A Shape Restricted Nonparametric Method with Financial Application

by

X.J. Jeng
Purdue University

P.A. Mykland
University of Chicago

Technical Report #08-08

Department of Statistics
Purdue University

November 2008

A Shape Restricted Nonparametric Method with Financial Application

X. Jessie Jeng and Per A. Mykland
Purdue University and University of Chicago

November 20, 2008

Abstract

Many high-frequency financial data can be described by unimodal stationary processes. In this paper, we propose to apply the Grenander maximum likelihood estimation for estimating unimodel densities of stationary processes. We derive analogous asymptotic results as those for independent and identically distributed (i.i.d.) samples. In particular, the efficient convergence rate of $n^{1/3}$ is shown to be achieved under suitable mixing condition. We demonstrate this procedure for financial applications using the S&P 500 Index.

Keywords: Unimodal, Stationary process, mixing, Triangular array, MLE.

A Shape Restricted Nonparametric Method with Financial Application

X. Jessie Jeng and Per A. Mykland

November 19, 2008

Abstract

Many high-frequency financial data can be described by unimodal stationary processes. In this paper, we propose to apply the Grenander maximum likelihood estimation for estimating unimodel densities of stationary processes. We derive analogous asymptotic results as those for independent and identically distributed (i.i.d.) samples. In particular, the efficient convergence rate of $n^{1/3}$ is shown to be achieved under suitable mixing condition. We demonstrate this procedure for financial applications using the S&P 500 Index.

Keywords: Unimodal, Stationary process, mixing, Triangular array, MLE.

1 Introduction

In empirical finance, estimating returns on assets such as stocks, market indexes and exchange rates is of broad interest. Classical Black-Scholes model assumes that the logarithmic return of a stock price can be represented by an Ito stochastic differential equation:

$$d \log S_t = \mu dt + \sigma dW_t,$$

where S_t is a stochastic process representing the stock price at time t, μ the expected rate of return, σ the volatility and W_t a standard Brownian motion. Motivated by this pioneering work, different continuous-time processes have been proposed to capture the dynamics of the process S_t. Parametric model helps to reveal potential relationships among various financial factors. It also has the flexibility to adopt restrictions based on economics theory into the formulation (Hull, 2000). However, parametric model is usually sensitive to its assumptions. Empirical research has brought forth a considerable amount of stylized facts indicating erratic behavior of financial assets (Franses and Dijl, 2000). In this paper, we are interested in nonparametric methods that are suitable for analyzing financial data. Without assuming a particular model, nonparametric methods capture the empirical feature of the data set more precisely and are free of the joint hypotheses on asset-price dynamics and risk premia that are typical in parametric models. A commonly used nonparametric method is the kernel estimation (Scott, 1992), which estimates density function using smoothing procedures.

We observed some shape related features that are common for high-frequency data. For example, the histograms of logarithmic returns of S&P 500 Index often show unimodal shapes that have peaks around 0 (See Figure 1). This motivates us to consider the well-known Grenander estimator, which estimates a non-increasing continuous density by the slope of the concave majorant of the empirical process. One advantage of Grenander estimator compared with other nonparametric methods is that it is free of smoothing parameters because of the built-in shape constraint. Grenander estimator has been shown to achieve efficient convergence rate when sampling from i.i.d. observations (Rao, 1969). Also it can be easily modified to estimate a unimodal continuous density. For financial data, however, the independent assumption is usually unnatural. Stochastic processes that allow dependency among observations are often used to model the data. In order to incorporate the dependency, we study the Grenander estimator in the case of sampling from a stationary process. We argue that the stationary assumption is reasonable, especially for short-term and high-frequency data. As a nonparametric method, Grenander estimator can be applied to many financial instruments. One example is demonstrated on the S&P 500 Index data in Section 4. We find that Genander estimator captures the empirical features much better than the Black-Scholes estimate.

The rest of the paper is organized as follows. In section 2, the Grenander estimator is introduced with an estimation efficiency result for i.i.d observations. Then we generalize this result to the case of sampling from a stationary process. In Section 3, we apply Grenander estimator to the S&P 500 index data. We conclude in Section 4.

2 Method and asymptotic results

Given observations $Y_i, 1 \leq i \leq n$, from a unimodal distribution. If the mode, m_0, is known, the Grenander estimator of density $f(y)$ is defined as

$$\hat{f}_n(y; m_0) = \hat{p}\hat{f}_n(y; y \leq m_0) + (1 - \hat{p})\hat{f}_n(y; y \geq m_0),$$

where $\hat{p}$ is the proportion of observations taking values in $(-\infty, m_0]$, $\hat{f}_n(y; y \geq m_0)$ is the slope of the concave majorant of the empirical distribution F_n at point y given $y \geq m_0$, i.e.

$$\hat{f}_n(y; y \geq m_0) = \sup_{v>y} \inf_{u<y} \frac{F_n(v) - F_n(u)}{v - u},$$

and $\hat{f}_n(y; y \leq m_0)$ is the slope of the convex minorant of the empirical distribution F_n at point y given $y \leq m_0$ (Grenander (1956), Grenander (1983), Groeneboom (1985)).

In our case, the true mode m_0 is unknown. Let $\hat{m}$ be a consistent estimator of m_0. Then the density estimator $\hat{f}_n(y;\hat{m})$ has the same asymptotic property as the estimator $\hat{f}_n(y;m_0)$. The consistent mode estimator can be calculated by MLE method as follows. Let

$$\hat{j} = \arg\max_j \sum_{i \neq j} log(\hat{f}_n(Y_i; Y_j), \tag{2.1}$$

then, $\hat{m} = Y_{\hat{j}}$ (Bickel and Fan, 1996).

The following asymptotic property of Grenander estimator is derived in Rao (1969).

Lemma 2.1 *Given $X_1, ..., X_n$ independently observed from a distribution with decreasing density f on $[0, \infty)$, which has a nonzero derivative $f'(x)$ at a point $x \in (0, \infty)$. If $\hat{f}_n$ is the Grenander estimator of f based on $X_1, ..., X_n$, then*

$$n^{1/3}|\frac{1}{2}f(x)f'(x)|^{-1/3}(\hat{f}_n(x) - f(x)) \xrightarrow{L} 2Z \tag{2.2}$$

where Z is distributed as the random location of the maximum of the process $(W(u) - u^2, u \in \Re)$, and W is a standard two-sided Brownian Motion on $\Re$ originating from zero (i.e., $W(0) = 0$).

2.1 Asymptotic results in the case of sampling from a stationary process

In empirical finance, it is common to assume that the logarithmic return of a asset or volatility follows a stationary process. Therefore, we are motivated to study Grenander estimate under the stationary assumption. Since the independence assumption in Lemma 2.1 does not hold anymore, we assume the following mixing condition to describe and control the dependency among observations (Hall and Heyde, 1980). Define the α-mixing coefficient of a triangular array as

$$\alpha(j) = \max_{n,i:1 \leq i \leq n-j} \sup_{A \in \chi_{i+j}^n, B \in \chi_1^i} (|P(A \cap B) - P(A)P(B)|)$$

where $\chi_1^i$and χ_{i+j}^n denote the σ-fields generated by $\{X_k, k = 1, ..., i\}$ and $\{X_k, k = i + j, ..., n\}$, respectively. We assume the mixing condition,

$$\sum_{j=1}^{\infty} \alpha(j) < \infty. \tag{2.3}$$

We will show that the asymptotic result in Lemma 2.1 still holds under this more general setting. The idea is to reduce the problem of calculating the asymptotic distribution of the slope of the concave majorant of $F_n(y)$ over the interval $[x - 2cn^{-1/3}, x + 2cn^{-1/3}]$ at $y = x$ to the corresponding problem of a Brownian Motion over $[-2c, 2c]$. Define

$$Y_n(\delta_n) = n^{1/2}[F_n(x + \delta_n) - F(x + \delta_n)] - n^{1/2}[F_n(x) - F(x)] \tag{2.4}$$

and

$$D = -f'(x)/2 > 0. \tag{2.5}$$

Let $\alpha_n(x) + \delta_n\beta_n(x)$ denote the tangent to the concave majorant of $n^{-1/2}Y_n(\delta_n) - D\delta_n^2[1 + o(1)]$ at $\delta_n = 0$. Then $\beta_n(x)$ is the slope of the concave majorant of $n^{-1/2}Y_n(\delta_n) - D\delta_n^2[1 + o(1)]$ at $\delta_n = 0$. Therefore, we have

$$\hat{f}_n(x;\hat{m}) = f(x) + \beta_n(x). \tag{2.6}$$

Next, define

$$W_n(\zeta) = n^{-1/2}Y_n(\delta_n)[r_n^2 D]^{-1}, \tag{2.7}$$

where

$$r_n = [fD^{-2}n^{-1}]^{1/3} \tag{2.8}$$

and $\delta_n = r_n\zeta$, then

$$\begin{aligned} & n^{-1/2}Y_n(\delta_n) - D\delta_n^2[1+o(1)] - \alpha_n(x) - \delta_n\beta_n(x) \\ = \ & r_n^2 D\left(W_n(\zeta) - (\zeta + \frac{\beta_n}{2r_nD})^2 - (\frac{\alpha_n}{r_n^2D} - \frac{\beta_n^2}{4r_n^2D^2}) - \zeta^2 o(1)\right). \end{aligned}$$

It is obvious that $\beta_n[r_nD]^{-1}$ is the slope of the concave majorant at $\zeta = 0$ of the process

$$X_n(\zeta) = W_n(\zeta) - \zeta^2[1+o(1)] \tag{2.9}$$

on $[-q,q]$, where $q = 2cD^{2/3}f^{-1/3}$.

The above reformulates the problem by some suitable normalization. It can be shown that within the neighborhood $\zeta \in [-q,q]$, the process $X_n(\zeta)$ converges in distribution to the process $X(\zeta) = W(\zeta) - \zeta^2$, where W is a standard two-sided Brownian Motion. The key step is to show the asymptotic normality of $W_n(\zeta)$. But first of all, the asymptotic variance of $W_n(\zeta)$ should be derived under the setting of stationary process.

Lemma 2.2 *Let $X_1, ..., X_n$ be a time reversible stationary process with common distribution function $F(x)$. The joint density functions of X_1 and X_k, $2 \le k < \infty$, are continuously differentiable. If the α-mixing condition is satisfied with $\sum_{j=1}^{\infty}\alpha(j) < \infty$, then, for any ζ in $[-q,q]$, $W_n(\zeta)$ has asymptotic variance $|\zeta|$.*

In order to prove asymptotic normality of $W_n(\zeta)$, we will use the results in Liebscher (2001). Suppose $T_n = \sum_{k=1}^{n} U_{nk}$, where $\{U_{nk}, k = 1, ..., n\}$ is a triangular array with $EU_{nk} = 0$. Let $\Gamma_n = \max_{1\le k\le n} EU_{nk}^2$. The following condition is imposed on U_{nk} in Liebscher (2001).

Condition $C'(\infty)$: $\Gamma_n = O(n^{-1})$, $\sum_{j=1}^{\infty}\alpha(j) < \infty$, and $\operatorname{ess\,sup}|U_{nk}| := C_{nk} < \infty$ for all $k = 1, ..., n$. There exists a sequence m_n of positive integers tending to ∞ such that

$$nm_n\gamma_n = o(1)$$

where

$$\gamma_n := \max_{j,k:1\le j,k\le n, j\ne k} E|U_{nj}U_{nk}|$$

and

$$\sum_{j=m_n+1}^{\infty}\alpha(j)\sum_{k=1}^{n} C_{nk}^2 = o(1).$$

The following lemma can be found in Liebscher (2001), Section 2.

Lemma 2.3 *Assume $\{U_{ni}\}$ is an α-mixing array, and Condition $C'(\infty)$ holds. Suppose that there is a sequence τ_n of positive real numbers such that $\tau_n = o(\sqrt{n})$, $\tau_n \to \infty$,*

$$\tau_n \le (\max_{1\le k\le n} C_{nk})^{-1}$$

and

$$\frac{n}{\tau_n}\alpha(\lfloor\epsilon\tau_n\rfloor) \to 0 \quad \textit{for all } \epsilon > 0.$$

Moreover, assume that

$$\lim_{n\to\infty}\sum_{i=1}^{n} EU_{ni}^2 = \sigma^2 > 0.$$

Then

$$T_n \longrightarrow N(0,\sigma^2) \quad \textit{in distribution.}$$

With the help of the above two lemmas, we have the following theorem.

Theorem 1 *For any ζ in $[-q,q]$, $W_n(\zeta)$ is asymptotically normal with mean 0 and variance $|\zeta|$.*

Next, we show that $W_n(\zeta)$ converges in distribution to a standard Brownian motion $W(\zeta)$. Two sufficient conditions for this purpose has been derived in Chenstov (1956) and Sethuraman (1965). They essentially constrain the variance-covariance matrix and the tightness of the process. We verify these conditions under our situation in the following Lemmas.

Lemma 2.4 *For any collection ζ_i, $1 \le i \le p$, $|\zeta_i| \le q$, the joint distribution of $[W_n(\zeta_1), \ldots, W_n(\zeta_p)]$ converges to the multivariate normal distribution with mean 0 and the variance-covariance matrix*

$$[\delta(\zeta_i, \zeta_j) min(|\zeta_i|, |\zeta_j|)],$$

where

$$\delta(a, b) = 1\{sign(a) = sign(b)\}.$$

Lemma 2.5 *For any $\zeta_1 < \zeta_2 < \zeta_3$ in $[-q, q]$,*

$$E\big[|W_n(\zeta_1) - W_n(\zeta_2)||W_n(\zeta_2) - W_n(\zeta_3)|\big] \le C|\zeta_3 - \zeta_1|,$$

where C is a constant independent of n, ζ_1, ζ_2 and ζ_3.

Now, we are ready to apply the following Lemma, which is due to Chenstov (1956). A proof can be found in Sethuraman (1965).

Lemma 2.6 *Let X_n be a sequence of stochastic processes in $D[a, b]$ and X be another process in $D[a, b]$ such that*

1) for any $(t_1, ..., t_k)$ in $[a, b]$, the joint distribution of $[X_n(t_1), ..., X_n(t_k)]$ converges weakly to the joint distribution of $[X(t_1), ..., X(t_k)]$, and

2) for any $t_1 < t_2 < t_3$ in $[a, b]$,

$$E\big[|W_n(\zeta_1) - W_n(\zeta_2)|^{\gamma_1}|W_n(\zeta_2) - W_n(\zeta_3)|^{\gamma_2}\big] \le C|\zeta_3 - \zeta_1|^{1+\gamma_3}$$

for some numbers γ_i, $i = 1, 2, 3$ and a constant $C > 0$ independent of n, t_1, t_2 and t_3.

Let ν_n and ν be measures of X_n and X on $D[a, b]$. Then, ν_n converges weakly to ν.

We have shown that within the neighborhood $\zeta \in [-q, q]$, the process $X_n(\zeta)$ converges in distribution to the process $X(\zeta) = W(\zeta) - \zeta^2$. Since the slope of concave majorant of process $X(\zeta)$ has the same distribution as that of $2Z$, where Z is the location of the maximum of process $X(\zeta)$, then, by combining (2.9) with (2.6), (2.8) and (2.5), the following theorem is straightforward.

Theorem 2 *Given $X_1, ..., X_n$ observed from a time reversible stationary process with a decreasing density f on $[0, \infty)$, which has a nonzero derivative $f'(x)$ at a point $x \in (0, \infty)$. Assume mixing condition (2.3), then the Grenander estimate $\hat{f}_n$ satisfies*

$$n^{1/3}\left|\frac{1}{2}f(x)f'(x)\right|^{-1/3}\left(\hat{f}_n(x) - f(x)\right) \xrightarrow{L} 2Z \tag{2.10}$$

where Z is distributed as the random location of the maximum of the process $(W(u) - u^2, u \in \Re)$, and W is a standard two-sided Brownian Motion on $\Re$ originating from zero (i.e., $W(0) = 0$).

2.2 Penalized maximum likelihood estimate of $f(\hat{m}+)$

When sampling from i.i.d observations, it is known that, even though $\hat{f}_n(x)$ is a consistent estimator of $f(x)$ for $x > m$, it does not follow that $\hat{f}_n(m+)$ is a consistent estimator of $f(m+)$ (Woodroofe and Sun, 1993). If the mode m is unknown, a consistent estimator $\hat{m}$ can be found by (2.1), but $\hat{f}_n$ often show spikiness near the location $\hat{m}+$.

One way to reduce the the size of $\hat{f}_n(\hat{m}+)$ is to penalize the nonparametric likelihood function for large value of $f(\hat{m}+)$. In Woodroofe and Sun (1993), a penalized MLE is proposed, which is consistent at $\hat{m}+$. The approach is as the follows. First, a smoothing parameter $\alpha > 0$ is introduced, and the penalized likelihood function is of the form

$$l_\alpha(f) = \sum_{i=1}^{n} logf(x_i) - n\alpha f(\hat{m}+).$$

For a fixed α and $\gamma \in [0,1]$, let

$$f_1(\gamma) = max_{1\leq s\leq n}\frac{s/n}{\alpha + \gamma x_s},$$

and let $\hat{\gamma}$ denote the solution of the equation

$$\gamma = 1 - \alpha f_1(\gamma),$$

so that

$$\hat{\gamma} = min_{1\leq s\leq n}\left(\frac{1}{2}(1-\frac{\alpha}{x_s}) + \sqrt{(\frac{\alpha}{2x_s})^2 + \frac{\alpha}{2x_s}(1-\frac{2s}{n}) + \frac{1}{4}}\right).$$

Then, the consistent estimator $\tilde{f}_n(\hat{m}+)$ at $\hat{m}+$ is

$$\tilde{f}_n(\hat{m}+) = \frac{1-\hat{\gamma}}{\alpha}.$$

Once $\hat{\gamma}$ is found, computing the penalized MLE $\tilde{f}_n$ is no more difficult than computing the unpenalized MLE $\hat{f}_n$. In fact, the penalized MLE is equal to an unpenalized MLE with a transformed data set $\alpha + \hat{\gamma}x_k$, $k = 1, ..., n$.

Note that the smoothing parameter α should satisfy the following conditions:

$$0 < \alpha = \alpha(n) \longrightarrow 0$$

and

$$n\alpha(n) \longrightarrow \infty.$$

For example, $\alpha(n)$ can be $\frac{logn}{2n}$, or cn^{-pq}, where $1 < p < \infty$ and $q = \frac{1}{2p-1}$.

3 Density estimation for the S&P 500 Index

In order to examine the empirical relevance of the unimodal density estimator, we study the daily log returns of the S&P 500 Index obtained from CBOE. Figure 1 is the histogram of two data sets. The left one is the daily log returns of S&P 500 Index from January 1st, 2002 to December 31,2002, and the right one is from January 1st, 1986 to September 30, 2003. Both the one-year and longer-term data show peaks near 0, and frequencies decrease with increasing absolute log returns. This fact motivates us to assume that the daily log returns of S&P 500 Index have unimodal density function with mode near 0.

S&P 500 Index is among the most actively traded assets in the world. Since jumps are less likely to occur in indices than in individual equities due to diversification, the dynamics of S&P 500 Index likely follows a stationary process. The main features of the data set are shown in Table 1 by calculating the first four moments. We can see that both the skewness and the kurtosis in 2002 is quite unusual compared with those from the longer-term data set. The positive skewness means that large positive returns tend to occur more often than large negative returns, and the small kurtosis implies thin tails.

Both one-year and longer-term data show evidence of unimodal distribution of the log returns of S&P 500 Index. We apply the Grenander estimator to estimate the density. We compare the behavior of our nonparametric approach with the Black-Scholes model. In Figure 2, the estimated densities of Grenander estimate and Normal distribution are shown for the 2002 data. The normal density is implied by Black-Scholes model, and its parameters are estimated by matching the first two moments.

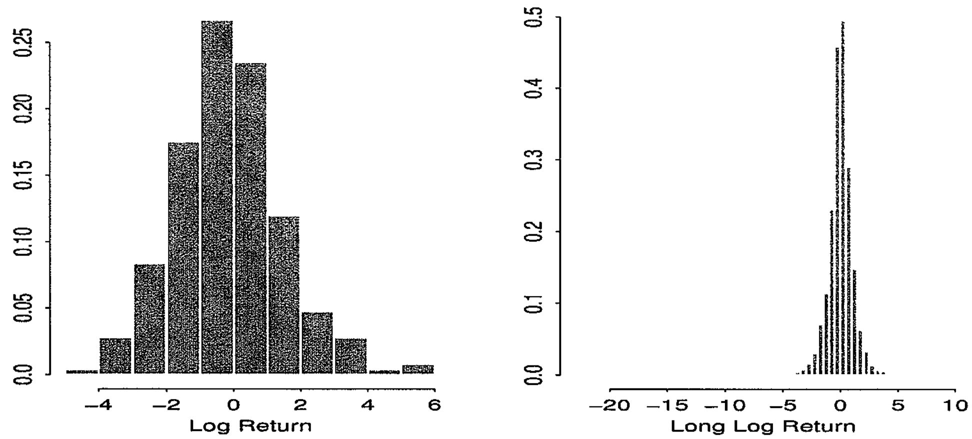

Figure 1: Histogram of daily log returns of S&P 500 Index.

	Mean	Var	Skew	Kurt
2002	-0.1083	2.6831	0.4268	3.6252
86-03	0.0348	1.2749	-2.0689	45.0045

Table 1: Summary statistics of daily log returns of S&P 500 Index.

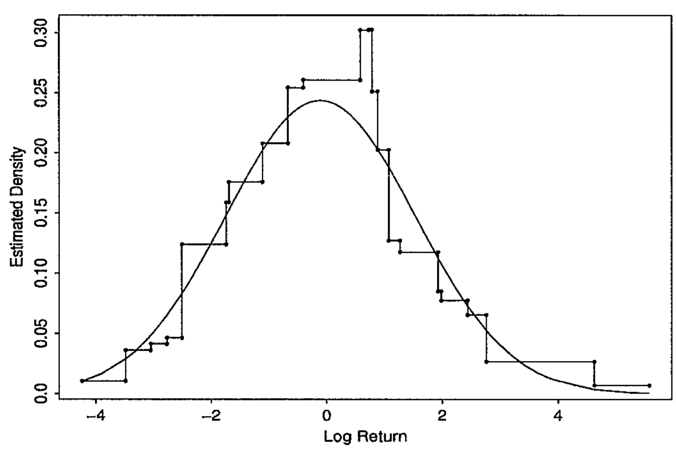

Figure 2: Grenander estimate versus Black-Scholes estimate.

The summary statistics of Grenander estimate and Black-Scholes estimate (BS) are compared in Table 2. We can see that the Grenander estimate matches the skewness and kurtosis of the data more successfully.

Based on Theorem 2, it is possible to construct a confidence interval for the density estimator of log returns. The limiting distribution Z is simulated with the 5% and 95% quantiles being -0.14 and 0.14. The 90% confidence interval plots are shown in Figure 3.

	Mean	Var	Skew	Kurt
Data	-0.1083	2.6831	0.4268	3.6252
Grenander estimate	-0.1973	2.6825	0.4640	5.3666
BS	-0.1083	2.6831	0	3

Table 2: Summary statistics of Grenander estimate and Normal estimate.

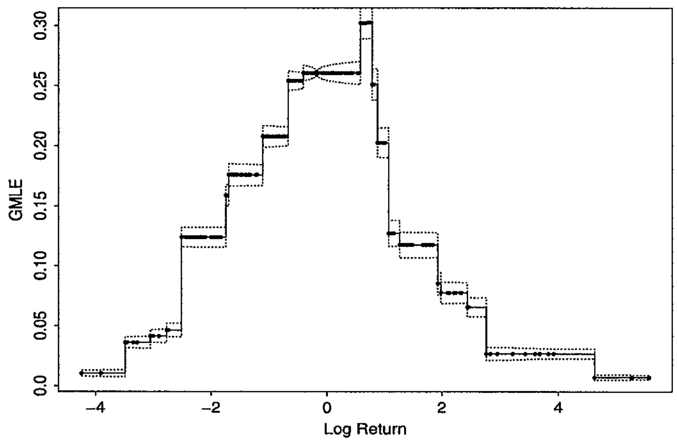

Figure 3: Confidence interval plot for Grenander estimate.

4 Conclusion

In this paper, we study the asymptotic properties of Grenander maximum likelihood estimator in the setting of sampling from a stationary process. We find that both the convergence rate and the limiting distribution remain the same as those in the i.i.d. setting under suitable mixing condition. We apply the Grenander estimator on the S&P 500 Index to incorporate the unimodal shape of the frequency. Compared with the parametric method using Geometric Brownian Motion, our approach captures more empirical features of the data. Because of its robustness, the proposed nonparametric method may be useful for modelling other financial quantities, for example interest rate and volatility.

5 Appendix

5.1 Proof of Lemma 2.2

By (2.7), (2.8) and (2.4), we have

$$W_n(\zeta) = f^{-2/3}D^{1/3}n^{1/6}\left(\sqrt{n}(F_n(x+\delta_n)-F(x+\delta_n))-\sqrt{n}(F_n(x)-F(x))\right).$$

Let

$$M_n(x)=\sqrt{n}(F_n(x)-F(x)),$$

then

$$W_n(\zeta)=f^{-2/3}D^{1/3}n^{1/6}\left(M_n(x+\delta_n)-M_n(x)\right).$$

Next, we calculate $Var(M_n(x))$, $Var(M_n(x+\delta))$ and $Cov(M_n(x), M_n(x+\delta_n))$, respectively. First,

$$\begin{aligned} Var(M_n(x)) &= Var\left(\frac{\sum_{i=1}^{n} I(X_i \le x)}{\sqrt{n}}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} Var(I(X_i \le x)) + 2\frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(I(X_i \le x), I(X_j \le x)) \\ &= Var(I(X_1 \le x)) + 2\frac{1}{n}\sum_{i=0}^{n-2}\sum_{k=2}^{n-i} Cov(I(X_1 \le x), I(X_k \le x)) \\ &= Var(I(X_1 \le x)) + 2\sum_{k=2}^{n}\left(Cov(I(X_1 \le x), I(X_k \le x))\frac{n-k+1}{n}\right), \end{aligned}$$

where the third equality is by the stationary property. Similarly,

$$\begin{aligned} Var(M_n(x+\delta_n)) &= Var(I(X_1 \le x+\delta_n)) \\ &\quad +2\sum_{k=2}^{n}\left(Cov(I(X_1 \le x+\delta_n), I(X_k \le x+\delta_n))\frac{n-k+1}{n}\right) \end{aligned}$$

and,

$$\begin{aligned} Cov(M_n(x), M_n(x+\delta_n)) &= Cov(I(X_1 \le x), I(X_1 \le x+\delta_n)) \\ &\quad +2\sum_{k=2}^{n}\left(Cov(I(X_1 \le x), I(X_k \le x+\delta_n))\frac{n-k+1}{n}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} & Var(M_n(x+\delta_n) - M_n(x)) \\ = & \left(Var(I(X_1 \le x+\delta_n)) + 2\sum_{k=2}^{n} Cov(I(X_1 \le x+\delta_n), I(X_k \le x+\delta_n))\frac{n-k+1}{n}\right) \\ + & \left(Var(I(X_1 \le x)) + 2\sum_{k=2}^{n} Cov(I(X_1 \le x), I(X_k \le x))\frac{n-k+1}{n}\right) \\ - & 2\left(Cov(I(X_1 \le x), I(X_1 \le x+\delta_n)) + 2\sum_{k=2}^{n} Cov(I(X_1 \le x), I(X_k \le x+\delta_n))\frac{n-k+1}{n}\right). \end{aligned}$$

Without loss of generality, assume $\delta_n \ge 0$. Since

$$\begin{aligned} & Var(I(X_1 \le x+\delta_n)) + Var(I(X_1 \le x)) - 2cov(I(X_1 \le x+\delta_n), I(X_1 \le x)) \\ = & EI(X_1 \le x+\delta_n) - (EI(X_1 \le x+\delta_n))^2 + EI(X_1 \le x) - (EI(X_1 \le x))^2 \\ & -2\left(EI(X_1 \le x) - EI(X_1 \le x+\delta_n)EI(X_1 \le x)\right) \\ = & F(x+\delta_n) - F(x) - (F(x+\delta_n) - F(x))^2 \\ = & f(x)\delta_n + o(\delta_n) \end{aligned}$$

and,

$$
\begin{aligned}
& 2\sum_{k=2}^{n}\Big\{\big(Cov(I(X_1 \le x+\delta_n), I(X_k \le x+\delta_n)) + Cov(I(X_1 \le x), I(X_k \le x)) \\
& -2Cov(I(X_1 \le x), I(X_k \le x+\delta_n))\big)\frac{n-k+1}{n}\Big\} \\
= \; & 2\sum_{k=2}^{n}\Big\{\big(EI(X_1 \le x+\delta_n)I(X_k \le x+\delta_n) + EI(X_1 \le x)I(X_k \le x) \\
& -2EI(X_1 \le x)I(X_k \le x+\delta_n) - (F(x+\delta_n)-F(x))^2\big)\frac{n-k+1}{n}\Big\} \\
= \; & 2\sum_{k=2}^{n}\Big\{\big(P(X_1 \le x+\delta_n, X_k \le x+\delta_n) + P(X_1 \le x, X_k \le x) \\
& -2P(X_1 \le x, X_k \le x+\delta_n) - (F(x+\delta_n)-F(x))^2\big)\frac{n-k+1}{n}\Big\} \\
= \; & 2\sum_{k=2}^{n}\Big\{\big(P(x < X_1 \le x+\delta_n, X_k \le x+\delta_n) - P(X_1 \le x, x < X_k \le x+\delta_n) \\
& -(F(x+\delta_n)-F(x))^2\big)\frac{n-k+1}{n}\Big\},
\end{aligned}
$$

Then

$$
\begin{aligned}
\lim_{n\to\infty} Var(W_n(\zeta)) \;=\; & |\zeta| + \lim_{n\to\infty} f^{-4/3}D^{2/3}n^{1/3}2\sum_{k=2}^{n}\Big\{\big(P(x < X_1 \le x+\delta_n, X_k \le x+\delta_n) \\
& -P(X_1 \le x, x < X_k \le x+\delta_n) - (F(x+\delta_n)-F(x))^2\big)\frac{n-k+1}{n}\Big\}.
\end{aligned}
$$

It is left to show that the second term of the right hand side is 0. Without loss of generality, assume $\delta_n = n^{-1/3}$, then we have

$$
\begin{aligned}
& n^{1/3}\sum_{k=2}^{n}\Big\{\big(P(x < X_1 \le x+n^{-1/3}, X_k \le x+n^{-1/3}) \\
& -P(X_1 \le x, x < X_k \le x+n^{-1/3}) - (F(x+n^{-1/3})-F(x))^2\big)\frac{n-k+1}{n}\Big\} \\
= \; & n^{1/3}\sum_{k=2}^{n}\Big\{\big(P(x < X_1 \le x+n^{-1/3}, x < X_k \le x+n^{-1/3}) \\
& -(F(x+n^{-1/3})-F(x))^2\big)\frac{n-k+1}{n}\Big\} \\
= \; & n^{1/3}\sum_{k=2}^{n}\Big\{\big(P(x < X_k \le x+n^{-1/3} | x < X_1 \le x+n^{-1/3})(F(x+n^{-1/3})-F(x)) \\
& -(F(x+n^{-1/3})-F(x))^2\big)\frac{n-k+1}{n}\Big\} \\
= \; & n^{1/3}(F(x+n^{-1/3})-F(x))\sum_{k=2}^{n}\Big\{\big(P(x < X_k \le x+n^{-1/3} | x < X_1 \le x+n^{-1/3}) \\
& -P(x < X_k \le x+n^{-1/3})\big)\frac{n-k+1}{n}\Big\},
\end{aligned}
$$

where the first equality is implied by the time reversibility property. Since

$$
\begin{aligned}
& \sum_{k=2}^{\infty}\Big|P(x < X_k \le x+n^{-1/3} | x < X_1 \le x+n^{-1/3}) - P(x < X_k \le x+n^{-1/3})\Big| \\
\le \; & \sum_{k=2}^{\infty}\alpha(k-1) < \infty
\end{aligned}
$$

by mixing condition (2.3), then by Dominated Convergence Theorem, we get

$$\begin{aligned}
&\lim_{n\to\infty}\left\{n^{1/3}(F(x+n^{-1/3})-F(x))\sum_{k=2}^{n}(P(x<X_k\le x+n^{-1/3}|x<X_1\le x+n^{-1/3})\right.\\
&\left.-P(x<X_k\le x+n^{-1/3}))\frac{n-k+1}{n}\right\}\\
=&\ f(x)\sum_{k=2}^{\infty}\lim_{n\to\infty}\left(P(x<X_k\le x+n^{-1/3}|x<X_1\le x+n^{-1/3})-P(x<X_k\le x+n^{-1/3})\right)\\
=&\ f(x)\sum_{k=2}^{\infty}\lim_{n\to\infty}\frac{P(x<X_k\le x+n^{-1/3},x<X_1\le x+n^{-1/3})}{P(x<X_1\le x+n^{-1/3})}\\
=&\ f(x)\sum_{k=2}^{\infty}\lim_{n\to\infty}\frac{f_{1,k}(x,x)n^{-2/3}}{f(x)n^{-1/3}}\\
=&\ 0,
\end{aligned}$$

where $f_{1,k}(x,x)$ is the joint probability density of X_1 and X_k. Therefore, result follows.

5.2 Proof of Theorem 1

We will verify Condition $C'(\infty)$ and other conditions in Lemma 2.3.

Let $m_n=\lfloor n^{1/6}\rfloor$, then we have

(1) $\Gamma_n=\max_{1\le i\le n}EU_{nk}^2=EU_{n1}^2=\frac{f^{-4/3}D^{2/3}}{n}EZ_{nk}^2=O(n^{-1})$

(2) $\sum_{k=1}^{\infty}\alpha(k)<\infty$ is given.

(3) $C_{nk}:=esssup|U_{nk}|=esssup|\frac{f^{-2/3}D^{1/3}}{\sqrt{n}}Z_{nk}|$, where Z_{nk} is a non-degenerate random variable wtih $P(|Z_{nk}|>n^{\alpha})\to 0$ as $n\to\infty$ for any $\alpha>0$. Pick $\alpha<\frac{1}{2}$, then $C_{nk}<\infty$ for sufficient large n.

(4) $nm_n\gamma_n=nn^{1/6}\max_{j\neq k}E|U_{nj}U_{nk}|=n^{1/6}f^{-4/3}D^{2/3}\max_{j\neq k}E|Z_{nj}Z_{nk}|$, and

$$\begin{aligned}
&E|Z_{nj}Z_{nk}|\\
=&\ n^{1/3}E\left|(I(x<X_j\le x+\delta_n)-(F(x+\delta_n)-F(x)))-(I(x<X_k\le x+\delta_n)-(F(x+\delta_n)-F(x))\right|\\
\approx&\ n^{1/3}|P(x<X_j\le x+\delta_n,x<X_k\le x+\delta_n)-(f(x))^2(\delta_n)^2|\\
=&\ O(n^{-1/3}),
\end{aligned}$$

where the last equation is from 2-dimensional Taylor expansion. Then we have

$$nm_n\gamma_n=O(n^{-1/6})=o(1).$$

(5) Since

$$\left(\sum_{m_n+1}^{\infty}\alpha(j)\right)\left(\sum_{k=1}^{n}C_{nk}^2\right)=\left(\sum_{m_n+1}^{\infty}\alpha(j)\right)\left(\sum_{k=1}^{n}(\text{ess sup}\left|\frac{f^{-2/3}D^{1/3}Z_{nk}}{\sqrt{n}}\right|)^2\right),$$

where Z_{nk} is a non-degenerate random variable and $\sum_{m_n+1}^{\infty}\alpha(j)\to 0$. Then we have

$$\left(\sum_{m_n+1}^{\infty}\alpha(j)\right)\left(\sum_{k=1}^{n}C_{nk}^2\right)=o(1).$$

Condition $C'(\infty)$ is verified by Fact (1)-(5).

Next, we check other conditions in Lemma 2.3. Let $\tau_n=n^{5/12}$, then

(6) $\tau_n\le\frac{1}{\max_{1\le k\le n}C_{nk}}=O(\frac{\sqrt{n}}{\text{ess sup}|Z_{nk}|})$.

(7) $\frac{n}{\tau_n}\alpha(\lfloor\epsilon\tau_n\rfloor)=n^{7/12}\alpha(\lfloor\epsilon n^{5/12}\rfloor)$. For simplicity, assume $\alpha(j)=O(\frac{1}{j^2})$, then it is true that

$$\frac{n}{\tau_n}\alpha(\lfloor\epsilon\tau_n\rfloor)\to 0\ \text{ for all }\epsilon>0.$$

Result follows by applying Fact (6)-(7), Condition $C'(\infty)$, and Lemma 2.3.

5.3 Proof of Lemma 2.4

Recall

$$W_n(\zeta) = f^{-2/3} D^{1/3} n^{1/6} Y_n(\delta_n).$$

It is clear that the limiting distribution has mean 0. The variance-covariance matrix can be calculated by similar procedure as in proof of Lemma 2.2.

$$\begin{aligned}
Cov\left(Y_n(\delta_n^i), Y_n(\delta_n^j)\right) &= Cov\left(M_n(x+\delta_n^i) - M_n(x), M_n(x+\delta_n^j) - M_n(x)\right) \\
&= Cov\left(M_n(x+\delta_n^i), M_n(x+\delta_n^j)\right) - Cov\left(M_n(x+\delta_n^i), M_n(x)\right) \\
&\quad - Cov\left(M_n(x), M_n(x+\delta_n^j)\right) + Var\left(M_n(x)\right) \\
&\approx f(x)\delta_i.
\end{aligned}$$

Therefore, by similar procedure as in Theorem 1 we have asymptotic normality on

$$\begin{aligned}
\lambda^T \begin{pmatrix} W_n(\zeta_1) \\ \vdots \\ W_n(\zeta_p) \end{pmatrix} &= \sum_{i=1}^{p} \lambda_i W_n(\zeta_i) \\
&= \sum_{i=1}^{p} \lambda_i \frac{\sum_{k=1}^{n} Z_{nk}(\zeta_i)}{\sqrt{n}} \\
&= \frac{\sum_{k=1}^{n} \left(\sum_{i=1}^{p} \lambda_i Z_{nk}(\zeta_i)\right)}{\sqrt{n}} \\
&\xrightarrow{L} \Lambda
\end{aligned}$$

where Λ is a normal random variable and λ is any conformable nonzero vector.

Finally, by Cramer-Wold device, we have

$$\begin{pmatrix} W_n(\zeta_1) \\ \vdots \\ W_n(\zeta_p) \end{pmatrix} \xrightarrow{L} \begin{pmatrix} W(\zeta_1) \\ \vdots \\ W(\zeta_p) \end{pmatrix},$$

where the random vector on the right hand side follows a multivariate normal distribution.

5.4 Proof of Lemma 2.5

By direct calculation

$$\begin{aligned}
& E\left(|W_n(\zeta_1) - W_n(\zeta_2)||W_n(\zeta_2) - W_n(\zeta_3)|\right) \\
\leq\ & \sqrt{var(W_n(\zeta_1) - W_n(\zeta_2))}\sqrt{var(W_n(\zeta_2) - W_n(\zeta_3))} \\
\leq\ & C\sqrt{\zeta_1 + \zeta_2 - cov(W_n(\zeta_1), W_n(\zeta_2))}\sqrt{\zeta_2 + \zeta_3 - cov(W_n(\zeta_2), W_n(\zeta_3))} \\
\leq\ & C\sqrt{\zeta_1 - \zeta_2}\sqrt{\zeta_2 - \zeta_3} \\
\leq\ & C|\zeta_3 - \zeta_1|
\end{aligned}$$

for some constant C independent of n.

References

Bickel, P. J. and J. Fan (1996). Some problems of the estimation of unimodal densities. *Statist. Sinica 6*, 23–45.

Chenstov, N. N. (1956). Weak convergence of stochastic processes whose trajectories have no discontinuities of the second kind and the 'heuristic approach to the kolmogorov-smirnov tests'. *Theory of Prob. and Its Appl. 1*, 140–144.

Franses, P. and D. Dijl (2000). *Non-linear Time Series Models in Empirical Finance.* Cambridge University Press, UK.

Grenander, P. (1956). On the theory of mortality measurement, part ii. *Skand. Akt. 39*, 125–153.

Grenander, P. (1983). *Nonparametric Function Estimation.* John Wiley.

Groeneboom, P. (1985). Estimating a monotone density. *In Proc. Conf. in Honor of Jerzy Neyman and Jack Kiefer (Edited by L. Lecam and R. Olshen)*, 539–555.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application.* Academic Press, NY.

Hull, J. (2000). *Options, Futures and Other Derivatives. 4th Edition.* Prentile-Hall, NJ.

Liebscher, E. (2001). Central limit theorems for α-mixing triangular arrays with applications to nonparametric statistics. *Mathematical Methods of Statistics 10*, 194–214.

Rao, P. (1969). Estimation of unimodal densities. *Sankhyā Ser. A 31*, 23–36.

Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice, and Visualization.* Wiley, NY.

Sethuraman, J. (1965). Limit theorems for stochastic processes. Technical Report 10, Department of Statistics, Stanford University.

Woodroofe, M. and J. Sun (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statist. Sinica 3*, 501–515.