsilon_{t-j} \tag{4.2.4}$$

The impulse response can be calculated as

$$\begin{aligned} IRF^h &= \Psi_h A_0^{-1} \\ IRF^\infty &= (A_0 - \sum_{j=1}^{p} A_j)^{-1} = (I_n - \sum_{j=1}^{p} B_j)^{-1} A_0^{-1} \end{aligned}$$

Eq. (4.2.1), (4.2.2), (4.2.3), (4.2.4) represent the four different parameterizations of an SVAR that we study in this paper.

4.2.2 Identification

Using Eq. (4.2.1) and (4.2.2), for a particular $(B_1, \ldots, B_p, \Sigma)$, we can find more than one $(A_0, A_1, \ldots, A_p)$, such that $B_j = A_0^{-1}A_j$ and $\Sigma = A_0^{-1}(A_0^{-1})'$. The observational equivalence can be expressed as finding an orthogonal matrix P, such that $(PA_0, PA_1, \ldots, PA_p)$ is an observationally equivalent point. Denote Σ_{ch} as the unique Cholesky decomposition of Σ, where $\Sigma = \Sigma_{ch}\Sigma_{ch}'$. Since any A_0 of the form $A_0 = P\Sigma_{ch}^{-1}$ satisfies $\Sigma = (A_0'A_0)^{-1}$, in the absence of any identification restrictions $\{A_0 = P\Sigma_{ch}^{-1} : \mathrm{P} \in O(n)\}$ forms the set of A_0 that are consistent with a specific set of reduced-form parameters. Here $O(n)$ denotes the set of $n \times n$ orthonormal matrices and is referred as the orthogonal group.

As a result, the complete reduced-form parameterization should be $(B_1, \ldots, B_p, \Sigma, P)$. Once the orthogonal matrix P is identified, we can pin down a unique $A_0 = P\Sigma_{ch}^{-1}$ and $A_j = A_0B_j$. Conversely, given a unique $(A_0, A_1, \ldots, A_p)$, we can pin down a unique $\Sigma = A_0^{-1}(A_0^{-1})'$, $B_j = A_0^{-1}A_j$ and $\mathrm{P} = A_0\Sigma_{ch}$. This shows that the mapping between the reduced-form parameterization $(B_1, \ldots, B_p, \Sigma, P)$ and the structural parameterization $(A_0, A_1, \ldots, A_p)$ is bijective.

We assume the reduced-form VAR(p) can be mapped into a VMA(∞) model as Eq. (4.2.3). As a result, the mapping between $(B_1, \ldots, B_p, \Sigma)$ and $(\Psi_1, \ldots, \Psi_p, \Sigma)$ is bijective[5]. So the mapping between the complete reduced-form parameterization $(B_1, \ldots, B_p, \Sigma, P)$ and the complete reduced-form MA parameterization $(\Psi_1, \ldots, \Psi_p, \Sigma, P)$ is also bijective.

Lastly, the mapping between the reduced-form MA parameterization and the IRF parameterization is characterized by $IRF^h = \Psi_h A_0^{-1} = \Psi_h \Sigma_{ch} P'$. Conversely, given $(IRF^0, \ldots, IRF^p)$, $\Psi_h = IRF^h A_0 = IRF^h(IRF^0)^{-1}$. This shows that the mapping between the reduced-form MA parameterization $(\Psi_1, \ldots, \Psi_p, \Sigma, P)$ and the IRF parameterization $(IRF^0, IRF^1, \ldots, IRF^p)$ is bijective.

All arguments above indicate that the identification of an SVAR model boils down to the identification of the orthogonal matrix P. With sufficient restrictions placed on P, i.e., P is identified, we would have bijective mappings between any two parameterizations.

[5] See [40], P.260, for example.

4.2.3 Sign Restrictions

Since there could be infinitely many structural parameters corresponding to one set of reduced-form parameters, different methods were proposed to achieve the identification of the Structural VAR model. The approach of using the prior beliefs about the signs of the impacts of certain shock was pioneered by [34], [25] and [75].

For example, a prior belief about the sign of the response of i-th variable to j-th shock in period h being negative could be formulated as $-e_i' IRF^h e_j > 0$, where e_i represents i-th column of the identity matrix I_n. Similarly, a prior belief about the sign of (i, j)-th entry of the impact matrix A_0 being positive could be formulated as $e_i' A_0 e_j > 0$.

Given mappings between any two parameterizations are bijective, sign restrictions could be transformed across different parameterizations as well. On the other hand, inequality constraints are generally weaker than equality constraints, which indicates that given the sign restrictions are non-degenerate, the set of parameters satisfying the restrictions will be of positive Lebesgue measure. For this reason, most sign restricted SVARs are estimated based on a rejection sampler, i.e., only accepting draws that the sign restrictions hold.

4.3 Main result

The main goal of this section is to make probability statement about posterior density of all four parameterizations of a sign restricted SVAR model. We intend to answer the following questions: if we estimate a reduced-form sign restricted VAR model, what is the corresponding posterior distribution of the impulse responses and the posterior distribution of the structural parameters. To do this, we use the change of variables formula based on the exterior differential product.

4.3.1 Invariant Measure

We employ the theory of exterior product extensively in this section, a minimum background is given in Appendix C.1. Before presenting the result, it is important to note that, throughout this paper, probability density functions on the orthogonal group are over the invariant measure, probability density functions on the spaces of $n \times n$ random matrices or symmetric matrices are over the Lebesgue measure.

The reason that we use the invariant measure instead of the Lebesgue measure is threefold. Firstly, using the invariant measure leads to an analytically tractable expression, which is computationally efficient. Secondly, the invariant measure or the invariant exterior differential form is well developed in the statistics literature, and it is the standard way to apply the change of variables formula in multivariate statistical analysis. Last but not least, with the invariant measure, the commonly used "agnostic" prior for the orthogonal matrix is a uniform distribution, while in terms of the Lebesgue measure, it is no longer uniform. For this reason, we suspect results in the literature where the "agnostic" prior is treated as a uniform distribution over the Lebesgue measure.

The following proposition formally defines the commonly used "agnostic" prior distribution for the orthogonal matrix P.

Proposition 4.3.1. *The uniform distribution (Haar measure) on the orthogonal group O(n) is defined as*

$$\mu(A) = \int_A \frac{\Gamma_n(\frac{1}{2}n)}{2^n \pi^{n^2/2}} (P'\, dP) \tag{4.3.1}$$

where $P \in O(n)$, and $(P'\, dP) \equiv \bigwedge_{i<j}^{n} p_j'\, dp_i$, p_i is the i-th column of P.

Proof. See Appendix C.2. $\square$

To better understand this invariant measure, from $PP' = I_n$, taking total differentiation on both side, we obtain $P'\, dP + dP'\, P = 0$ or equivalently $P'\, dP$ = -$(P'\, dP)'$, which indicates that $P'\, dP$ is skew-symmetric. Consider the case that n = 2, and

$$P = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} = \begin{bmatrix} p_1 & p_2 \end{bmatrix}$$

we have

$$(P'\, dP) = p_2'\, dp_1 = \begin{pmatrix} -sin\theta & cos\theta \end{pmatrix} \begin{pmatrix} -sin\theta\, d\theta \\ cos\theta\, d\theta \end{pmatrix} = d\theta$$

If researchers are interested in uniformly sampling on a unit circle, one would uniformly draw from the polar coordinate (the invariant measure) instead of from the real coordinate(the Lebesgue measure).

To make random draws from the "agnostic" prior Eq. (4.3.1), we refer to the Proposition 4.3.3 below. We show that the QR decomposition of a spherically distributed matrix is uniformly distributed (has Haar measure) over $O(n)$.

Definition 4.3.2. A $n \times 1$ random vector X is said to have a spherical distribution if X and PX have the same distribution for all $P \in O(n)$.

If X is spherically distributed, then the above definition indicates that the density function depends on X only through the value $X'X$. IF X is n-rowed and m-columned, and each element X_{ij} follows an independent standard normal distribution, then the density function of X could be written as

$$
\begin{aligned}
& (2\pi)^{-mn/2} exp\{-\frac{1}{2} vec(X)' vec(X)\} \\
= \quad & (2\pi)^{-mn/2} etr\{-\frac{1}{2} X'X\}
\end{aligned} \tag{4.3.2}
$$

Clearly X is spherically distributed[6]. Another example is the multivariate t distribution. If a vector X follows multivariate t distribution, X also belongs to the family of spherical distribution.

Proposition 4.3.3. *If X is n-rowed and n-columned, spherically distributed, then Q defined uniquely by $X = QR'$, $Q'Q = I_n$, $R_{ii} > 0$, $i = 1, \ldots, n$, and $R_{ij} = 0$, $i < j$ is uniformly distributed on $O(n)$.*

Proof. See Appendix C.2. □

4.3.2 Change of Variables Formula

The most common change of variables formula is

$$\int_V f(v) dv = \int_U (f \circ g)(u) \mid det J_g(u) \mid du$$

where V, U are sets in $\mathbb{R}^n$, g is function: $U \to V$ that is well behaved, differentiable and bijective, and f is integrable on V. du here denote the differential form $\mid du_1 \wedge du_2 \wedge \cdots \wedge du_n \mid$, where $\wedge$ is the wedge product, J_g is the Jacobian matrix of g.

To calculate the posterior density of a different parameterization, we need to calculate $\mid det J_g(u) \mid$ between any two parameterizations, which is given in the following theorem.

[6]This example along with Proposition 4.3.3, imply Theorem 3.2 in [73], the commonly used algorithm for generating random orthogonal matrix.

Theorem 4.3.4.

$$det(\frac{\partial \Psi_1, \ldots, \Psi_p, \Sigma, P}{\partial B_1, \ldots, B_p, \Sigma, P}) = 1 \tag{4.3.3}$$

$$det(\frac{\partial A_0, A_1, \ldots, A_p}{\partial B_1, \ldots, B_p, \Sigma, P}) = 2^{-n} \mid \Sigma \mid^{-\frac{np+2n+1}{2}} \tag{4.3.4}$$

$$det(\frac{\partial IRF^0, IRF^1, \ldots, IRF^p}{\partial \Psi_1, \ldots, \Psi_p, \Sigma, P}) = 2^{-n} \mid \Sigma \mid^{\frac{np-1}{2}} \tag{4.3.5}$$

$$det(\frac{\partial IRF^0, IRF^1, \ldots, IRF^p}{\partial B_1, \ldots, B_p, \Sigma, P}) = 2^{-n} \mid \Sigma \mid^{\frac{np-1}{2}} \tag{4.3.6}$$

$$det(\frac{\partial A_0, A_1, \ldots, A_p}{\partial \Psi_1, \ldots, \Psi_p, \Sigma, P}) = 2^{-n} \mid \Sigma \mid^{-\frac{np+2n+1}{2}} \tag{4.3.7}$$

$$det(\frac{\partial A_0, A_1, \ldots, A_p}{\partial IRF^0, IRF^1, \ldots, IRF^p}) = \mid \Sigma \mid^{-(np+n)} \tag{4.3.8}$$

Proof. See Appendix C.2. □

Theorem 4.3.4 provides a way to calculate the posterior density of any parameterization. If one estimates a reduced-form sign restricted VAR, Eq. (4.3.4) could be used to calculate the posterior density of the structural parameterization, Eq. (4.3.6) could be used to calculate the posterior density of the impulse response parameterization. Similarly, if the estimation is implemented directly on the structural parameterization, we can use Eq. (4.3.8) to calculate the posterior density of the impulse response parameterization.

Theorem 4.3.4 could also be applied to transform prior to any parameterization that is computational efficient. Prior is commonly placed on the structural parameters that has economic meaning, whereas the estimation is often implemented on the reduced-form parameterization. With Eq. (4.3.4) we can solve this problem, the prior can be transformed onto reduced-form parameters.

Theorem 4.3.4[7] can be easily generalized to a partially identified sign restricted SVAR model. The main difference between a partially identified SVAR and a SVAR model considered in this paper is the parameter space. When it comes to a partially identified SVAR, only a subset of the structural shocks are of particular interest and thus identified. Suppose only k $(0 < k \leq n)$ structural shocks are of particular interest, then instead of drawing all columns of P, one only needs to draw the first k columns of P. As a result, the parameter space is no longer $O(n)$ but the Stiefel manifold $V_m(n)$, the set of all orthonormal m-frames in $\mathbb{R}^n$ with $\dim(V_m(n)) = mn$ - $\frac{1}{2}m(m+1)$.

[7] As well as the Propositions and Lemmas given in Appendix C.1.

4.4 Applications of the Main Result

It is commonly known that the problem of adopting a sign restricted SVAR model is that the data provide no information to distinguish between any two pairs of structural parameters in the admissible set. In this section, we tackle this problem using two applications of Theorem 4.3.4. In the first application, we try to combine sign restrictions with more restrictive restrictions, such as short-run and long-run exclusive restrictions. In the second application, we solve the same problem from a different perspective by constructing a credible set based on the posterior density that is robust to the prior specification of the orthogonal matrix.

4.4.1 Application 1: Combine Sign Restrictions and Zero Restrictions

As mentioned above, since the data is not informative about the orthogonal matrix P, its prior is never updated. A growing literature is dedicated to make inferences from an SVAR that is set identified using sign restrictions and zero restrictions.

Having the same view as in [19], instead of treating short-run and long-run exclusive restrictions as hard restrictions, we suggest treating zero restrictions as prior information. Instead of restricting parameter a_i to be zero, we suggest putting a distribution on $a_i \sim N(\bar{a}_i, \sigma_i^2)$, where the hard restriction is $a_i = \bar{a}_i$, and σ_i represents the confidence one has in this restriction. A smaller σ_i close to 0 indicates one is confident about the restriction, whereas $\sigma_i \to \infty$ indicates one is skeptical about such a restriction.

Importance sampling could be used to make random draws from the posterior combined with Theorem 4.3.4, where the proposal density could be chosen as the posterior density obtained without such additional prior information[8].

Suppose we would like to estimate a purely sign restricted SVAR model, for illustration purposes, we put a flat prior[9] on the structural parameterization such that $\pi_{A_0,A_1,\ldots,A_p}(\cdot) \propto 1$. As indicated by Theorem 4.3.4, this is equivalent to a prior on the reduced-form parameterization such that $\pi_{B_1,\ldots,B_p,\Sigma,P}(\cdot) \propto |\, \Sigma \,|^{-\frac{np+2n+1}{2}}$, which takes the form of Normal-inverse-Wishart-Haar conjugate prior[10]. Aside from the sign restrictions we imposed, if we would like to further impose zero restrictions that $[A_0]_{ij} = 0$, we treat this as additional prior information $[A_0]_{ij} \sim N(0, \sigma_i^2)$. The Algorithm is formulated as follows:

Algorithm 4.4.1. *(step 1) Draw reduced-form parameter* $(B_1, \ldots, B_p, \Sigma)$ *jointly from the Normal-inverse-Wishart posterior distribution where the prior is* $\pi_{B_1,\ldots,B_p,\Sigma}(\cdot) \propto |\, \Sigma \,|^{-\frac{np+2n+1}{2}}$.

[8]Generally, such a proposal density is the posterior density of a conjugate prior and is computationally efficient to sample from.

[9]A flat prior here reflects the fact that we do not have any information available for identification, and is considered to be an objective prior.

[10]The prior distribution only depends on Σ, indicating that the prior for P follows Eq. (4.3.1).

(step 2) Draw the orthogonal matrix P using Proposition 4.3.3.

(step 3) If $(B_1, \ldots, B_p, \Sigma, P)$ satisfy sign restrictions, retain this P. Otherwise, repeat step 2 at most N times. If none of N numbers of draws of P satisfies sign restrictions, return to step 1.

(step 4) For each success draw, calculate its importance weight as $\phi([A_0]_{ij}, 0, \sigma_i^2)$, where $\phi(\cdot)$ represents the normal density function.

4.4.2 Application 2: Robust Joint Credible Set of Impulse Responses

An unresolved question is how to represent the results of using pure sign restrictions when the set of admissible models includes conflicting interpretations. The most common approach is to report the vector of point-wise posterior quantiles of the impulse responses. This approach has two shortcomings. Firstly, as [36] and [54] argue, the quantile of vectors is not equal to the vector of the quantiles. As a result, the vector of point-wise posterior quantiles of the impulse responses has no economic interpretation, due to they are most likely from different structural models. Secondly, as mentioned by [47], it is well established in the statistics literature that the median response functions are not a valid measure of the central tendency of the admissible set of impulse response function.

Using Theorem 4.3.4, we can make probability statements about impulse responses, i.e., one could specify a particular loss function and minimize the corresponding risk function. In this subsection, we construct a joint credible set such that it should be robust to the choice of the prior associated with the orthogonal matrix that the data is not informative about.

The construction of the joint credible set of the impulse responses is based on Theorem 4.3.4 and note that none of the determinant of the Jacobian matrix depends on the orthogonal matrix P. Thus, under pure sign restrictions, if the prior of the orthogonal matrix is the same as in Proposition 4.3.1, i.e., a flat prior, since it is never updated, the posterior will be flat over $O(n)$ as well. As a result, the credible set constructed as the highest posterior density set for all parameterizations will be independent of P and robust to different choices of prior. Such a credible set was first proposed by [47] and [48] by working with the Lebesgue measure over the orthogonal group[11].

Another reason that the highest posterior density set under the flat prior is robust is that the constructed credible set could be viewed as a generalization of the robust Bayesian approach proposed by [38], which considers multiple priors for the orthogonal matrix. The difference is that our robust credible set allows for joint inference and the construction of the most likely model. We denote $l_\alpha^h(HPD)$ and $u_\alpha^h(HPD)$ as the lower and upper bound of the interested impulse response in period h of the $100\alpha\%$ highest posterior region. Given the prior for the

[11] They focus on identifying the most likely admissible model that satisfies the sign restrictions. However, based on our Theorem 4.3.4 the highest posterior density set can not pin done a single most likely model due to its independence of P.

orthogonal matrix is flat, $\pi(P) \propto 1$, the posterior density of impulse responses according to Theorem 4.3.4 is

$$\pi_{IRF^0,IRF^1,\ldots,IRF^p|Y,X}(\cdot) \propto \mid \Sigma \mid^{-\frac{np-1}{2}} \pi_{(B_1,\ldots,B_p,\Sigma|Y,X}(\cdot)$$

Note that it is independent of orthogonal matrix P. As a result, the $100\alpha\%$ highest posterior region can only pin down a set

$$HPD_\alpha = \{(B_1, \ldots, B_p, \Sigma) : \pi_{IRF^0,IRF^1,\ldots,IRF^p|Y,X}(\cdot) > c_\alpha\}$$

where c_α is the largest constant such that $P(HPD_\alpha) \geq 1$ - α, and we have

$$\begin{aligned} l^h_\alpha(HPD) &= inf\{l^h : (B_1, \ldots, B_p, \Sigma) \in HPD_\alpha\} \\ u^h_\alpha(HPD) &= sup\{u^h : (B_1, \ldots, B_p, \Sigma) \in HPD_\alpha\} \end{aligned}$$

Under posterior convergence, $[l^h_\alpha(HPD), u^h_\alpha(HPD)]$ converges to the same probability limit as the credible set constructed using the robust Bayesian approach proposed by [38][12]. We will further illustrate this point with an empirical example.

4.5 An Empirical Example: Identifying Monetary Policy Shock

In this section, we illustrate our method with the model of US monetary policy proposed by [75]. We follow the same setup and fit a VAR with 12 lags without intercept or a time trend. We use the same dataset as [75] that consists of interpolated monthly real GDP, the GDP deflator, a commodity price, total reserves, non-borrowed reserves and the federal funds rate from January 1965 to December 2003[13]. The VAR(12) is fitted in levels of the logs of the series except the federal funds rate.

4.5.1 Pure Sign Restrictions

In this subsection, we implement a pure sign restricted structural VAR. To identify the effects of monetary policy shock, we impose a contractionary monetary policy shock does not lead to increase in prices, increase in non-borrowed reserves, or decrease in the federal funds rate at horizons 0 - 5[14].

[12] The convergence of credible set constructed using the robust Bayesian approach is discussed in Proposition 6.1 in [38].

[13] See a more detailed description of the data in [75].

[14] These are the same sign restrictions imposed by [75]. A contraction monetary policy shock here does not necessary mean it will leads to a contractioinary effect on GDP, but rather a monetary tightening shock.

The prior we will be using is the flat prior we derived earlier $\pi_{B_1,\ldots,B_p,\Sigma,P}(\cdot) \propto |\Sigma|^{-\frac{np+2n+1}{2}}$. According to Theorem 4.3.4, this prior is equivalent to $\pi_{A_0,A_1,\cdot,A_p}(\cdot) \propto 1$ indicating that we do not have any additional information. Working with which parameterization is just a matter of personal preference. Since $\pi_{B_1,\ldots,B_p,\Sigma,P}(\cdot) \propto |\Sigma|^{-\frac{np+2n+1}{2}}$ is a conjugate prior for the reduced-form parameterization, it is computationally efficient to draw inference directly from the Normal-inverse-Wishart distribution. It also indicates that the prior for orthogonal matrix is $\pi_P(\cdot) \propto 1$, which is the uniform prior in Proposition 4.3.1.

Figure 4.1 reports the quantiles of the impulse responses to a contractionary monetary policy shock with one standard deviation in size in a pure sign restricted structural VAR based on 10000 draws. We could interpret the result as:

1. The federal funds rate reacts positively immediately, but drops below zero after one year.

2. The GDP deflator is dropping in the five year course considered here.

3. With a 2/3 probability, the response for real GDP is within the $\pm 0.2\%$ interval around zero.

4. The commodity price index reaches a plateau of a 1.5% drop after one year.

5. Non-borrowed reserves and total reserves drop immediately, around 1% and 0.6% respectively.

As a comparison, Figure 4.2 reports the 68% highest posterior region robust credible set of the associated impulse responses to a contractionary monetary policy shock with one standard deviation in size. An immediate difference is that the credible sets for all impulse responses reported are much wider. This is no surprise, since the credible set is robust to the choice of different prior specification as demonstrated in section 4.4.2. We can reinterpret the result as:

1. The federal funds rate reacts positively immediately because of the sign restriction we imposed, no clear inference can be made after one year.

2. The GDP deflator is most likely to drop in the five year course considered.

3. With a 2/3 probability, the response for real GDP is within the $\pm 0.8\%$ interval around zero.

4. There is a lot of fluctuation in the commodity price index, ranging from -8% to 3%.

5. Non-borrowed reserves drop immediately because of the sign restrictions that we imposed, total reserves could react positively or negatively.

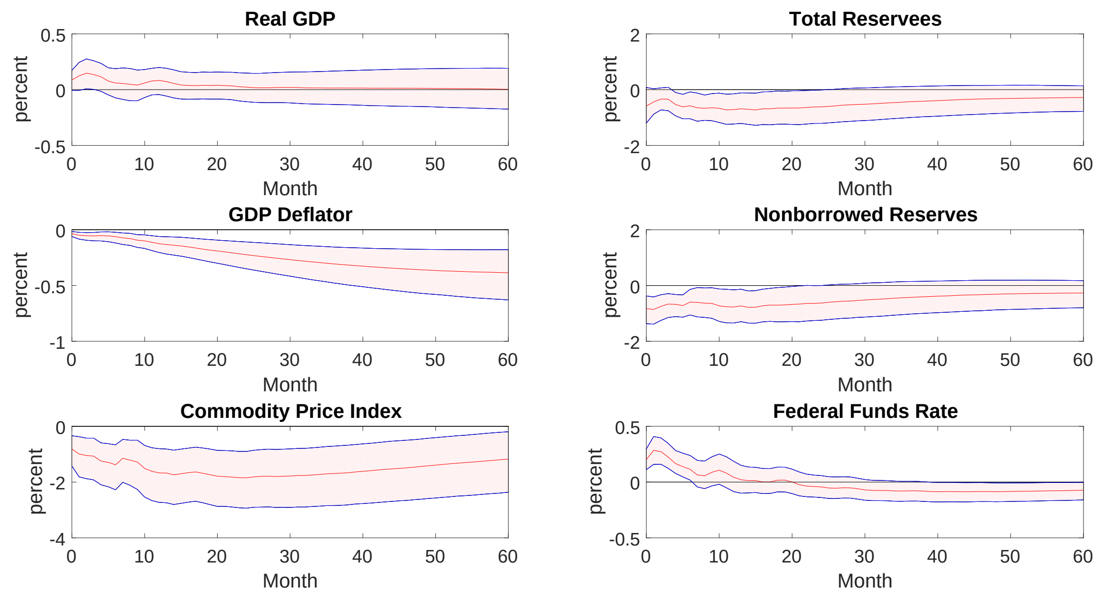

Figure 4.1 Responses to a contractionary monetary policy shock in pure sign identified model: 68% quantile of impulse response

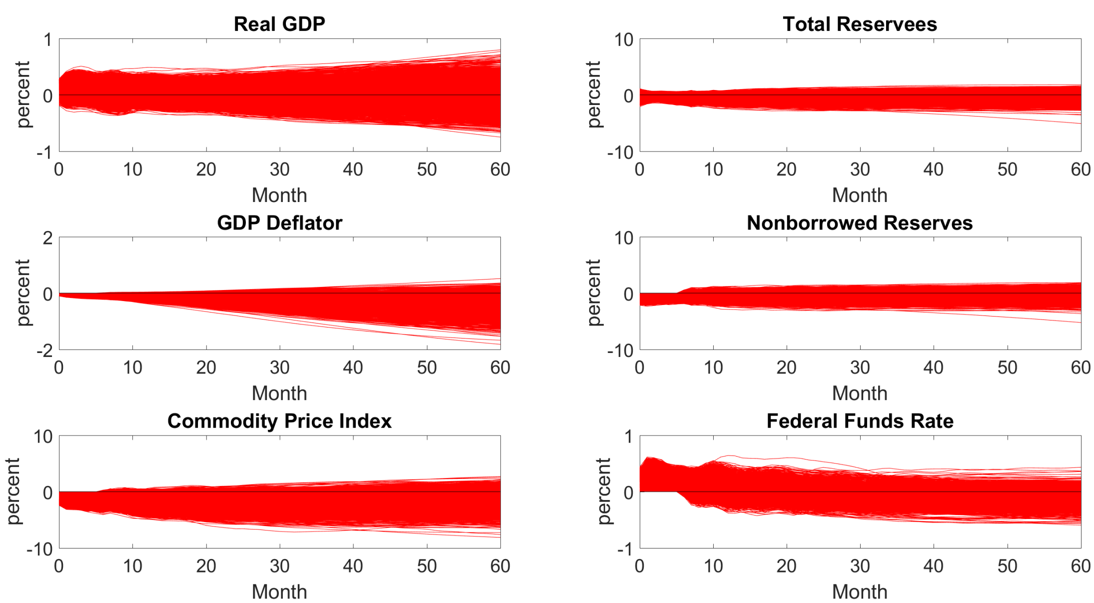

Figure 4.2 Responses to a contractionary monetary policy shock in pure sign identified model: 68% highest posterior region of impulse response

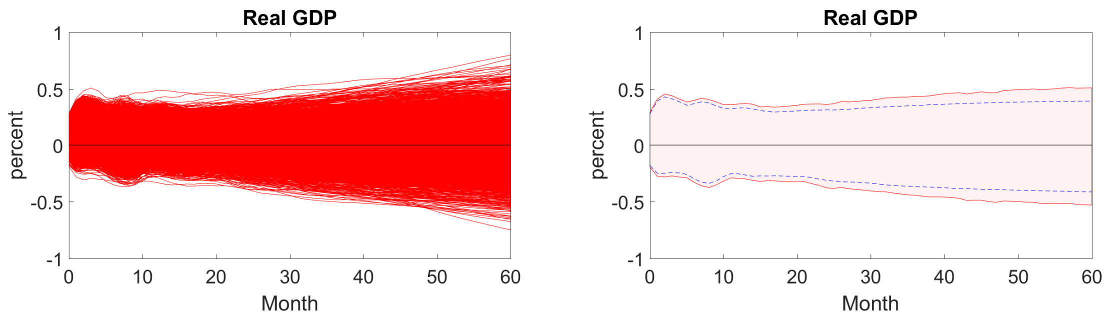

Figure 4.3 Responses to a contractionary monetary policy shock in pure sign identified model: 68% highest posterior region and 68% robustified credible region of impulse response.

Lastly, we implement the robust Bayesian approach proposed by [38] to illustrate that the highest posterior region robust credible set constructed here is asymptotically equivalent to their approach which allows multiple priors for P. The algorithm used to estimate the set identified SVAR with the robust Bayesian approach is as follows[15].

Algorithm 4.5.1. *(step 1) Draw reduced-form parameter* $(B_1, \ldots, B_p, \Sigma)$ *jointly from Normal-inverse-Wishart posterior distribution.*

(step 2) Set every success draws of P *from "agnostic" prior as initial values of the MATLAB optimization routine to calculate the lower bound* $l^h(B_1, \ldots, B_p, \Sigma)$ *and upper bound* $u^h(B_1, \ldots, B_p, \Sigma)$ *for the impulse response subject to sign restrictions.*

(step 3) Repeat step 1 and step 2, $N = 1000$ *times, and obtain* N *draws of the intervals. Approximate the posterior mean bound by the sample average of* $l_n^h(B_1, \ldots, B_p, \Sigma)$ *and* $u_n^h(B_1, \ldots, B_p, \Sigma)$*,* $n = 1, \cdots, N$*.*

(step 4) Obtain the approximated robustified credible region with credibility $\alpha \in (0,1)$ *using Proposition 5.1 of [55].*
Define $d_n^h(B_1, \ldots, B_p, \Sigma, r) = max\{\left|r - l_n^h\right|, \left|r - u_n^h\right|\}$*, let* $\hat{z}_\alpha(r)$ *be the sample* α*-th quantile of* $(d_n^h(B_1, \ldots, B_p, \Sigma, r) : n = 1, \cdots, N)$*. An approximated robustified credible region for* r *is obtained as an interval centered at* $arg\ min_r \hat{z}_\alpha(r)$ *with radius* $min_r \hat{z}_\alpha(r)$*.*

Figure 4.3 depicts the impulse response of real GDP to a contractionary monetary shock. The left panel is the constructed robust credible set based on the highest posterior region, and it is exactly the same as the upper-left panel of Figure 4.2. The right panel is obtained from the robust Bayesian approach using algorithm 4.5.1[16]. From the graph, the results from both methods are similar, and it is straightforward that under the current prior specification, two credible sets are asymptotically equivalent[17].

[15] This algorithm was developed in [38].

[16] The red lines represent the robustified credible region and the blue lines represent posterior mean bounds.

[17] See Proposition 6.1 in [38] for a discussion about the asymptotic convergence of posterior mean bound. The

4.5.2 Sign Restrictions With Zero Restrictions

Many economists would disagree with the conclusion that a contractionary monetary policy shock will not necessarily lead to a decrease in real GDP. In the pure sign restrictions approach, we impose sign restrictions that directly eliminate the price and liquidity puzzle. Here, aside from the sign restrictions, we further impose that federal funds rate does not react to total reserves and non-borrowed reserves contemporaneously, and the contemporaneous reaction of federal funds rate to real GDP and GDP deflator is positive[18]. All the zero restrictions are placed on the impact matrix A_0, and are $[A_0]_{64} = 0$, $[A_0]_{65} = 0$, $[A_0]_{61} < 0$ and $[A_0]_{62} < 0$[19].

We treat hard restrictions $[A_0]_{64} = 0$ and $[A_0]_{65} = 0$ as soft restrictions, $\pi([A_0]_{64}) \sim N(0, V_1)$ and $\pi([A_0]_{65}) \sim N(0, V_2)$. We use Algorithm 4.4.1 to estimate, with a slight modification of step (4) and set importance weight as $\phi([A_0]_{64}, 0, V_1) \times \phi([A_0]_{65}, 0, V_2)$, where $\phi(\cdot)$ represent the normal density function. We report the monetary equation

$$ffr_t = \psi_y y_t + \psi_p p_t + \psi_{pc} pc_t + \psi_{tr} tr_t + \psi_{nbr} nbr_t + \sigma_{6,t} \tag{4.5.1}$$

where $\psi_y = -[A_0]_{61}/[A_0]_{66}$, $\psi_p = -[A_0]_{62}/[A_0]_{66}$, $\psi_{pc} = -[A_0]_{63}/[A_0]_{66}$, $\psi_{tr} = -[A_0]_{64}/[A_0]_{66}$, $\psi_{nbr} = -[A_0]_{65}/[A_0]_{66}$ and $\sigma_{6,t} = \epsilon_{6,t}/[A_0]_{66}$. We set $V_1 = v_1\sigma_1^2$ and $V_2 = v_2\sigma_2^2$, where σ_1^2 and σ_2^2 are calibrated by the variance of $[A_0]_{64}$ and $[A_0]_{65}$ under pure sign restrictions considered in Section 4.5.1.

Table 4.1 reports the results we obtained using importance sampling with proposal density that is equal to pure sign restricted posterior density based on 10000 random draws[20]. As v_1 and v_2 become smaller, the medians of ψ_{tr} and ψ_{nbr} both move toward zero, and their 68 and 95 percent posterior density intervals become tighter, indicating that our methodology is a strict generalization of the traditional hard zero restrictions.

Under the identification scheme, the posterior median of ϕ_y ranges from 0.33 to 0.51, indicating that the federal funds rate reacts $\frac{1}{3}$ to $\frac{1}{2}$ to a contemporaneous change in output. The posterior median of ϕ_p ranges from 2.32 to 2.85, indicating that the federal funds rate reacts

---

convergence of robust credible set constructed based on the highest posterior region follows from the convergence of the Normal-inverse-Wishart posterior distribution.

[18]This identification scheme is proposed by [8].

[19]To see this, we can rewrite the monetary equation, ignoring the lag variables

$$[A_0]_{61}y_t + [A_0]_{62}p_t + [A_0]_{63}pc_t + [A_0]_{64}tr_t + [A_0]_{65}nbr_t + [A_0]_{66}ffr_t = \epsilon_{6,t}$$

or equivalently

$$ffr_t = -([A_0]_{61}y_t + [A_0]_{62}p_t + [A_0]_{63}pc_t + [A_0]_{64}tr_t + [A_0]_{65}nbr_t)/[A_0]_{66} + \epsilon_{6,t}/[A_0]_{66}$$

Restrictions we imposed are $-[A_0]_{64}/[A_0]_{66} = 0$, $-[A_0]_{65}/[A_0]_{66} = 0$, $-[A_0]_{61}/[A_0]_{62} > 0$ and $-[A_0]_{62}/[A_0]_{66} > 0$. From sign restrictions in Section 4.5.1, $[A_0]_{66} > 0$.

[20]The sample size is set to 10000 to make sure the efficiency of the importance sampler.

Table 4.1 Contemporaneous Coefficients in the Monetary Equations

(v_1, v_2)		ψ_y	ψ_p	ψ_{pc}	ψ_{tr}	ψ_{nbr}
(0.01,0.01)	Median	0.33	2.32	0.06	0.00	0.01
	68% CI	[0.09,0.70]	[1.19,4.00]	[0.01,0.15]	[-0.01,0.03]	[-0.01,0.02]
	95% CI	[0.02,1.19]	[0.45,7.01]	[-0.01,0.31]	[-0.03,0.05]	[-0.02,0.04]
(0.1,0.1)	Median	0.39	2.33	0.07	0.03	0.03
	68% CI	[0.11,0.85]	[1.15,4.28]	[0.02,0.17]	[-0.02,0.09]	[-0.01,0.08]
	95% CI	[0.02,1.43]	[0.38,7.30]	[-0.01,0.32]	[-0.08,0.18]	[-0.05,0.17]
(1,1)	Median	0.45	2.49	0.08	0.04	0.07
	68% CI	[0.13,1.00]	[1.17,4.85]	[0.02,0.19]	[-0.13,0.20]	[-0.05,0.27]
	95% CI	[0.02,1.89]	[0.36,10.11]	[-0.01,0.41]	[-0.43,0.46]	[-0.15,0.80]
(10,10)	Median	0.50	2.74	0.09	0.00	0.14
	68% CI	[0.14,1.16]	[1.17,6.12]	[0.02,0.21]	[-0.40,0.23]	[-0.07,0.59]
	95% CI	[0.02,2.87]	[0.30,18.28]	[-0.02,0.58]	[-1.96,0.54]	[-0.23,2.54]
(100,100)	Median	0.51	2.85	0.09	-0.02	0.15
	68% CI	[0.15,1.22]	[1.22,6.44]	[0.02,0.22]	[-0.49,0.23]	[-0.07,0.68]
	95% CI	[0.02,3.30]	[0.32,21.49]	[-0.02,0.71]	[-2.66,0.53]	[-0.23,3.36]

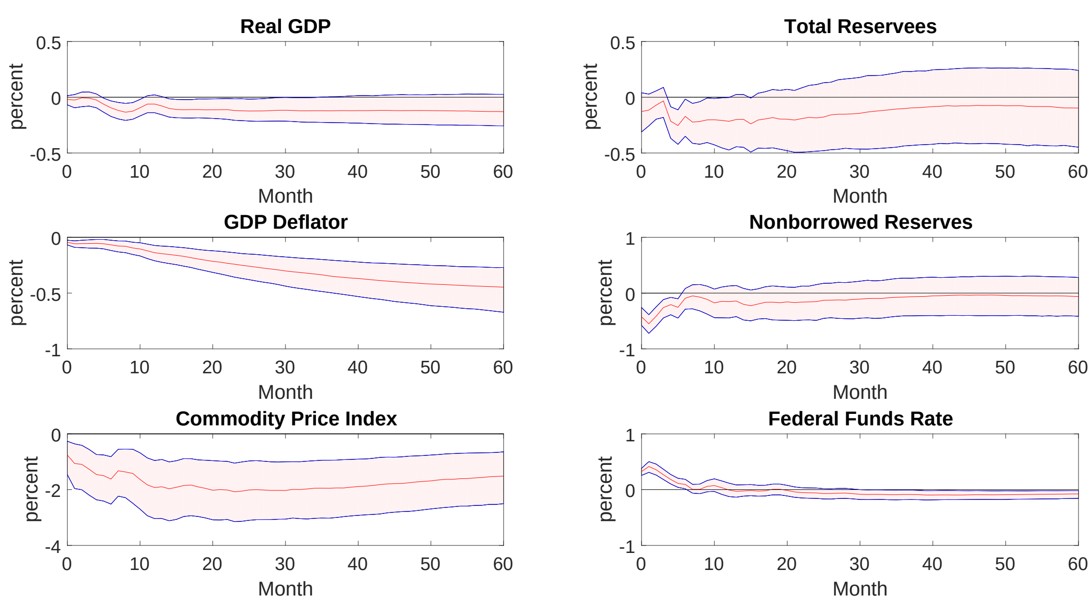

Figure 4.4 Responses to a contractionary monetary policy shock: 68% quantile of impulse response, $v_1 = 0.01$, $v_2 = 0.01$

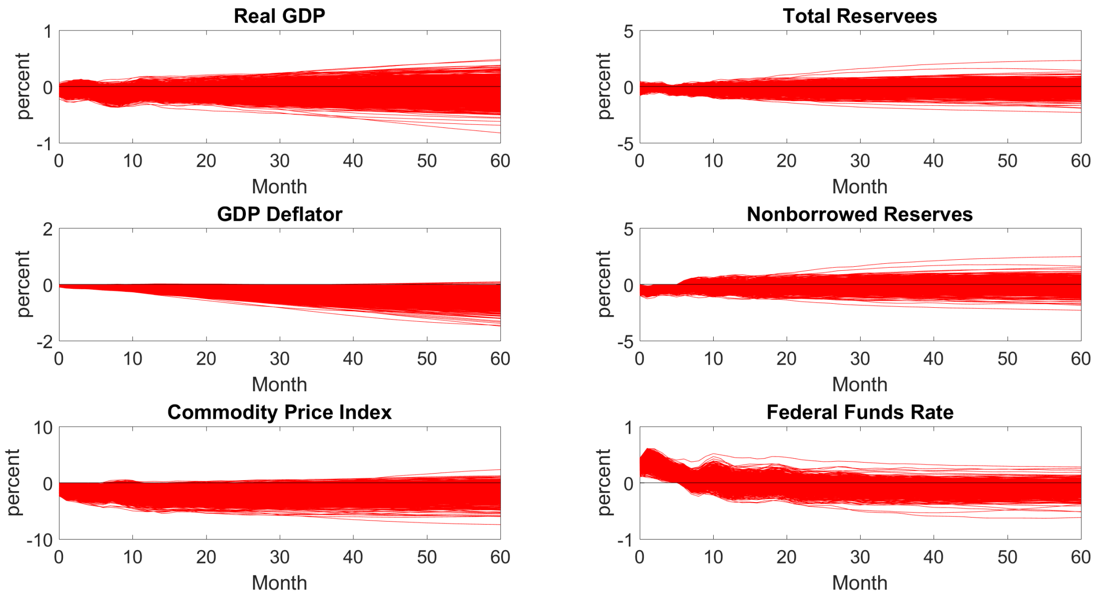

Figure 4.5 Responses to a contractionary monetary policy shock: 68% highest posterior region of impulse response, $v_1 = 0.01$, $v_2 = 0.01$

more than one to one to contemporaneous change in GDP deflator. The posterior medians of ϕ_{pc}, ϕ_{tr} and ϕ_{nbr} are close to 0, indicating that they have minor effects.

Figure 4.4 report quantiles of impulse responses to a contractionary monetary policy shock, we can interpret the result as:

1. The federal funds rate reacts positively immediately and drops below zero after one year.

2. The GDP deflator is dropping in the five year course that we considered.

3. The response for real GDP is hump shaped, and below zero.

4. The commodity price index reaches a plateau of a 2% drop after one year.

5. Non-borrowed reserves and total reserves drop immediately, around 0.4% and 0.1% respectively, and then increase in the five year course.

Figure 4.5 reports the 68% highest posterior region robust credible set of the associated impulse responses to a contractionary monetary policy shock. As argued in Section 4.5.1, the credible set is considered to be robust to different prior specifications, and thus we can reinterpret the result as:

1. The federal funds rate reacts positively immediately.

2. The GDP deflator drops in the five year course considered.
3. With a 2/3 probability, the response for real GDP is within the $\pm 0.5\%$ interval around zero.
4. The response of commodity price index is negative.
5. Non-borrowed reserves drop immediately. Total reserves could react positively or negatively.

4.6 Conclusion

This paper studies four different parameterizations of SVARs with the goal of making probability statement about the posterior density across all four parameterizations. The main contribution of our study is the calculation of the determinants of the Jacobian matrices using the change of variables formula. We use the exterior differential form extensively in the derivation. We demonstrate our main result with two applications: combining sign restrictions with zero exclusive restrictions and constructing joint credible sets that is robust to different prior specifications. Both applications are intended to solve the problem that the data provide no information to distinguish between any two pairs of structural parameters in the admissible set. We demonstrate our methodology by revisiting the work of [75] and [8].

Chapter 5

Conclusions

The three essays in the dissertation are concerned with financial and time series econometrics. The first two essays consider the problem of estimating volatility based on the high frequency financial data and the last essay consider the problem of using sign restrictions in SVARs model. In this dissertation, we have proposed efficient approaches for solving these problems.

In Chapter 2, we study a class of volatility measures based on the Laplace transform, which is robust to the presence of microstructure noise and the time endogeneity. Asymptotic properties and feasible central limit theorems are established. Given the presence of time endogeneity, our bias-corrected version Laplace estimator allows taking advantage of the informational content of time endogeneity, which leads to an efficiency gain and narrower confidence bounds. Through Monte Carlo study, we study the optimal choice of the convolution frequency in the presence of microstructure noise. We also compare the performance of the Laplace estimator with other popular estimators through forecasting exercises by employing high frequency data. We conclude that the naive and the bias-corrected Laplace estimator perform better than most estimators in terms of forecasting equity return volatility.

In Chapter 3, we have provided a detailed analysis of the performance of the realized kernel estimator of the Integrated Volatility under time endogeneity and the presence of market microstructure noise due to frictions. We derive central limit theorems for the realized kernel estimators in a nonparametric setting. Our framework contains various assumptions about the kernel weight and a general endogenous setting. We further show that the realized kernel estimator can be made to converge at the fastest possible rate, robust to different specifications of the noise and robust to dynamics in the noise process. Evidence from Monte Carlo experiments confirmed our theoretical results. We have also applied our methodology to the millisecond time stamped trade data and illustrated that the realized kernel estimator could be adopted in a model that are sampled at the highest possible frequency.

In Chapter 4, we study four different parameterizations of SVARs with the goal of making probability statement about the posterior density across all four parameterizations. The main contribution of our study is the calculation of the determinants of the Jacobian matrices using the change of variables formula. We use the exterior differential form extensively in the derivation. We demonstrate our main result with two applications: combining sign restrictions with zero exclusive restrictions and constructing joint credible sets that is robust to different prior specifications. Both applications are intended to solve the problem that the data provide no information to distinguish between any two pairs of structural parameters in the admissible set. We demonstrate our methodology by revisiting the work of [75] and [8].

There are several research directions worthy of further studies in the future. A very promising extension of Chapter 3 that will be considered in the future paper is inferencing for a multivariate parameter. There are two challenges. The first is how to model and capture time endogeneity in a multivariate setting, and the second challenge is the effect of the asynchronous trading. Another possible extension is a consistent estimate of instantaneous (spot) volatility. We gave two consistent estimators that are robust to microstructure noise and time endogeneity in Section 3.5.4. The difficulty arises from deriving a stable convergence in law results. The third possible extension is to study the most efficient endogenous sampling scheme. As the asymptotic variance of the limiting process depends on a_t, b_t, $\Psi(t)$ and $H(t)$, an interesting question is to design the most efficient sampling scheme and choose the realized kernel $k(\cdot)$ to minimize asymptotic variance.

BIBLIOGRAPHY

[1] Aït-Sahalia, Y. & Jacod, J. *High-frequency financial econometrics*. Princeton University Press, 2014.

[2] Ait-Sahalia, Y. & Mykland, P. A. "The Effects of Random and Discrete Sampling when Estimating Continuous–Time Diffusions". *Econometrica* **71** (2 2003), pp. 483–549.

[3] Aldous, D. J., Eagleson, G., et al. "On mixing and stability of limit theorems". *The Annals of Probability* **6**.2 (1978), pp. 325–331.

[4] Andersen, T. G. & Bollerslev, T. "Answering the skeptics: Yes, standard volatility models do provide accurate forecasts". *International economic review* (1998), pp. 885–905.

[5] Andersen, T. G. et al. "The distribution of realized exchange rate volatility". *Journal of the American statistical association* **96**.453 (2001), pp. 42–55.

[6] Andersen, T. G. et al. "Modeling and forecasting realized volatility". *Econometrica* **71**.2 (2003), pp. 579–625.

[7] Andersen, T. G. et al. "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility". *The review of economics and statistics* **89**.4 (2007), pp. 701–720.

[8] Arias, J. et al. "The systematic component of monetary policy in SVARs: an agnostic identification procedure" (2015).

[9] Arias, J. E. et al. "Inference Based on Structural Vector Autoregressions Identified With Sign and Zero Restrictions: Theory and Applications". *Econometrica* **86**.2 (2018), pp. 685–720.

[10] Aït-Sahalia, Y. et al. "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise". *The Review of Financial Studies* **18**.2 (2005), pp. 351–416.

[11] Bandi, F. M. & Russell, J. R. "Separating microstructure noise from volatility". *Journal of Financial Economics* **79**.3 (2006), pp. 655–692.

[12] Bandi, F. M. & Russell, J. R. "Microstructure noise, realized variance, and optimal sampling". *The Review of Economic Studies* **75**.2 (2008), pp. 339–369.

[13] Barndorff-Nielsen, O. E. & Shephard, N. "Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**.2 (2001), pp. 167–241.

[14] Barndorff-Nielsen, O. E. & Shephard, N. “Econometric analysis of realized volatility and its use in estimating stochastic volatility models”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**.2 (2002), pp. 253–280.

[15] Barndorff-Nielsen, O. E. & Shephard, N. “Power and bipower variation with stochastic volatility and jumps”. *Journal of financial econometrics* **2**.1 (2004), pp. 1–37.

[16] Barndorff-Nielsen, O. E. et al. “Designing realized kernels to measure the ex post variation of equity prices in the presence of noise”. *Econometrica* **76**.6 (2008), pp. 1481–1536.

[17] Barndorff-Nielsen, O. E. et al. “Realized kernels in practice: Trades and quotes”. *The Econometrics Journal* **12**.3 (2009), pp. C1–C32.

[18] Barndorff-Nielsen, O. E. et al. “Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading”. *Journal of Econometrics* **162**.2 (2011), pp. 149–169.

[19] Baumeister, C. & Hamilton, J. D. “Sign restrictions, structural vector autoregressions, and useful prior information”. *Econometrica* **83**.5 (2015), pp. 1963–1999.

[20] Bernanke, B. S. *Alternative explanations of the money-income correlation*. 1986.

[21] Bibinger, M. & Mykland, P. A. “Inference for Multi-dimensional High-frequency Data with an Application to Conditional Independence Testing”. *Scandinavian Journal of Statistics* **43**.4 (2016), pp. 1078–1102.

[22] Black, F. “Noise”. *The journal of finance* **41**.3 (1986), pp. 528–543.

[23] Blanchard, O. J. & Quah, D. *The dynamic effects of aggregate demand and supply disturbances*. 1988.

[24] Bollerslev, T. et al. “Exploiting the errors: A simple approach for improved volatility forecasting”. *Journal of Econometrics* **192**.1 (2016), pp. 1–18.

[25] Canova, F. & De Nicolo, G. “Monetary disturbances matter for business fluctuations in the G-7”. *Journal of Monetary Economics* **49**.6 (2002), pp. 1131–1159.

[26] Clement, E. & Gloter, A. “Limit theorems in the Fourier transform method for the estimation of multivariate volatility”. *Stochastic Processes and their Applications* **121**.5 (2011), pp. 1097–1124.

[27] Comte, F. & Renault, E. “Long memory in continuous-time stochastic volatility models”. *Mathematical finance* **8**.4 (1998), pp. 291–323.

[28] Corsi, F. “A simple approximate long-memory model of realized volatility”. *Journal of Financial Econometrics* **7**.2 (2009), pp. 174–196.

[29] Cui, W. "Laplace Estimator of Integrated Volatility When Sampling Times Are Endogenous". *Working Paper* (2019).

[30] Curato, I. V. et al. "Spot volatility estimation using the Laplace transform". *Econometrics and Statistics* **6** (2016), pp. 22 –43.

[31] Delbaen, F. & Schachermayer, W. "A general version of the fundamental theorem of asset pricing". *Mathematische annalen* **300**.1 (1994), pp. 463–520.

[32] Delbaen, F. & Schachermayer, W. "A General Version of the Fundamental Theorem of Asset Pricing (1994)". *The Mathematics of Arbitrage*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 149–205.

[33] Fan, J. & Wang, Y. "Multi-Scale Jump and Volatility Analysis for High-Frequency Financial Data". *Journal of the American Statistical Association* **102** (2007), pp. 1349–1362.

[34] Faust, J. "The robustness of identified VAR conclusions about money". *Carnegie-Rochester Conference Series on Public Policy*. Vol. 49. Elsevier. 1998, pp. 207–244.

[35] Foster, D. P. & Nelson, D. B. "Continuous Record Asymptotics for Rolling Sample Variance Estimators". *Econometrica* **64**.1 (1996), p. 139.

[36] Fry, R. & Pagan, A. "Sign restrictions in structural vector autoregressions: A critical review". *Journal of Economic Literature* **49**.4 (2011), pp. 938–960.

[37] Fukasawa, M. & Rosenbaum, M. "Central limit theorems for realized volatility under hitting times of an irregular grid". *Stochastic Processes and their Applications* **122** (12 2012), pp. 3901–3920.

[38] Giacomini, R. & Kitagawa, T. "Robust inference about partially identified SVARs". *Manuscript, University College London* (2015).

[39] Glosten, L. R. & Milgrom, P. R. "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders". *Journal of financial economics* **14**.1 (1985), pp. 71–100.

[40] Hamilton, J. D. *Time series analysis*. Vol. 2. Princeton university press Princeton, 1994.

[41] Hansen, P. R. & Lunde, A. "Realized variance and market microstructure noise". *Journal of Business & Economic Statistics* **24**.2 (2006), pp. 127–161.

[42] Harris, L. "Stock price clustering and discreteness". *The Review of Financial Studies* **4**.3 (1991), pp. 389–415.

[43] Hasbrouck, J. *Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading*. Vol. 1. Oxford University Press, USA, 2007.

[44] Hayashi, T., Yoshida, N., et al. "On covariance estimation of non-synchronously observed diffusion processes". *Bernoulli* **11**.2 (2005), pp. 359–379.

[45] Hayashi, T. et al. "Irregular sampling and central limit theorems for power variations : the continuous case". en. *Annales de l'I.H.P. Probabilités et statistiques* **47**.4 (2011), pp. 1197–1218.

[46] Ikeda, S. S. "A bias-corrected estimator of the covariation matrix of multiple security prices when both microstructure effects and sampling durations are persistent and endogenous". *Journal of Econometrics* **193**.1 (2016), pp. 203 –214.

[47] Inoue, A. & Kilian, L. "Inference on impulse response functions in structural VAR models". *Journal of Econometrics* **177**.1 (2013), pp. 1–13.

[48] Inoue, A. & Kilian, L. "Corrigendum to "Inference on Impulse Response Functions in Structural VAR Models" [J. Econometrics 177 (2013), 1-13] " (2018). forthcoming: Journal of Econometrics.

[49] Jacod, J. & Shiryaev, A. *Limit theorems for stochastic processes*. Vol. 288. Springer Science & Business Media, 2013.

[50] Jacod, J. et al. "Microstructure noise in the continuous case: the pre-averaging approach". *Stochastic processes and their applications* **119**.7 (2009), pp. 2249–2276.

[51] James, A. T. et al. "Normal multivariate analysis and the orthogonal group". *The Annals of Mathematical Statistics* **25**.1 (1954), pp. 40–75.

[52] Kalnina, I. & Linton, O. "Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error". *Journal of econometrics* **147**.1 (2008), pp. 47–59.

[53] Karatzas, I. & Shreve, S. *Brownian motion and stochastic calculus*. Vol. 113. Springer Science & Business Media, 2012.

[54] Kilian, L. & Murphy, D. P. "Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models". *Journal of the European Economic Association* **10**.5 (2012), pp. 1166–1188.

[55] Kitagawa, T. "Estimation and inference for set-identified parameters using posterior lower probability". *Manuscript, UCL* (2012).

[56] Koike, Y. "Time endogeneity and an optimal weight function in pre-averaging covariance estimation". *Statistical Inference for Stochastic Processes* **20**.1 (2017), pp. 15–56.

[57] Li, Y. et al. "Volatility inference in the presence of both endogenous time and microstructure noise". *Stochastic Processes and their Applications* **123**.7 (2013), pp. 2696–2727.

[58] Li, Y. et al. "Realized volatility when sampling times are possibly endogenous". *Econometric Theory* **30**.3 (2014), pp. 580–605.

[59] Malliavin, P. & Mancino, M. E. "Fourier series method for measurement of multivariate volatilities". *Finance and Stochastics* **6**.1 (2002), pp. 49–61.

[60] Malliavin, P., Mancino, M. E., et al. "A Fourier transform method for nonparametric estimation of multivariate volatility". *The Annals of Statistics* **37**.4 (2009), pp. 1983–2010.

[61] Mancino, M. E. & Sanfelici, S. "Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise". *Computational Statistics & data analysis* **52**.6 (2008), pp. 2966–2989.

[62] Muirhead, R. J. *Aspects of multivariate statistical theory*. Vol. 197. John Wiley & Sons, 2009.

[63] Mykland, P. A. & Zhang, L. "Inference for continuous semimartingales observed at high frequency". *Econometrica* **77**.5 (2009), pp. 1403–1445.

[64] Mykland, P. A. & Zhang, L. "ANOVA for diffusions and Ito processes". *The Annals of Statistics* **34**.4 (2006), pp. 1931–1963.

[65] Newey, W. K. & West, K. D. *A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix*. 1986.

[66] O'hara, M. *Market microstructure theory*. Vol. 108. Blackwell Cambridge, 1995.

[67] Park, S. et al. "Estimating the quadratic covariation matrix for asynchronously observed high frequency stock returns corrupted by additive measurement error". *Journal of Econometrics* **191**.2 (2016), pp. 325–347.

[68] Phillips, P. C. & Yu, J. "Comment on" Realized Variance and Market Microstructure Noise" by Peter R. Hansen and Asger Lunde" (2005).

[69] Potiron, Y. & Mykland, P. A. "Estimation of integrated quadratic covariation with endogenous sampling times". *Journal of Econometrics* **197**.1 (2017), pp. 20–41.

[70] Roll, R. "A simple implicit measure of the effective bid-ask spread in an efficient market". *The Journal of finance* **39**.4 (1984), pp. 1127–1139.

[71] Rubio-Ramirez, J. F. et al. "Structural vector autoregressions: Theory of identification and algorithms for inference". *The Review of Economic Studies* **77**.2 (2010), pp. 665–696.

[72] Sims, C. A. "Macroeconomics and reality". *Econometrica: Journal of the Econometric Society* (1980), pp. 1–48.

[73] Stewart, G. W. "The efficient generation of random orthogonal matrices with an application to condition estimators". *SIAM Journal on Numerical Analysis* **17**.3 (1980), pp. 403–409.

[74] Swanson, N. R. & White, H. "Forecasting economic time series using flexible versus fixed specification and linear versus nonlinear econometric models". *International journal of Forecasting* **13**.4 (1997), pp. 439–461.

[75] Uhlig, H. "What are the effects of monetary policy on output? Results from an agnostic identification procedure". *Journal of Monetary Economics* **52**.2 (2005), pp. 381–419.

[76] Xiu, D. "Quasi-maximum likelihood estimation of volatility with high frequency data". *Journal of Econometrics* **159**.1 (2010), pp. 235–250.

[77] Zhang, L. et al. "A tale of two time scales: Determining integrated volatility with noisy high-frequency data". *Journal of the American Statistical Association* **100**.472 (2005), pp. 1394–1411.

[78] Zhang, L. et al. "Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach". *Bernoulli* **12**.6 (2006), pp. 1019–1043.

APPENDICES

Appendix A

Supplemental Material for Chapter 2

Throughout this section, we follow the method presented in [63] by using Girsanov's Theorem to suppress μ_t, and act as if X_t is a martingale. We considered a probability measure P^*, under which X_t is a local martingale

$$dX_t = \sigma_t \, dW_t^*,$$

where W_t^* is a P^*-Brownian motion, with

$$dW_t^* = dW_t + \sigma_t^{-1} \mu_t \, dt.$$

Following the Girsanov's Theorem, the likelihood ratio (Radon-Nikodym derivative)

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

By Proposition 1 in [63], we can carry out the analysis under P^*, and by the mutually absolute continuity between P^* and P: if Z_n is a sequence of random variables which converges stably to $b + aN(0,1)$ under P^*, where $N(0,1)$ is a standard normal variable independent of $\mathcal{F}_T$, a and b are also $\mathcal{F}_T$-measurable. Then Z_n converges stably in law to $b + aN(0,1)$ under P, where $N(0,1)$ remains independent of $\mathcal{F}_T$ under P.

A.1 Proofs

In this section, we first prove Theorem 2.3.4, and Theorem 2.3.2 is a special case of it with $b_v = 2\Psi(v)$.

Proof of Theorem 2.3.4. Using Itô formula, the residual process is

$$\bar{C}_T - \langle X, X\rangle_T = 2 \times \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right) dX_u\, dX_v. \tag{A.1.1}$$

We use notation A_1 to denote half of the residual process as in Eq. (A.1.1), where

$$A_1 = \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right) dX_u\, dX_v.$$

Apply Itô formula to Eq. (2.3.3) we have,

$$\begin{aligned} & H^{-1/2} \sum_{t_{j+1}\le T} \left(\sum_{t_{i+1}\le T} k\left(\frac{t_i - t_j}{H}\right) \Delta X_{t_i} \right) \Delta t_j \\ &= H^{-1/2} \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right) dX_u\, dv + H^{-1/2} \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right) du\, dX_v. \end{aligned} \tag{A.1.2}$$

We show the second term in Eq. (A.1.2), $H^{-1/2} \int_0^T \int_0^v k\left(\frac{t^*(u)-t^*(v)}{H}\right) du\, dX_v \xrightarrow{p} 0$ in probability. Consider the expected quadratic variation,

$$\begin{aligned} & E\langle H^{-1/2} \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right) du\, dX_v, H^{-1/2} \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right) du\, dX_v\rangle_t \\ &= H^{-1} E \int_0^T \left(\int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right) du \right)^2 d\langle X, X\rangle_v \\ &\le ||\sigma^2||_\infty H^{-1} E \int_0^T \left(\int_0^T |k\left(\frac{t^*(u) - t^*(v)}{H}\right)|\, du \right)^2 dv \\ &\le ||\sigma^2||_\infty H^{-1} E \int_0^T \left(\int_0^T |k\left(\frac{t^*(u) - t^*(v)}{H}\right)|^p\, du \right)^{\frac{2}{p}} dv \\ &\le ||\sigma^2||_\infty \left(H^{-1}\right)^{1-\frac{2}{p}} E \int_0^T \left(\int_0^T H^{-1} |k\left(\frac{t^*(u) - t^*(v)}{H}\right)|^p\, du \right)^{\frac{2}{p}} dv, \end{aligned}$$

where we use Hölder's inequality in the second to last line for any $p > 1$. By Lemma A.2.1 and taking $1 < p < 2$, it converges to 0 in probability. This concludes that from Eq. (2.3.3) we obtain,

$$H^{-1/2} \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right) dX_u \, dv \xrightarrow{p} \int_0^T a_v \sigma_v \, dv.$$

And further leads to

$$\langle H^{-1/2} A_1, X \rangle_T \xrightarrow{p} \int_0^T a_v \sigma_v^3 \, dv. \tag{A.1.3}$$

Similarly, apply Itô formula to Eq. (2.3.4) we have

$$\begin{aligned}
& H^{-1} \sum_{t_{j+1} \leq T} \left(\sum_{t_{i+1} \leq T} k\left(\frac{t_i - t_j}{H}\right) \Delta X_{t_i} \right)^2 \Delta t_j \\
&= H^{-1} \int_0^T \left(\int_0^T k\left(\frac{t^*(u) - t^*(v)}{H}\right) dX_u \right)^2 dv \\
&= H^{-1} \int_0^T \left(\int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right) dX_u \right)^2 dv + H^{-1} \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right)^2 du d\langle X, X \rangle_v \\
&+ 2 \times H^{-1} \int_0^T \left(\int_0^v \left(\int_0^v k\left(\frac{t^*(u) - t^*(u_1)}{H}\right) dX_u \right) k\left(\frac{t^*(u_1) - t^*(v)}{H}\right) du_1 \right) dX_v. \quad \text{(A.1.4)}
\end{aligned}$$

By Assumption H2.3 we have the following convergence in probability

$$H^{-1} \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right)^2 du \, d\langle X, X \rangle_v \xrightarrow{p} \int_0^T \Psi(v) \sigma_v^2 \, dv. \tag{A.1.5}$$

Next we show the last term in Eq. (A.1.4) converges in probability to 0,

$$H^{-1} \int_0^T \left(\int_0^v \left(\int_0^v k\left(\frac{t^*(u) - t^*(u_1)}{H}\right) dX_u \right) k\left(\frac{t^*(u_1) - t^*(v)}{H}\right) du_1 \right) dX_v \xrightarrow{p} 0. \tag{A.1.6}$$

To start consider the expected quadratic variation,

$$
\begin{aligned}
&= H^{-2}E\int_0^T\left(\int_0^v\left(\int_0^v k\left(\frac{t^*(u)-t^*(u_1)}{H}\right)\,dX_u\right)k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)\,du_1\right)^2\,d\langle X,X\rangle_v\\
&\le ||\sigma^2||_\infty H^{-2}E\int_0^T\left(\int_0^v\left(\int_0^v k\left(\frac{t^*(u)-t^*(u_1)}{H}\right)\,dX_u\right)k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)\,du_1\right)^2\,dv\\
&\le ||\sigma^2||_\infty H^{-1}E\int_0^T\left(\int_0^v\left(H^{-1/2}\int_0^v k\left(\frac{t^*(u)-t^*(u_1)}{H}\right)dX_u\right)k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)du_1\right)^2\,dv.
\end{aligned}
$$

By the Burkholder-Davis-Gundy inequality, and Assumption H2.3,

$$
\begin{aligned}
H^{-1/2}E\left(\int_0^v k\left(\frac{t^*(u)-t^*(u_1)}{H}\right)\,dX_u\right) &\le CH^{-1/2}E\left(\int_0^v k\left(\frac{t^*(u)-t^*(u_1)}{H}\right)^2\,du\right)^{\frac{1}{2}}\\
&\le CE\left(\int_0^v H^{-1}k\left(\frac{t^*(u)-t^*(u_1)}{H}\right)^2\,du\right)^{\frac{1}{2}}\\
&\le C',
\end{aligned}
$$

where C and C' denote constant of $O(1)$. Combining together, we have the expected quadratic variation of the last term is bounded by

$$
\begin{aligned}
&= H^{-2}E\int_0^T\left(\int_0^v\left(\int_0^v k\left(\frac{t^*(u)-t^*(u_1)}{H}\right)\,dX_u\right)k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)\,du_1\right)^2\,d\langle X,X\rangle_v\\
&\le C'||\sigma^2||_\infty H^{-1}E\int_0^T\left(\int_0^v k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)\,du_1\right)^2\,dv\\
&\le C'||\sigma^2||_\infty H^{-1}E\int_0^T\left(\int_0^T |k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)|^p\,du_1\right)^{\frac{2}{p}}\,dv\\
&\le C'||\sigma^2||_\infty\left(H^{-1}\right)^{1-\frac{2}{p}}E\int_0^T\left(H^{-1}\int_0^T |k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)|^p\,du_1\right)^{\frac{2}{p}}\,dv,
\end{aligned}
$$

where we use Hölder's inequality in the second to last line for any $p > 1$. By Lemma A.2.1 and taking $1 < p < 2$, the whole term converges to 0 in probability. This concludes that by Eq.

(2.3.4) we obtain

$$H^{-1}\int_0^T \left(\int_0^v k\left(\frac{t^*(u)-t^*(v)}{H}\right) dX_u\right)^2 dv \xrightarrow{p} \int_0^T (b_v - \Psi(v))\sigma_v^2\, dv. \tag{A.1.7}$$

And further leads to

$$\langle H^{-1/2}A_1, H^{-1/2}A_1\rangle_T \xrightarrow{p} \int_0^T (b_v - \Psi(v))\sigma_v^4\, dv. \tag{A.1.8}$$

Putting all pieces together, from Assumption H2.4, we have derived 2 results (Eq. (A.1.3) and Eq. (A.1.8)):

$$\begin{aligned}
\langle H^{-1/2}A_1, X\rangle_T &\xrightarrow{p} \int_0^T a_v\sigma_v^3\, dv,\\
\langle H^{-1/2}A_1, H^{-1/2}A_1\rangle_T &\xrightarrow{p} \int_0^T (b_v - \Psi(v))\sigma_v^4\, dv.
\end{aligned}$$

For the last step, consider an alternative process defined as

$$A_1^* = H^{-1/2}A_1 - \int_0^T a_v\sigma_v\, dX_v. \tag{A.1.9}$$

We have,

$$\begin{aligned}
\langle A_1^*, X\rangle_T &= \langle H^{-1/2}A_1, X\rangle_T - \int_0^T a_v\sigma_v\, d\langle X, X\rangle_v\\
&\xrightarrow{p} 0, && \text{(A.1.10)}\\
\langle A_1^*, A_1^*\rangle_T &= \langle H^{-1/2}A_1, H^{-1/2}A_1\rangle_T + \int_0^T a_v^2\sigma_v^2\, d\langle X, X\rangle_v - 2\int_0^T a_v\sigma_v\, d\langle H^{-1/2}A_1, X\rangle_v\\
&\xrightarrow{p} \int_0^T (b_v - \Psi(v) - a_v^2)\sigma_v^4\, dv. && \text{(A.1.11)}
\end{aligned}$$

Eq. (A.1.10) and Eq. (A.1.11) indicate the following stable convergence in law

$$H^{-1/2}(\bar{C}_T - \langle X, X\rangle_T) - 2\int_0^T a_v\sigma_v\, dX_v \xrightarrow{\mathcal{L}-s} 2\int_0^T \sqrt{b_v - \Psi(v) - a_v^2}\sigma_v^2\, dB_v.$$

This concludes the proof. □

Proof of Theorem 2.3.2. Under the case of no time endogeneity, we have

$$Cov(\Delta X_{t_i}, \Delta t_i) = 0,$$

this indicates $a_v = 0$. Nest we show that $b_v = 2\Psi(v)$. Consider the first term in the decomposition in Eq. (A.1.4),

$$H^{-1} \int_0^T \left(\int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right) dX_u \right)^2 dv = T_1 + T_2,$$

where we use Itô formula,

$$T_1 = H^{-1} \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right)^2 \sigma_u^2 \, du \, dv, \tag{A.1.12}$$

$$T_2 = 2H^{-1} \int_0^T \int_0^v \left(\int_0^{u_1} k\left(\frac{t^*(u) - t^*(v)}{H}\right) \sigma_u \, dW_u^* \right) k\left(\frac{t^*(u_1) - t^*(v)}{H}\right) \sigma_{u_1} \, dW_{u_1}^* \, dv. \tag{A.1.13}$$

We have

$$\begin{aligned} &|H^{-1} \int_0^T \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right)^2 \sigma_u^2 \, du \, dv - \int_0^T \Psi(v) \sigma_v^2 \, dv \\ &\leq \int_0^T |H^{-1} \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right)^2 \sigma_u^2 \, du - \Psi(v)\sigma_v^2| \, dv. \end{aligned}$$

By Assumption H2.3 and the property of the realized kernel,

$$|H^{-1} \int_0^v k\left(\frac{t^*(u) - t^*(v)}{H}\right)^2 \sigma_u^2 \, du - \Psi(v)\sigma_v^2| = o_p(1).$$

This concludes the convergence in probability

$$T_1 \xrightarrow{p} \int_0^T \Psi(v) \sigma_v^2 \, dv.$$

Next, we show that $T_2 \xrightarrow{p} 0$ in probability. Consider

$$\begin{aligned} F(v) &= H^{-1} \int_0^v \left(\int_0^{u_1} k\left(\frac{t^*(u) - t^*(v)}{H} \right) \sigma_u \, dW_u^* \right) k\left(\frac{t^*(u_1) - t^*(v)}{H} \right) \sigma_{u_1} \, dW_{u_1}^* \\ &= H^{-1} \int_0^v F_1(u_1) k\left(\frac{t^*(u_1) - t^*(v)}{H} \right) \sigma_{u_1} \, dW_{u_1}^*, \end{aligned}$$

where

$$F_1(u_1) = \int_0^{u_1} k\left(\frac{t^*(u) - t^*(v)}{H} \right) \sigma_u \, dW_u^*.$$

By Itô isometry,

$$\begin{aligned} E(F_1(u_1)^2) &= \int_0^{u_1} k\left(\frac{t^*(u) - t^*(v)}{H} \right)^2 \sigma_u^2 \, du \\ &\leq ||\sigma^2||_\infty \int_0^{u_1} k\left(\frac{t^*(u) - t^*(v)}{H} \right)^2 \, du, \end{aligned}$$

which is $O(H)$ by Lemma A.2.1. Now, consider

$$\begin{aligned} E(T_2)^2 &= 4E\left(\int_0^T F(v) \, dv \right)^2 \\ &\leq 4||\sigma^2||_\infty \int_0^T \int_0^t E(F(v)F(v')) \, dv \, dv'. \end{aligned}$$

By Itô isometry,

$$\begin{aligned} E(F(v)F(v')) &= H^{-2} \int_0^{v \wedge v'} E\left(F_1(u_1)^2 k\left(\frac{t^*(u_1) - t^*(v)}{H} \right) k\left(\frac{t^*(u_1) - t^*(v')}{H} \right) \sigma_{u_1}^2 \right) du_1 \\ &\leq ||\sigma^2||_\infty H^{-1} \int_0^{v \wedge v'} E\left(H^{-1} F_1(u_1)^2 k\left(\frac{t^*(u_1) - t^*(v)}{H} \right) k\left(\frac{t^*(u_1) - t^*(v')}{H} \right) \right) du_1. \end{aligned}$$

By Fubini's Theorem, and Hölder's inequality

$$
\begin{aligned}
&4E\left(\int_0^T F(v)\,dv\right)^2 \\
\leq &C||\sigma^2||_\infty^2 H^{-1}\int_0^T\int_0^T\left(\int_0^{v\wedge v'} k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)k\left(\frac{t^*(u_1)-t^*(v')}{H}\right)du_1\right)dv\,dv' \\
\leq &C||\sigma^2||_\infty^2 H^{-1}\int_0^T\left(\int_0^T |k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)|\,dv\right)\left(\int_0^T |k\left(\frac{t^*(u_1)-t^*(v')}{H}\right)|\,dv'\right)du_1 \\
= &C||\sigma^2||_\infty^2 H^{-1}\int_0^T\left(\int_0^T |k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)|\,dv\right)^2 du_1 \\
\leq &C||\sigma^2||_\infty^2 H^{-1}\int_0^T\left(\int_0^T |k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)|^p\,dv\right)^{\frac{2}{p}} du_1 \\
= &C||\sigma^2||_\infty^2 (H^{-1})^{1-\frac{2}{p}}\int_0^T\left(\int_0^T H^{-1}|k\left(\frac{t^*(u_1)-t^*(v)}{H}\right)|^p\,dv\right)^{\frac{2}{p}} du_1,
\end{aligned}
$$

for any $p > 1$. By taking $1 < p < 2$, and using Lemma A.2.1, it converges in probability to 0. This concludes that

$$
H^{-1}\int_0^T\left(\int_0^v k\left(\frac{t^*(u)-t^*(v)}{H}\right)dX_u\right)^2 dv \xrightarrow{p} \int_0^T \Psi(v)\sigma_v^2\,dv. \tag{A.1.14}
$$

Comparing Eq. (A.1.14) and (A.1.7) gives the desired result $b_v = 2\Psi(v)$. □

Proof of Theorem 2.3.7. The result follows from Theorem 2.3.2 and Lemma B.5 in [30], as $\frac{R}{N} \to c$ and R, $N \to \infty$, where c is a positive constant

$$
N\int_0^v \left(D(Rt^*(u)-Rt^*(v))\right)^2 du \to \frac{T}{2}(1+2\eta(c)),
$$

this indicates that

$$
\Psi(v) = \frac{cT}{2}(1+2\eta(c)).
$$

□

Proof of Proposition 2.4.1. This is mainly a modification of the Lemma 1 in [58]. And the further development is straightforward by noticing that whether the observation times are endogenous or not, the fastest convergence rate is always of order $N^{1/2}$, hence using blocking method will leads to the uniform convergence on any compact interval. See the proof of Lemma 1 in [58]. □

Proof of Proposition 2.4.2. (1) and (2) follow from Proposition 2.4.1, Assumptions H2.3 and H2.4. (3) follows from Proposition 2.4.1 and Theorem VI.6.22 of [49]. □

A.2 Auxiliary Lemmas

Lemma A.2.1. *Under the Assumptions H2.1 – H2.4, and that $\frac{1}{HN} \to c$ and $H \to 0$, $N \to \infty$, where c is a nonnegative constant, it holds that for $p > 1$*

$$\limsup_{t_j} H^{-1} \sum_{i=1}^{N} |k\left(\frac{t_i - t_j}{H}\right)|^p (t_i - t_{i-1}) \leq C,$$

where C is a suitable constant.

Proof. Recall that if a positive valued function $g(x)$: $\mathbb{R} \to \mathbb{R}^+$, decreasing for $x > 0$ and increasing for $x < 0$, and if $s_0 < s_1 < \cdots < s_N$ is a finite sequence of real numbers such that $s_i - s_{i-1} < r$ for all i, then

$$\sum_{i=1}^{N} g(s_i)(s_i - s_{i-1}) \leq rg(0) + \int_{-\infty}^{\infty} g(x)dx. \tag{A.2.1}$$

consider g as before with $g(0) = 1$ and $|g| \geq |k|^p$ for $p > 1$. Let $s_i = \frac{t_i - t_j}{H}$, $r = \max_i(\Delta t_i)/H$ and $C_p = \int_{-\infty}^{\infty} g(x)dx$. Then

$$\begin{aligned}\left(\max_i(\Delta t_i)\right)^{-1} \sum_{i=1}^{N} |k\left(\frac{t_i - t_j}{H}\right)|^p (t_i - t_{i-1}) &\leq 1 + \frac{HC_p}{\max_i(\Delta t_i)} \\ H^{-1} \sum_{i=1}^{N} |k\left(\frac{t_i - t_j}{H}\right)|^p (t_i - t_{i-1}) &\leq \max_i(\Delta t_i)/H + C_p\end{aligned}$$

□

Appendix B

Supplemental Material for Chapter 3

We prove all the Theorems and Propositions in the appendix. Throughout this section, we follow the method presented in [63] by using Girsanov's Theorem to suppress μ_t. We considered a probability measure P^*, under which X_t is a local martingale

$$dX_t = \sigma_t \, dW_t^*,$$

where W_t^* is a P^*-Brownian motion, with

$$dW_t^* = dW_t + \sigma_t^{-1}\mu_t \, dt.$$

B.1 Proofs

Proof of Theorem 3.2.1. See [29]. □

Proof of Proposition 3.3.1. For simplicity, in the proof, we only consider the terms that are not affected by the end points, terms involving the end points are of smaller order due to the effect of jittering. In the proof below, we use $\sum_i$ to denote $\sum_{i=1}^{N-1}$ for notation simplicity.

$$\begin{aligned} E(B) &= E\left(\sum_i \sum_j k\left(\frac{t_i - t_j}{H}\right) \Delta X_{t_i}(U_{t_{j+1}} - U_{t_j})\right) \\ &= E\left(\sum_i \sum_j \left[k\left(\frac{t_i - t_{j-1}}{H}\right) - k\left(\frac{t_i - t_j}{H}\right)\right] \Delta X_{t_i} U_{t_j}\right) = 0, \end{aligned}$$

and

$$
\begin{aligned}
HNE(B^2) &= HN \times E\left(\sum_i\sum_j\left[k\left(\frac{t_i - t_{j-1}}{H}\right) - k\left(\frac{t_i - t_j}{H}\right)\right]^2 \Delta X_{t_i}^2 U_{t_j}^2\right) \\
&= HN \times E\left(\sum_i\sum_j\left[k'\left(\frac{t_i - t_{j-1}}{H}\right)\frac{\Delta t_{j-1}}{H} + o_p(\frac{1}{HN})\right]^2 \Delta X_{t_i}^2 U_{t_j}^2\right),
\end{aligned}
$$

we focus on the leading term, and

$$
\begin{aligned}
&HN \times E\left(\sum_i\sum_j\left[k'\left(\frac{t_i - t_{j-1}}{H}\right)\right]^2 \frac{\Delta t_{j-1}^2}{H^2}\Delta X_{t_i}^2 U_{t_j}^2\right) \\
=&\epsilon^2 HN\sum_i\sum_j\left[k'\left(\frac{t_i - t_{j-1}}{H}\right)\right]^2 H^{-2}E(\Delta t_{j-1}^2)E(\Delta X_{t_i}^2) \\
&+ \epsilon^2 HNk'(0)^2 H^{-2}\sum_i cov(\Delta t_i^2, \Delta X_{t_i}^2) \\
=&\epsilon^2 TE\left(\int_0^T\left(H^{-1} \times \frac{N}{T}\int_0^T\left[k'\left(\frac{t^*(i) - t^*(j)}{H}\right)\right]^2 d(j - t^*(j))^2\right) d(X_i - X_{t^*(i)})^2\right) \\
=&\epsilon^2 TE\left(\int_0^T\left(H^{-1} \times \frac{N}{T}\int_0^i\left[k'\left(\frac{t^*(i) - t^*(j)}{H}\right)\right]^2 d(j - t^*(j))^2\right) d(X_i - X_{t^*(i)})^2\right) \\
&+ \epsilon^2 TE\left(\int_0^T\left(H^{-1} \times \frac{N}{T}\int_i^T\left[k'\left(\frac{t^*(i) - t^*(j)}{H}\right)\right]^2 d(j - t^*(j))^2\right) d(X_i - X_{t^*(i)})^2\right) \\
&\xrightarrow{p} 2\epsilon^2 T\left(\int_0^\infty k'(t)^2\, dH(t)\right)\langle X, X\rangle_T,
\end{aligned}
$$

in the last line, we use the the change of variables formula, definition of the "quadratic variation of time" in [64] and that $E(\int_0^T d(X_i - X(t_i^*)^2) \xrightarrow{p} \langle X, X\rangle_T$ (see [58] for this under time endogeneity). It is also important here to notice that due to $k'(0) = 0$, so that the covariance between $\Delta X_{t_i}^2$ and Δt_{j-1}^2 whenever $i = j - 1$ has been assigned zero weight, so that we can treat them as if there is no time endogeneity from the third line. Under Assumption H3.5, the convergence in distribution follows from the usual martingale central limit theorem. The stable convergence follows from Proposition 5 and Lemma 1 in [16]. □

Proof of Proposition 3.3.2.

$$
\begin{aligned}
C &= \sum_i \sum_j k\left(\frac{t_i - t_j}{H}\right)(U_{t_{i+1}} - U_{t_i})(U_{t_{j+1}} - U_{t_j}) \\
&= \sum_i \sum_j \underbrace{\left[k\left(\frac{t_i - t_j}{H}\right) + k\left(\frac{t_{i-1} - t_{j-1}}{H}\right) - k\left(\frac{t_{i-1} - t_j}{H}\right) - k\left(\frac{t_i - t_{j-1}}{H}\right)\right]}_{k(i,j)} U_{t_i} U_{t_j} \qquad \text{(B.1.1)} \\
&= \sum_i \sum_j \left[k'\left(\frac{t_{i-1} - t_j}{H}\right)\frac{\Delta t_{i-1}}{H} - k'\left(\frac{t_{i-1} - t_{j-1}}{H}\right)\frac{\Delta t_{i-1}}{H} + o_p(\frac{1}{HN})\right] U_{t_i} U_{t_j} \\
&= -\sum_i \sum_j \left[k''\left(\frac{t_{i-1} - t_{j-1}}{H}\right)\frac{\Delta t_{i-1}\Delta t_{j-1}}{H^2} + o_p(\frac{1}{HN})\right] U_{t_i} U_{t_j},
\end{aligned}
$$

we focus on the leading terms,

$$
\begin{aligned}
&H^2 N \times E\left(\sum_i \sum_j \left[k''\left(\frac{t_{i-1} - t_{j-1}}{H}\right)\frac{\Delta t_{i-1}\Delta t_{j-1}}{H^2}\right] U_{t_i} U_{t_j}\right) \\
=&H^2 N \times E\left(\sum_i \left[k''(0)\frac{\Delta t_{i-1}^2}{H^2}\right] U_{t_i}^2\right) \\
=&\epsilon^2 k''(0) N \int_0^T d(i - t^*(i))^2 \xrightarrow{p} \epsilon^2 k''(0) T H(T).
\end{aligned}
$$

Next consider

$$
H^4 N^2 \times E(C^2) = H^4 N^2 \times E\left(\sum_i \sum_j \left[k''\left(\frac{t_{i-1} - t_{j-1}}{H}\right)\frac{\Delta t_{i-1}\Delta t_{j-1}}{H^2} + o_p(\frac{1}{HN})\right] U_{t_i} U_{t_j}\right)^2,
$$

again, we focus on the leading term,

$$
\begin{aligned}
&H^4N^2 \times E\left(\sum_i\sum_j\left[k''\left(\frac{t_{i-1}-t_{j-1}}{H}\right)\frac{\Delta t_{i-1}\Delta t_{j-1}}{H^2}\right]U_{t_i}U_{t_j}\right)^2\\
=&2H^4N^2E\left(\sum_{i\neq j}\left[k''\left(\frac{t_{i-1}-t_{j-1}}{H}\right)^2\frac{\Delta t_{i-1}^2\Delta t_{j-1}^2}{H^4}\right]U_{t_i}^2U_{t_j}^2\right)\\
&+H^4N^2E\left(\sum_i\left[k''(0)^2\frac{\Delta t_{i-1}^4}{H^4}\right]U_{t_i}^4\right)+H^4N^2E\left(\sum_{i\neq j}k''(0)^2\frac{\Delta t_{i-1}^2\Delta t_{j-1}^2}{H^4}U_{t_i}^2U_{t_j}^2\right)\\
=&\underbrace{2H^4N^2E\left(\sum_i\sum_j\left[k''\left(\frac{t_{i-1}-t_{j-1}}{H}\right)^2\frac{\Delta t_{i-1}^2\Delta t_{j-1}^2}{H^4}\right]\epsilon^4\right)}_{C_1}\\
&+\underbrace{H^4N^2E\left(\sum_i\left[k''(0)^2\frac{\Delta t_{i-1}^4}{H^4}\right]\left(U_{t_i}^4-3\epsilon^4\right)\right)}_{C_2}+\underbrace{H^4N^2E\left(\sum_i\sum_j\left[k''(0)^2\frac{\Delta t_{i-1}^2\Delta t_{j-1}^2}{H^4}\right]\epsilon^4\right)}_{C_3}.
\end{aligned}
$$

First, we consider C_1,

$$
\begin{aligned}
C_1 &= 2H^4N^2E\left(\sum_i\sum_j\left[k''\left(\frac{t_{i-1}-t_{j-1}}{H}\right)^2\frac{\Delta t_{i-1}^2\Delta t_{j-1}^2}{H^4}\right]\epsilon^4\right)\\
&= 2T^2H\epsilon^4E\left(\frac{N}{T}\int_0^T H^{-1}\frac{N}{T}\int_0^T k''\left(\frac{t^*(i)-t^*(j)}{H}\right)^2 d(j-t^*(j))^2\,d(i-t^*(i))^2\right)\\
&= 2T^2H\epsilon^4E\left(\frac{N}{T}\int_0^T H^{-1}\frac{N}{T}\int_0^i k''\left(\frac{t^*(i)-t^*(j)}{H}\right)^2 d(j-t^*(j))^2\,d(i-t^*(i))^2\right)\\
&\quad+2T^2H\epsilon^4E\left(\frac{N}{T}\int_0^T H^{-1}\frac{N}{T}\int_i^T k''\left(\frac{t^*(i)-t^*(j)}{H}\right)^2 d(j-t^*(j))^2\,d(i-t^*(i))^2\right)\\
&\xrightarrow{p} 4T^2H\epsilon^4H(T)\left(\int_0^\infty k''(t)^2\,dH(t)\right)\xrightarrow{p}0,
\end{aligned}
\tag{B.1.2}
$$

where we use the change of variables formula. Since $C_2 = O_p(\frac{1}{N})$, we next consider the effect of

C_3,

$$C_3 = H^4 N^2 E\left(\sum_i \sum_j \left[k''(0)^2 \frac{\Delta t_{i-1}^2 \Delta t_{j-1}^2}{H^4}\right]\epsilon^4\right)$$
$$= \epsilon^4 k''(0)^2 N^2 \int_0^T \int_0^T d(i - t^*(i))^2\, d(j - t^*(j))^2 \xrightarrow{p} \epsilon^4 k''(0)^2 T^2 H(T)^2.$$

Combining all above together,

$$Var\left(H^2 NC\right) = H^4 N^2 E(C^2) - H^4 N^2 E(C)^2 = O_p(H) + O_p(\frac{1}{N}),$$

the consistency follows from Chebyshev's inequality. □

Proof of Theorem 3.3.3. The result follows from Theorem 3.2.1, Proposition 3.3.1 and Proposition 3.3.2. □

Proof of Theorem 3.4.1. We first show that $H^{-1/2}(A - A_F) \xrightarrow{p} 0$, as $\frac{1}{H^2 N} \to c$ as $H \to 0$ and $N \to \infty$, where c is a nonnegative constant.

$$H^{-1/2}(A - A_F) = 2H^{-1/2}\left(\sum_i \left[k\left(\frac{t_i - t_{i-1}}{H}\right) - 1\right]\Delta X_{t_i}\Delta X_{t_{i-1}}\right)$$
$$= 2H^{-1/2}\left(\sum_i \left[k'(0)\frac{\Delta t_{i-1}}{H} + \frac{1}{2}k''(0)\frac{\Delta t_{i-1}^2}{H^2} + o_p(\frac{1}{H^2N^2})\right]\Delta X_{t_i}\Delta X_{t_{i-1}}\right),$$

we focus on the $k''(0)$ term, due to $k'(0) = 0$,

$$H^{-1/2}\left(\sum_i k''(0)\frac{\Delta t_{i-1}^2}{H^2}\Delta X_{t_i}\Delta X_{t_{i-1}}\right) \le H^{-5/2}\max_i(\Delta t_i)^2 k''(0)\sum_i |\Delta X_{t_i}\Delta X_{t_{i-1}}|$$
$$= O_p(\frac{1}{H^{5/2}N^2}) \xrightarrow{p} 0.$$

Nest we show that $H^{-1/2}(B - B_F) \xrightarrow{p} 0$, as $\frac{1}{H^2 N} \to c$ as $H \to 0$ and $N \to \infty$, where c is a nonnegative constant.

$$
\begin{aligned}
&H^{-1/2}(B - B_F) \\
&= \underbrace{H^{-1/2}\left(\sum_i \left[k\left(\frac{t_i - t_{i-1}}{H}\right) - 1\right]\Delta X_{t_i} U_{t_i}\right)}_{B_{F1}} - \underbrace{H^{-1/2}\left(\sum_i \left[k\left(\frac{t_i - t_{i-1}}{H}\right) - 1\right]\Delta X_{t_i} U_{t_{i-1}}\right)}_{B_{F2}} \\
&+ \underbrace{H^{-1/2}\left(\sum_i \left[k\left(\frac{t_i - t_{i+1}}{H}\right) - 1\right]\Delta X_{t_i} U_{t_{i+1}}\right)}_{B_{F3}} - \underbrace{H^{-1/2}\left(\sum_i \left[k\left(\frac{t_i - t_{i+1}}{H}\right) - 1\right]\Delta X_{t_i} U_{t_i}\right)}_{B_{F4}},
\end{aligned}
$$

we show that $B_{F1} \xrightarrow{p} 0$, the convergence of other terms can be proved analogously.

$$
\begin{aligned}
B_{F1} &= H^{-1/2}\left(\sum_i \left[k\left(\frac{t_i - t_{i-1}}{H}\right) - 1\right]\Delta X_{t_i} U_{t_i}\right) \\
&= H^{-1/2}\left(\sum_i \left[\frac{1}{2}k''(0)\frac{\Delta t_{i-1}^2}{H^2} + o_p(\frac{1}{H^2N^2})\right]\Delta X_{t_i} U_{t_i}\right),
\end{aligned}
$$

focusing on the main term, we have

$$
E(B_{F1}^2) \leq \frac{1}{4H^5}k''(0)^2\epsilon^2 \max_i(\Delta t_i)^4 E\left(\sum_i \Delta X_{t_i}^2\right) = O_p(\frac{1}{H^5N^4}).
$$

Last, we work on C_F,

$$
\begin{aligned}
C_F &= \sum_i \sum_j k^F\left(\frac{t_i - t_j}{H}\right)(U_{t_{i+1}} - U_{t_i})(U_{t_{j+1}} - U_{t_j}) \\
&= \sum_i \sum_j \underbrace{\left[k^F\left(\frac{t_i - t_j}{H}\right) + k^F\left(\frac{t_{i-1} - t_{j-1}}{H}\right) - k^F\left(\frac{t_{i-1} - t_j}{H}\right) - k^F\left(\frac{t_i - t_{j-1}}{H}\right)\right]}_{k^F(i,j)} U_{t_i} U_{t_j},
\end{aligned}
$$

notice $k^F(i,i) = 0$, so that $E(C_F) = 0$. Next consider

$$\begin{aligned} H^{-1}C_F^2 &= 2H^{-1}\sum_{i\neq j} k^F(i,j)^2 U_{t_i}^2 U_{t_j}^2 + H^{-1}\sum_{i\neq j} k^F(i,i)k^F(j,j)U_{t_i}^2 U_{t_j}^2 + H^{-1}\sum_i k^F(i,i)^2 U_{t_i}^4 \\ &= 2H^{-1}\sum_{i\neq j} k^F(i,j)^2 U_{t_i}^2 U_{t_j}^2, \end{aligned}$$

consider the following term and also noting that when $|i-j| \notin \{0,1,2\}$, $k^F(i,j) = k(i,j)$, and when $|i-j| \in \{0,1,2\}$, $k^F(i,j) - k(i,j) = O_p(\frac{1}{H^2N^2})$ using the second order Taylor expansion, where $k(i,j)$ is defined in Eq. (B.1.1),

$$\begin{aligned} E\left(H^{-1}\sum_{i\neq j} k^F(i,j)^2 U_{t_i}^2 U_{t_j}^2 - H^{-1}\sum_i\sum_j k(i,j)^2\epsilon^4\right) &= H^{-1}\epsilon^4 E\left(\sum_i\sum_j k^F(i,j)^2 - k(i,j)^2\right) \\ &= O_p(\frac{1}{H^5N^3}) \xrightarrow{p} 0. \end{aligned}$$

As shown in Eq. (B.1.2), we have

$$E\left(H^{-1}\sum_i\sum_j k(i,j)^2\epsilon^4\right) \xrightarrow{p} 4c^2T^2\epsilon^4 H(T)\left(\int_0^\infty k''(t)^2\,dH(t)\right).$$

The convergence in distribution follows from the usual martingale central limit theorem, Theorem 3.2.1 and Proposition 3.3.1, the stable convergence result follows by Lemma 1 and Proposition 5 in [16]. □

Proof of Theorem 3.4.2. The result follows from Theorem 3.2.1, Eq. (3.4.6) and (3.4.7). □

Appendix C

Supplemental Material for Chapter 4

Since we use the theory of exterior product throughout the proof, Appendix C.1 provides a minimum background, proofs are given in Appendix C.2.

C.1 Exterior Differential Form

A multiple integral

$$A = \int_{\Delta} f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \tag{C.1.1}$$

where $\Delta \subset \mathbb{R}^n$. By the classical change of variables formula, if $x = x(y)$, we have

$$A = \int_{\Delta'} f(x(y)) det(\frac{\partial x_i}{\partial y_j})\, dy_1 \ldots dy_n \tag{C.1.2}$$

where Δ' is the image of Δ. Now substitute dx_i in Eq. (C.1.1) with its differential

$$dx_i = \frac{\partial x_i}{\partial y_1}\, dy_1 + \ldots + \frac{\partial x_i}{\partial y_n}\, dy_n$$

we have

$$A = \int_{\Delta'} f(x(y))(\frac{\partial x_1}{\partial y_1}\, dy_1 + \ldots + \frac{\partial x_1}{\partial y_n}\, dy_n) \ldots (\frac{\partial x_n}{\partial y_1}\, dy_1 + \ldots + \frac{\partial x_n}{\partial y_n}\, dy_n) \tag{C.1.3}$$

The equality between the left hand sides of equation Eq. (C.1.2) and (C.1.3) indicates their right hand sides should be the same. For simplicity, we assume $n = 2$, we will need to have

$$(\frac{\partial x_1}{\partial y_1}\frac{\partial x_2}{\partial y_2} - \frac{\partial x_1}{\partial y_2}\frac{\partial x_2}{\partial y_1})\, dy_1\, dy_2 = (\frac{\partial x_1}{\partial y_1}\, dy_1 + \frac{\partial x_1}{\partial y_2}\, dy_2)(\frac{\partial x_2}{\partial y_1}\, dy_1 + \frac{\partial x_2}{\partial y_2}\, dy_2) \tag{C.1.4}$$

To equate left and right hand side of the equation above, we must have $dy_1\, dy_2 = -dy_2\, dy_1$, or more generally $dy_i\, dy_j = -dy_j\, dy_i$. Such an anti-commutative product is called the exterior product or the wedge product and is denoted by the symbol $\wedge$, such that $dy_i \wedge dy_j = -dy_j \wedge dy_i$. In particular, $dy_i \wedge dy_i = -dy_i \wedge dy_i = 0$. The following lemma justifies this formal procedure.

Lemma C.1.1. *If dy is an $n \times 1$ vector of differentials and if $dy = B\, dx$, where B is a $n \times n$ nonsingular matrix, then*

$$\bigwedge_{i=1}^{n} dy_i = det(B) \bigwedge_{i=1}^{n} dx_i$$

Proof. See [51]. □

Following the convention, for an arbitrary $n \times m$ matrix X, we denote (dX) as the wedge product of the mn elements of dx_{ij}

$$(dX) = \bigwedge_{j=1}^{m} \bigwedge_{i=1}^{n} dx_{ij}$$

If X is a symmetric $m \times m$ matrix, (dX) denotes the wedge product of the $\frac{m(m+1)}{2}$ distinct element of dx_{ij}:

$$(dX) = \bigwedge_{1 \le i \le j \le m} dx_{ij}$$

If X is a skew symmetric matrix with X = -X', (dX) denotes the exterior product of the $\frac{m(m-1)}{2}$ distinct elements of dX, which could be either sub-diagonal or super-diagonal.

If X is an upper triangular $m \times m$ matrix, (dX) denotes the wedge product of the $\frac{m(m+1)}{2}$ distinct element of dx_{ij}[1]:

$$(dX) = \bigwedge_{1 \le i \le j \le m} dx_{ij}$$

Before calculating the determinants of the Jacobian matrices, it is important to make it explicit that it is unnecessary to keep track of the sign, since the main usage of them in this paper

[1]Lower triangular matrix is defined similarly.

is to integrate differential forms representing probability density functions. Next we summarize the important results use throughout the paper.

Proposition C.1.2. *If $Y = BX$, where X and Y are $n \times m$ matrices and B is a fixed nonsingular $n \times n$ matrix, then*

$$(dY) = det(B)^m (dX) \tag{C.1.5}$$

Proof. for each column X_j of X, using Lemma C.1.1, $(dY_j) = det(B)(dX_j)$, thus $(dY) = \bigwedge_{j=1}^{m}(dY_j) = \bigwedge_{j=1}^{m} det(B)(dX_j) = det(B)^m(dX)$. □

Proposition C.1.3. *If $Y = BXB'$, where X and Y are $n \times n$ symmetric matrices and B is a nonsingular $n \times n$ matrix, then $(dY) = \mid B \mid^{n+1}(dX)$.*

Proof. See [62]. □

Corollary C.1.4. *If $Y = BXB'$, where X and Y are $n \times n$ skew-symmetric matrices and B is a nonsingular $n \times n$ matrix, then $(dY) = \mid B \mid^{n-1}(dX)$.*

Proof. See [62]. □

Proposition C.1.5. *For the inverse matrix transformation $Y = X^{-1}$, we have $(dY) = \mid X \mid^{-2n}(dX)$ when X has no special structure. For the inverse matrix transformation $Y = X^{-1}$ with X symmetric, we have $(dY) = \mid X \mid^{-(n+1)}(dX)$.*

Proof. For an arbitrary $n \times n$ matrix $Y = X^{-1}$. $\frac{(dY)}{(dX)} = \mid \frac{dvec(Y)}{dvec(X)} \mid$.

The matrix function is $Y = X^{-1}$, $dY = -X^{-1}\, dX\, X^{-1}$. Applying the vec operator, we have $dvecY = -((X')^{-1} \otimes X^{-1})\, dvec(X)$, hence $\frac{(dY)}{(dX)} = \mid -((X')^{-1} \otimes X^{-1}) \mid = \mid X \mid^{-2n}$.

If X is symmetric and so is Y, $dY = -X^{-1}\, dX\, X^{-1}$, we can apply Proposition C.1.3. □

Proposition C.1.6. *If Y is an $n \times n$ positive definite matrix and X is its Cholesky decomposition lower triangular matrix with positive diagonal elements, $Y = XX'$, then $(dY) = 2^n \prod_{i=1}^{n} x_{ii}^{n+1-i}(dX) = \mid L(I_{n^2} + K_{nn})(X \otimes I_n)L' \mid (dX)$.*

Proof. The result follows directly from the special structure of Cholesky decomposition.

$$(dY) = \bigwedge_{i \le j \le n} dy_{ij} = 2^n x_{11}^{n} x_{22}^{n-1} \dots x_{nn} \bigwedge_{i \le j \le n} dx_{ij}$$

Another expression may be preferred for simplicity, $\frac{(dY)}{(dX)} = \mid \frac{dvech(Y)}{dvech(X)} \mid$. $dY = dX(X') + (X)dX'$, hence $dvec(Y) = [(I_{n^2} + K_{nn})(X \otimes I_n)]\, dvec(X)$, where K_{nn} is the commutation matrix.

Since X is lower triangular, we have $dvec(X) = L'\,dvech(X)$, where L is the elimination matrix such that $Lvec(\cdot) = vech(\cdot)$ for any square matrix. Thus $\frac{(dY)}{(dX)} = \mid L(I_{n^2} + K_{nn})(X \otimes I_n)L' \mid^2$. $\square$

Corollary C.1.7. *If Y is an $n \times n$ positive definite matrix and X is a lower triangular matrix with positive diagonal elements, and $Y = X'X$, then $(dY) = 2^n \prod_{i=1}^n x_{ii}^i (dX) = \mid L(I_{n^2} + K_{nn})(I_n \otimes X)L' \mid (dX)$.*

Proof. $\frac{(dY)}{(dX)} = \mid \frac{dvech(Y)}{dvech(X)} \mid$. $dY = dX'\,(X) + (X')\,dX$, hence $dvec(Y) = [(I_{n^2} + K_{nn})(I_n \otimes X)]\,dvec(X)$, we have $dvec(X) = L'\,dvech(X)$. Thus $\frac{(dY)}{(dX)} = \mid L(I_{n^2} + K_{nn})(I_n \otimes X)L' \mid$. $\square$

Proposition C.1.8. *Let Y be an $n \times n$ matrix of rank n, and its QR decomposition is $Y = QR$, where $Q = [q_1 \ldots q_n]$ is an $n \times n$ orthogonal matrix, and R is an $n \times n$ upper triangular matrix with positive diagonal elements. Then $(dY) = \prod_{i=1}^n R_{ii}^{n-i} (Q'\,dQ)(dR)$.*

Proof. See [62]. $\square$

Corollary C.1.9. *Let X be an $n \times n$ matrix of rank n, and its QR decomposition is $X = QR$, where $Q = [q_1 \ldots q_n]$ is an $n \times n$ orthogonal matrix, and R is an $n \times n$ upper triangular matrix with positive diagonal elements. $Y = X'X$, then $(dX) = 2^{-n} \mid Y \mid^{-\frac{1}{2}} (Q'\,dQ)(dY)$.*

Proof. From Proposition C.1.6, we have $(dY) = 2^n \prod_{i=1}^n R_{ii}^{n+1-i}(dR)$. From Proposition C.1.8 we have $(dX) = \prod_{i=1}^n R_{ii}^{n-i}(dR)(Q'\,dQ)$. Combine together, $(dX) = 2^{-n} \mid Y \mid^{-\frac{1}{2}} (dY)(Q'\,dQ)$. $\square$

C.2 Proofs

Before we prove Proposition 4.3.1, it is important to note that the problem of defining a uniformly distributed measure on the orthogonal group boils down to finding an invariant exterior differential form according to the following theorem.

Theorem C.2.1. *If a differential form is invariant, the corresponding measure is invariant.*

Proof. Suppose μ is a measure on a manifold $\mathcal{T}$ that corresponds to the differential form $\omega(x)$,

$$\mu(T) = \int_T \omega(x)$$

and f is an analytic homeomorphism that maps manifold $\mathcal{T}$ on itself, i.e., for a set $T \subset \mathcal{T}$ onto a set $\bar{T} = f(T) \subset \mathcal{T}$. Furthermore, f induces maps of μ on $\bar{\mu}$, and $\omega(x)$ on $\bar{\omega}(y)$.

$$\bar{\mu}(\bar{T}) = \mu(f^{-1}(\bar{T}))$$

[2]Notice one can not use chain rule to get the result since $\frac{1}{dvec(X)}$ is not properly defined.

Without loss of generality, suppose the differential form $\omega(x) = g(x)\ dx_1 \wedge \ldots \wedge dx_n$, then follows from Lemma C.1.1 we have,

$$\bar{\omega}(y) = g(f^{-1}(y)) det(\frac{\partial x_i}{\partial y_j})\, dy_1 \wedge \ldots \wedge dy_n$$

By the change of variables formula, we have

$$\bar{\mu}(\bar{T}) = \int_{\bar{T}} \bar{\omega}(y) = \int_{f^{-1}(\bar{T})} \omega(x)$$

If the differential form ω is invariant under f, i.e., $\bar{\omega}(y) = \omega(y)$, then we have

$$\int_{f^{-1}(\bar{T})} \omega(x) = \int_{\bar{T}} \bar{\omega}(y) = \int_{\bar{T}} \omega(y)$$

Thus measure μ is invariant under the transformation f, i.e., $\mu(\bar{T}) = \mu(f^{-1}(\bar{T}))$. $\square$

Proof of Proposition 4.3.1. First we show that $(P'\, dP)$ is the invariant exterior differential product on the orthogonal group by showing that $(P'\, dP)$ is invariant under orthogonal transformation.

It is easy to check that it is invariant under left and right translation[3],

$$(P'\, dP) = (P'Q'Q\, dP) = (QP'\, dP\, Q')$$

The first equality holds because Q is orthogonal and $Q'Q = I_n$. The second equality follows from Corollary C.1.4. By Theorem C.2.1, $(P'\, dP)$ defines a measure on $O(n)$, and this measure is the invariant measure or the Haar measure.

Next we show that the integration of this measure over the orthogonal group sums up to 1. Consider an X that is n-rowed and n-columned, and each element X_{ij} follows an independent standard normal distribution. The density function of X is given by Eq. (4.3.2),

$$(2\pi)^{-n^2/2} etr\{-\frac{1}{2}X'X\}$$

the integration of this density sums up to 1. Now make the change of variables formula $X =$

[3]Here P and $Q \in O(n)$.

PR' using Proposition C.1.8 $(dX) = \prod_{i=1}^{m} R_{ii}^{n-i}(dR)(P'\,dP)$, we have

$$
\begin{aligned}
(2\pi)^{n^2/2} &= \int etr\{-\frac{1}{2}X'X\}(dX) \\
&= \int etr\{-\frac{1}{2}RR'\}\prod_{i=1}^{n} R_{ii}^{n-i}(dR)\int_{O(n)}(P'\,dP) \\
&= \int exp(-\frac{1}{2}\sum_{i\geq j}^{n} R_{ij}^2)\prod_{i=1}^{n} R_{ii}^{n-i}(dR)\int_{O(n)}(P'\,dP) \qquad \text{(C.2.1)}
\end{aligned}
$$

Integrating (dR) term by term, we have

$$
\begin{aligned}
& \int exp(-\frac{1}{2}\sum_{i\geq j}^{n} R_{ij}^2)\prod_{i=1}^{n} R_{ii}^{n-i}(dR) \\
= & \int \ldots \int exp(-\frac{1}{2}\sum_{i\leq j}^{n} R_{ij}^2)\prod_{i=1}^{n} R_{ii}^{n-i}(\bigwedge_{i\geq j}^{n} dR_{ij}) \\
= & \prod_{i>j}^{n}[\int exp(-\frac{1}{2}R_{ij}^2)dR_{ij}]\prod_{i=1}^{n}[\int_0^{\infty} exp(-\frac{1}{2}R_{ii}^{n-i+2})dR_{ii}] \\
= & 2^{n^2/2-n}\Gamma_n(\frac{1}{2}n) \qquad \text{(C.2.2)}
\end{aligned}
$$

Substituting Eq. (C.2.2) into (C.2.1), we have

$$\int_{On)}(P'\,dP) = \frac{2^n\pi^{n^2/2}}{\Gamma_n(\frac{1}{2}n)} \qquad \text{(C.2.3)}$$

Using Eq. (C.2.3), the integration of Eq. (4.3.1) over $O(n)$ sums up to 1. □

Proof of Proposition 4.3.3. Since X has spherical distribution, its density function takes the form $g(X'X)$. Then it follows from Proposition C.1.8 $(dX) = \prod_{i=1}^{m} R_{ii}^{n-i}(dR)(Q'\,dQ)$ and the change of variables formula,

$$
\begin{aligned}
g(X'X)(dX) &= g(RR')\prod_{i=1}^{m} R_{ii}^{n-i}(dR)(Q'\,dQ) \\
&= [\frac{2^m\pi^{mn/2}}{\Gamma_m(\frac{1}{2}n)}\prod_{i=1}^{m} R_{ii}^{n-i}g(RR')(dR)][\frac{\Gamma_m(\frac{1}{2}n)}{2^m\pi^{mn/2}}(Q'\,dQ)] \qquad \text{(C.2.4)}
\end{aligned}
$$

From Eq. (C.2.4) Q is independent of R, and Q is uniformly distributed on $O(n)$. □

Proof of Theorem 4.3.4. Here we prove Eq. (4.3.3), (4.3.4), (4.3.5), the rest follows by the chain rule.

1. The Jacobian matrix between the reduced-form parameterization and the reduced-form MA parameterization

   Since $\Psi_s = B_1\Psi_{s-1} + B_2\Psi_{s-2} + \ldots + B_p\Psi_{s-p}$, with $\Psi_0 = I_n$ and $\Psi_s = \mathbf{0}$ for s < 0. We have

$$\begin{aligned}
d\Psi_1 &= dB_1 \\
d\Psi_2 &= dB_2 + \text{term involving } dB_1 \\
d\Psi_3 &= dB_3 + \text{term involving } dB_1 \text{ and } dB_2 \\
&\vdots \\
d\Psi_p &= dB_p + \text{term involving } dB_1,\, dB_2 \ldots dB_{p-1}
\end{aligned}$$

   Taking the exterior product on both sides of the above equations[4], we have

$$(d\Psi_1)\ldots(d\Psi_p) = (dB_1)\ldots(dB_p)$$

   which indicates that

$$(d\Psi_1)\ldots(d\Psi_p)(d\Sigma)(P'\,dP) = (dB_1)\ldots(dB_p)(d\Sigma)(P'\,dP)$$

   This proves Eq. (4.3.3).

2. The Jacobian matrix between the reduced-form parameterization and the structural parameterization

   The mapping between them is given by $A_0 = P\Sigma_{ch}^{-1}$, $A_i = P\Sigma_{ch}^{-1}B_i$.

$$\begin{aligned}
det(\frac{\partial A_0, A_1, \ldots, A_p}{\partial P\Sigma_{ch}^{-1}, B_1, \ldots, B_p}) &= \mid \frac{\partial vec(A_0, A_1, \ldots, A_p)}{\partial vec(P\Sigma_{ch}^{-1}, B_1, \ldots, B_p)} \mid \\
&= det(I_{np} \otimes P\Sigma_{ch}^{-1}) \\
&= \mid \Sigma \mid^{-\frac{np}{2}}
\end{aligned}$$

[4]Recall that exterior product is anti-commutative, so that $dB_i \wedge dB_i = -dB_i \wedge dB_i = 0$.

From Proposition C.1.5, we have

$$det(\frac{\partial P\Sigma_{ch}^{-1}}{\partial \Sigma_{ch}P'}) = \mid \Sigma \mid^{-n}$$

From Corollary C.1.9,

$$det(\frac{\partial \Sigma_{ch}P'}{\partial \Sigma, P}) = 2^{-n} \mid \Sigma \mid^{-\frac{1}{2}}$$

Combining all these together, we have

$$det(\frac{\partial A_0, A_1, \ldots, A_p}{\partial B_1, \ldots, B_p, \Sigma, P}) = 2^{-n} \mid \Sigma \mid^{-\frac{np+2n+1}{2}}$$

3. The Jacobian matrix between the reduced-form MA parameterization and the IRF parameterization

   The mapping between them is given by $IRF^0 = A_0^{-1} = \Sigma_{ch}P'$, and $IRP^i = \Psi_i \Sigma_{ch} P'$.

$$\begin{aligned} det(\frac{\partial IRF^0, IRF^1, \ldots, IRF^p}{\partial \Sigma_{ch}P', \Psi_1, \ldots, \Psi_p}) &= det((\Sigma_{ch}P')' \otimes I_{np}) \\ &= \mid \Sigma_{ch}P' \mid^{np} \\ &= \mid \Sigma \mid^{\frac{np}{2}} \end{aligned}$$

   From Corollary C.1.9, $det(\frac{\partial \Sigma_{ch}P'}{\partial \Sigma, P}) = 2^{-n} \mid \Sigma \mid^{-\frac{1}{2}}$. Thus

$$det(\frac{\partial IRF^0, IRF^1, \ldots, IRF^p}{\partial \Psi_1, \ldots, \Psi_p, \Sigma, P}) = 2^{-n} \mid \Sigma \mid^{\frac{np-1}{2}}$$

□