KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.947

"Convergence rates of sums of α-mixing triangular arrays: with an application to non-parametric drift function estimation of continuous-time processes"

Shin Kanaya

August 2016

KYOTO UNIVERSITY

KYOTO, JAPAN

Convergence rates of sums of α-mixing triangular arrays: with an application to non-parametric drift function estimation of continuous-time processes*

Shin Kanaya[†]
Aarhus University, CREATES, and IER

Abstract

The convergence rates of the sums of α-mixing (or strongly mixing) triangular arrays of heterogeneous random variables are derived. We pay particular attention to the case where central limit theorems may fail to hold, due to relatively strong time-series dependence and/or the non-existence of higher-order moments. Several previous studies have presented various versions of laws of large numbers for sequences/triangular arrays, but their convergence rates were not fully investigated. This study is the first to investigate the convergence rates of the sums of α-mixing triangular arrays whose mixing coefficients are permitted to decay arbitrarily slowly. We consider two kinds of asymptotic assumptions: one is that the time distance between adjacent observations is fixed for any sample size n; and the other, called the *infill* assumption, is that it shrinks to zero as n tends to infinity. Our convergence theorems indicate that an explicit trade-off exists between the rate of convergence and the degree of dependence. While the results under the infill assumption can be seen as a direct extension of those under the fixed-distance assumption, they are new and particularly useful for deriving sharper convergence rates of discretization biases in estimating continuous-time processes from discretely sampled observations. We also discuss some examples to which our results and techniques are useful and applicable: a moving-average process with long lasting past shocks, a continuous-time diffusion process with weak mean reversion, and a near-unit-root process.

This version: August 2016.
Keywords: Law of large numbers; rate of convergence; α-mixing triangular array; infill asymptotics; kernel estimation.
JEL codes: C14; C22; C58.

*The author thanks Bruce E. Hansen, Bent Nielsen, and Neil Shephard for helpful discussions and suggestions. I would also like to thank the Editor, Peter C. B. Phillips, and three anonymous referees for their constructive and valuable comments, which have greatly improved the original version of this paper. In particular, I would like to express my sincere gratitude to Professor Phillips for generous support and outstanding editorial input into this paper, which were considerable and far in excess of what I could expect. I gratefully acknowledge support from CREATES, Center for Research in Econometric Analysis of Time Series, funded by the Danish National Research Foundation (DNRF78), and from the Danish Council for Independent Research, Social Sciences (grant no. DFF - 4182-00279). Part of this research was conducted while I was visiting the Institute of Economic Research, Kyoto University (under the Joint Research Program of the KIER), the support and hospitality of which are gratefully acknowledged.

[†]Address: Department of Economics, Aarhus University, Fuglesangs Alle 4, Aarhus V, 8210, Denmark; the Institute of Economic Research (IER), Hitotsubashi University. E-mail: skanaya@econ.au.dk. Phone: +45 87165984.

1 Introduction

The purpose of this paper is to investigate the convergence rates of the sums of α-mixing (or strongly mixing) triangular arrays of heterogeneous random variables. We specifically consider the case where the degree of dependence is relatively strong (allowing for any arbitrarily slow decay rate of mixing coefficients), its higher-order moments do not exist and, thus, the central limit theorem (CLT) may fail to hold. For such cases, several versions of the laws of large numbers (LLN) for sequences/triangular arrays are known, but, in contrast, the convergence rates of the sums have not been fully investigated in the literature.

The LLN states that the sum of a zero-mean array divided by its sample size n converges to zero. The CLT, on the other hand, tells us that its convergence rate is the order of $n^{-1/2}$. While the weak LLN imposes no decay rate on the mixing coefficients (see Example 4 of Andrews, 1988, and Remark 2 below), requiring only that they converge to zero, the CLT requires that the decay rate is faster than a particular rate and the higher-order moments exist.[1] There is a gap between these two results: the convergence rates of the sums (or the sample means) are unknown when the CLT may fail to hold due to the slow decay of mixing coefficients and/or non-existence of the higher-order moments. In related studies, Davidson and de Jong (1997) and de Jong (1998) analyzed the convergence rates of the sums of mixingale arrays/sequences, considering such intermediate cases.[2] Analyses of this sort for α-mixing arrays have not been fully conducted in the literature. Since α-mixing processes are also mixingale, given a certain moment condition, the results of Davidson and de Jong are also applicable to the α-mixing case. However, convergence rates obtained from such mixingale results are generally not sharp for α-mixing arrays. This is because the mixingale size of a mixing array/sequence is generally shown to be lower than its mixing size (see subsequent discussions: "Comparison with previous mixingale LLN results"). Our results, which aim at α-mixing arrays, may lead to sharper rates. This complements the work of Davidson and de Jong. This study is the first to investigate the convergence rates of the sums of α-mixing triangular arrays whose mixing coefficients are allowed to decay arbitrarily slowly. We present an explicit trade-off between the rate of convergence and the degree of dependence, suggesting that stronger dependence implies slower convergence of the sums.

Our results also complement previous studies concerning convergence rates in the LLNs for α-mixing (or some other dependent) sequences/arrays. These studies include Shao (1993), Liebscher (1996), Louhichi (2000), and Louhichi and Soulier (2000). We take an approach similar to that in Liebscher (1996), relying on the Bernstein-type inequality derived in Rio (1995). All of these, including Liebscher's, consider *strong* LLNs and focus on the case where the degree of dependence is (relatively) weak, typically assuming that the decay rate of mixing coefficients is fast enough. In contrast, we work under a *weak* notion of the convergence in probability. This seems the price of allowing for (relatively) strong dependence. Additionally, our results (as in the previous studies) exhibit a trade-off between the rate of convergence and the order of the moment, that is, if the higher order moment exists, the convergence of the sum occurs more quickly.

The results are useful when we consider time series with high persistence, such as financial asset

[1] For the conditions under which the CLT for a mixing process holds (or fails to hold), see Ibragimov (1962), Chapter 18 of Ibragimov and Linnik (1971), Bradley (1985, 1988), Merlevède, Peligrad and Utev (2006), and the references therein.

[2] For LLNs in the mixingale case, see also McLeish (1975), Andrews (1988), Hansen (1991, 1992), Davidson (1993), and de Jong (1995, 1996).

returns. As argued in Chen, Hansen, and Carrasco (2010), some Markov processes, which look like long memory type processes, exhibit this property. These processes, whose mixing coefficients decay very slowly, are in the scope of our results. Heavy-tailed distributions may also be a key feature in a financial time series. Our results may be applied to processes with such a feature, like processes with infinite variance. We present some examples to which our results are relevant and applicable for both discrete and continuous time cases. Note that we do not assume the stationarity of the processes, allowing for some form of heterogeneity and/or non-stationarity. However, we can only deal with minor non-stationarity in that strongly non-stationary processes, such as unit-root and null-recurrent processes, are excluded. Our theorems are stated for triangular arrays and can be used to investigate the convergence rates of estimators/statistics in situations where distributions of random variables may vary with the sample size n. Such situations naturally arise, e.g., when estimating time varying models (e.g., Dahlhaus and Rao, 2006), using kernel-based methods (see, e.g., Chapters 5 and 6 of Fan and Yao, 2003), and investigating limit properties of near-unit-root processes (e.g., Stock, 1991; Elliott, Rothenberg, and Stock, 1996; Phillips and Magdalinos, 2007). We discuss the last two cases in detail (in Section 3 and Example 3, respectively).

We consider two kinds of asymptotic assumptions: one is that the time distance between the adjacent observations is fixed for any sample size n and the other is that it shrinks to zero as n tends to infinity. The latter is called *infill* asymptotics and is often necessary to estimate continuous-time models from discretely sampled data (e.g., Florens-Zmirou, 1989; Bandi and Phillips, 2003; Kristensen, 2010). An original suggestion and exploration of infill asymptotics were given in Phillips' (1987) seminal paper on time series regression with a unit root (see his Section 6 of "Continuous Record Asymptotics").[3] We do not consider a unit root case but share the motivation for infill asymptotics with Phillips (1987). These may provide a reasonable approximation to analyze high-frequency data. While several previous studies have investigated infill asymptotics, their derivations of limit theorems have often relied on a sort of (semi) martingale assumption, as in Phillips (1987) and in recent papers on volatility estimation (such as Barndorff-Nielsen and Shephard, 2002). In contrast, we intend to explore infill-asymptotic results under the mixing (and ergodic) environment without such a martingale assumption. Given the nature of infill asymptotics, the dependence between consecutive observations typically becomes stronger as $n \to \infty$. This leads to slower convergence rates of the sums (relative to those obtained under the fixed-distance assumption).

While our results for the infill case can be seen as a direct extension of results for the fixed-distance case, they are new and particularly useful to derive sharper convergence rates of estimators for continuous-time processes from discretely sampled observations. For this infill case, we also assume that the time horizon of an observation period tends to infinity (the long-span assumption). This double-asymptotic scheme is important for estimating continuous-time processes, where we note that the long-span assumption is often necessary for the drift estimation (see p. 243 of Bandi and Phillips, 2003; Section 5 of Kristensen, 2010). While certain parts of a model structure (such as volatility) may be identified and consistently estimated under only the infill assumption (over a fixed time span), several economically interesting parameters, including risk-attitude parameters and the market price of risk (e.g. Stanton, 1997), can be identified only through agents' behavior or price movement in the long

[3] Phillips (1987) showed that the infill limit of a regression estimator was represented as a functional of a Brownian motion. He discussed that such a representation was particularly useful to explain the estimator's (finite-sample) behavior.

run, necessitating the long-span assumption. Estimation of such parameters often requires that of drift components. We discuss and exemplify the usefulness of our convergence results for the infill case in a non-parametric drift estimation problem. We specifically consider a kernel-based estimator for Park's (2009) martingale regression model, which encompasses a wide class of continuous-time models found in the economics/finance literature, and we show that our new results lead to a sharper convergence rate of the non-parametric estimator's discretization bias.

In the next section, we introduce our basic framework and derive our primary convergence results for both the fixed-distance and infill cases. We also compare the previous mixingale results with ours and present some examples of processes to which our results are applicable. In Section 3, we consider the drift function estimation of continuous-time processes and discuss the usefulness of the derived results. Section 4 provides some concluding remarks. Proofs, as well as auxiliary discussions and results, can be found in the Appendix.

2 Primary convergence results

2.1 Basic setup

Let $(\Omega, \mathcal{F}, \Pr)$ denote a probability space and let $\{X_{n,i}\}$ $(:= \{X_{n,i} : i = 1, \ldots, n,\ n \in \mathbb{N}\})$ be a triangular array of random variables on $(\Omega, \mathcal{F}, \Pr)$. Let ${}^{n}\mathcal{F}_{j}^{k}$ be the σ-algebra generated by $X_{n,j}, X_{n,(j+1)}, \ldots, X_{n,k}$ and define the α-mixing coefficients of $\{X_{n,i}\}$ by

$$\alpha_n(m) := \sup_{1 \le k \le n-m} \sup \left\{ \left| \Pr(A \cap B) - \Pr(A)\Pr(B) \right| : A \in {}^{n}\mathcal{F}_{m+k}^{n},\ B \in {}^{n}\mathcal{F}_{1}^{k} \right\}.$$

This is a standard definition of triangular arrays (see Bradley, 2005). The mixing coefficients represent the degree of time-series dependence of $\{X_{n,i}\}$. We note that the coefficients may depend on n, since ${}^{n}\mathcal{F}_{m+k}^{n}$ and ${}^{n}\mathcal{F}_{1}^{k}$ are allowed to depend on n. The dependence on n is relevant in this study, since we consider the cases where the degree of time-series dependence between consecutive observations changes with n (as in Example 2 of a near-unit-root process). It is also relevant to the infill assumption under which we let time intervals between adjacent observations shrink to zero as the sample size n tends to infinity. To accommodate this case, we suppose that there exists some function $\bar{\alpha}(\cdot)$ which does not depend on m and n, such that $\alpha_n(m) \le \bar{\alpha}(m\Delta)$, where $\Delta = \Delta_n := T/n\,(> 0)$ is the time interval of adjacent observations and $T\,(:= T_n)$ is the time horizon of the observation period. We specifically consider two cases: (I) Δ is fixed; and (II) Δ shrinks to zero as $n \to \infty$. The array $\{X_{n,i}\}$ is said to be α-mixing if $\bar{\alpha}(m\Delta) \to 0$ as $m\Delta \to \infty$. The second case is often relevant in estimating continuous time processes from discretely sampled observations, where we need the infill assumption to kill biases (due to the discretization) and achieve consistent estimation.[4]

We derive the rate of the sum $\sum_{i=1}^{n} X_{n,i}$ for both cases. For our purposes, we impose the following conditions:

[4]The α-mixing coefficients of a continuous time process are defined analogously: for a process $\{Z_s : s \ge 0\}$ defined on a filtered probability space $(\Omega, \mathfrak{F}, \{\mathfrak{F}_s\}, \Pr)$,

$$\alpha_Z(t) := \sup\nolimits_{s \ge 0} \sup \left\{ \left| \Pr(A \cap B) - \Pr(A)\Pr(B) \right| : A \in \mathcal{F}_{t+s}^{\infty},\ B \in \mathcal{F}_{0}^{s} \right\},$$

where $\mathcal{F}_{t}^{u}$ is the σ-algebra generated by $\{Z_s : s \in [t, u]\}$. If $X_{n,i} = Z_{i\Delta}$, we can set $\bar{\alpha}(\cdot) = \alpha_Z(\cdot)$.

A1. Let $\{X_{n,i}\}$ be an α-mixing triangular array of mean-zero random variables that are L_p-integrable with $p \geq 1$. Its mixing coefficients satisfy

$$\alpha_n(m) \leq \bar{\alpha}(m\Delta) := A(m\Delta)^{-\beta}, \tag{1}$$

for some $A > 0$ and $\beta > 0$, where $\Delta (= T/n > 0)$ is the time interval between the adjacent observations and T is the time horizon of the observation period.

A2. For a triangular array $\{X_{n,i}\}$, there exist some sequences of positive real numbers, $\{b_n\}$ and $\{\rho_n\}$, such that as $n \to \infty$,

$$\rho_n^{-1}\sum\nolimits_{i=1}^{n} X_{n,i} 1\{|X_{n,i}| > b_n\} = O_p(1). \tag{2}$$

If Δ is fixed in Condition A1, we can set $\Delta = 1$ without loss of generality. When we let $\Delta \to 0$ (the infill assumption), we also let $T \to \infty$ as $n \to \infty$ (the long-span assumption). If we impose an additional assumption, such as a sort of martingale-difference or (semi) martingale one, it may be possible to establish the LLN and explore the rate of the sum under infill asymptotics without long-span asymptotics, as in volatility estimation (see, e.g., Barndorff-Nielsen and Shephard, 2002; Kristensen, 2010). However, given fixed T, we cannot exploit the asymptotic independence implied by the mixing condition, and it is uncertain whether we could establish a sensible LLN result for general α-mixing arrays under only infill asymptotics. We assume polynomial decay of $\alpha_n(m)$ and its dependence on n only through Δ in (1). This is only for simplicity. Our techniques based on the Bernstein-type inequality work even without these assumptions, as discussed and exemplified in Remark 1 and Example 2. The sequence $\{b_n\}$ in Condition A2 is used for truncating random variables, by which we control tail behavior of the sum of $\{X_{n,i}\}$ when applying the Bernstein-type inequality. If $\{X_{n,i}\}$ is almost surely bounded uniformly over any n and i, then (2) is satisfied by any ρ_n with $b_n = O(1)$.

2.2 Convergence rates of sums with fixed observation intervals

Given the stated conditions, we can obtain the following theorem:

Theorem 1. *Suppose that Conditions A1 and A2 hold and that Δ is a positive constant. Then, for any sequences $\{\rho_n\}$ and $\{b_n\}$ which satisfy (2) and*

$$\frac{b_n}{\rho_n} = \begin{cases} O\left(n^{-1/(1+\beta)}\right) & \text{for } \beta \in (0,1), \\ O(1/\sqrt{n \log n}) & \text{for } \beta \geq 1, \\ O(1/\sqrt{n}) & \text{for } \beta > 1, \end{cases} \tag{3}$$

it holds that $\sum_{i=1}^{n} X_{n,i} = O_p(\rho_n)$ as $n \to \infty$.

The rate of ρ_n may be written without b_n if the uniform moment bound of $\{X_{n,i}\}$ exists. Suppose that each $X_{n,i}$ is L_p integrable with $p > 1$ and let

$$g_n = g_n(p) \geq \max\nolimits_{1\leq i\leq n} \|X_{n,i}\|_p, \tag{4}$$

where $\|\cdot\|_p := \{E[|\cdot|^p]\}^{1/p}$. Then, Condition A2 is satisfied under $n g_n^p/\rho_n b_n^{p-1} = O(1)$, since we have

$$\begin{aligned}\Pr\left(\rho_n^{-1}\sum\nolimits_{i=1}^{n} X_{n,i} 1\{|X_{n,i}| > b_n\} \geq a\right) &\leq a^{-1}\rho_n^{-1} E\left[\left|\sum\nolimits_{i=1}^{n} X_{n,i}\{1\,|X_{n,i}| > b_n\}\right|\right] \\ &\leq a^{-1} n g_n^p/\rho_n b_n^{p-1},\end{aligned}$$

which can be made arbitrarily small for any sufficiently large $a > 0$. In this case, we can write ρ_n as:

$$\rho_n = \begin{cases} n^{(p+\beta)/p(1+\beta)} g_n & \text{for } \beta \in (0,1), \\ n^{(p+1)/2p} (\log n)^{(p-1)/2p} g_n & \text{for } \beta \geq 1, \\ n^{(p+1)/2p} g_n & \text{for } \beta > 1. \end{cases} \tag{5}$$

We note that rates in (5) can be obtained by solving $ng_n^p/\rho_n b_n^{p-1} = O(1) (\neq o(1))$ and (3) with respect to ρ_n for each case, where each rate on the right-hand side (RHS) of (3) should be interpreted as a big-$O(\cdot)$ order of the relevant rate (not a small-$o(\cdot)$ one). This (5) indicates an explicit trade-off between the degree of dependence and the rate of convergence: the smaller β (i.e., the stronger time-series dependence) implies the faster divergence rate of ρ_n (i.e., the slower convergence rate of the sample mean). If $\{X_{n,i}\}$ is uniformly bounded, the sum of the array has the rate of $O_p(\sqrt{n})$ for $\beta > 1$, which corresponds to the classical CLT (Theorem 18.5.4 of Ibragimov and Linnik, 1971). Some other remarks on Theorem 1 are in order:

Remark 1. The polynomial decay condition (1) on mixing coefficients can be relaxed in Theorem 1 (and also in Theorems 2 and 4 - 6). The decay rate may be arbitrarily slower than any polynomial rate. Given the condition that $\alpha_n(m) \leq \bar{\alpha}(m\Delta) \to 0$, as $m\Delta \to \infty$ in Theorem 1, instead of (1) in Condition A1, we can show that $\sum_{i=1}^n X_{n,i} = O_p(\rho_n)$ if ρ_n satisfies (2) and $(n^2 b_n^2/\rho_n^2)\, \bar{\alpha}(\rho_n^2/nb_n^2) = O(1)$ (the proof of this result proceeds in the same way as that of Theorem 1 and is omitted).

Remark 2. If $\{X_{n,i}\}$ is uniformly integrable (UI), then for any sequence $b_n \to \infty$, we can always find some ρ_n satisfying (2) with $\rho_n = o(n)$. Given this fact, together with the argument in Remark 1, we can show that $\sum_{i=1}^n X_{n,i} = o_p(n)$ if $\{X_{n,i}\}$ is UI and $\bar{\alpha}(m\Delta) \to 0$ as $m\Delta \to \infty$. No decay rate on $\bar{\alpha}(m\Delta)$ is imposed here. This is an α-mixing counterpart of Andrews' (1988) LLN, where he proved that the weak LLN holds if an L_1-mixingale zero-mean array is UI and the average of its scaling constants has a finite limsup (we again recall that no decay rate is imposed on the mixingale numbers in Andrews' LLN).

The result in (5) also indicates that the rate of ρ_n improves if a higher order moment of $\{X_{n,i}\}$ exists. If $p > 2$, we can obtain further improvements as follows:

Theorem 2. *Suppose that Condition A1 holds with $p > 2$ and Δ is a positive constant. Then, for a sequence $\{\rho_n\}$ such as*

$$\rho_n = \begin{cases} n^{(p+\beta)/p(1+\beta)} g_n & \text{for } \beta \in (0, p/(p-2)), \\ \sqrt{n \log n}\, g_n & \text{for } \beta \geq p/(p-2), \\ \sqrt{n} g_n & \text{for } \beta > p/(p-2), \end{cases} \tag{6}$$

where g_n is given in (4), it holds that $\sum_{i=1}^n X_{n,i} = O_p(\rho_n)$ as $n \to \infty$.

While the rate $\rho_n = n^{(p+\beta)/p(1+\beta)} g_n$ for $\beta \in (0, p/(p-2))$ in this theorem is the same as the one given in (5), "$\beta \geq 1$" is allowed here. The rate for $\beta > p/(p-2)$ also corresponds to the classical CLT: $\sum_{i=1}^n X_{n,i} = O_p(\sqrt{n})$ if $g_n = O(1)$ (see Theorem 18.5.3 of Ibragimov and Linnik, 1971 or Theorem 1.7 of Bosq, 1998). Note that for arrays whose degree of dependence is weak (i.e., ones with β large enough), various results on the strong convergence exist in the literature, as stated previously. Some other remarks on Theorems 1-2 follow:

Remark 3. From the viewpoints of the analysis in Remark 2 and the classical CLT results in Ibragimov and Linnik (1971), we can say that the results of Theorems 1 and 2 for $\beta \in (0,1]$ and $\beta \in (0, p/(p-2)]$, respectively, fill the gap between the LLN and the CLT. We also note that according to Bradley (1985, Remark 1) and Bosq, Merlevède, and Peligrad (1999, Remark 1.1), $\sum_{m=1}^{\infty} m\bar{\alpha}(m) < \infty$ (resp. $\sum_{m=1}^{\infty} m\bar{\alpha}^{(p-2)/p}(m)$ with $p > 2$) is essentially a minimal requirement for an α-mixing sequence/array of bounded (resp. L_p-bounded) random variables to satisfy a non-degenerate CLT under the condition that $\mathrm{Var}[\sum_{i=1}^{n} X_{n,i}] \to \infty$.

Remark 4. The degree of heterogeneity in $\{X_{n,i}\}$ is captured and controlled by the truncation constant b_n and the moment bound g_n, respectively, in Theorems 1 and 2, which are uniform over i. Overall, b_n and g_n play a role similar to the scaling constants in the mixingale cases, as in Andrews (1988), Hansen (1991, 1992), Davidson (1993), Davidson and de Jong (1997), and de Jong (1995, 1996, 1998). In some of these studies, such constants may depend on each of the observations. Analogously, we could potentially let b_n and/or g_n be dependent on each i. This manner of treating heterogeneity may allow for more flexibility, but we do not pursue this direction, since it makes conditions and proofs more complicated. Our treatment of heterogeneity seems sufficient in many applications.

Remark 5. The results for $\beta \geq p/(p-2)$ in Theorem 2 can be strengthened to the L_2 convergence, which is a direct consequence of Davydov's inequality for covariances (see the proof of Theorem 2). While the same inequality also allows us to derive an L_2-convergence rate for $\beta \in (0, p/(p-2))$, it is inferior to $\rho_n = n^{(p+\beta)/p(1+\beta)} g_n$ in (6), which seems to illustrate the price of using the stronger notion of the convergence.

Comparison with previous mixingale LLN results. Here, we compare our convergence results of Theorems 1 and 2 with previous mixingale LLN results and discuss advantages of our results in mixing cases. As stated in the Introduction, Davidson and de Jong (1997) and de Jong (1998) investigated convergence rates of sums of mixingale sequences/arrays. Since α-mixing arrays can also be mixingale, we can apply Davidson and de Jong's results (and other mixingale LLN results) to the α-mixing cases. To understand this point, we recall the following fact: if $\{X_{n,i}\}$ is a zero-mean array with each $X_{n,i}$ being L_p-bounded ($p > 1$) and measurable with respect to ${}^{n}\mathcal{F}_{1}^{i}$, and if its mixing coefficients satisfy Conditions A2 (i.e., $\alpha_n(m) \leq Am^{-\beta}$, where we set $\Delta = 1$ for simplicity), then it is also L_q-mixingale with its mixingale numbers ζ_m satisfying

$$\zeta_m \leq \bar{A} m^{-\beta(1/q - 1/p)} \quad \text{for } 1 \leq q < p, \tag{7}$$

which follows from Lemma 2.1 of McLeish (1975, p. 834; see also Section 16.1 of Davidson, 1994). To the author's knowledge, this inequality is the best available result that relates an α-mixing rate to a mixingale one. For the definition of L_q-mixingale arrays, see Andrews (1988) or Section 16 of Davidson (1994).

The inequality (7) implies that we can also apply mixingale LLN results to α-mixing arrays. However, if we do so, some stronger conditions will be needed and/or less sharp rates will be obtained than if we would directly apply the LLN results tailored to the mixing arrays as developed here. For example, to apply Andrews' (1988) LLN to an α-mixing array $\{X_{n,i}\}$, we need to impose the L_p-boundedness with $p > 1$, which is stronger than necessary, while the original Andrews' LLN holds under the L_1-boundedness (see also Remark 2 above).

To provide another example, we compare our LLN rates with those by Davidson and de Jong. While their results and ours are not necessarily immediately comparable (as they consider the almost-sure and L_p convergence concepts), we compare our L_2 rates in Remark 5 with those in de Jong (1998), which seems to lead to a fair comparison, illustrating advantages of our LLN theorems rather than their mixingale LLNs in the mixing environment. Suppose that $\{X_{n,i}\}$ satisfies Condition A1 with $\beta > p/(p-2)$ $(p>2)$ and $g_n(p) = O(1)$. In this case, by Remark 5, it holds that $\|\sum_{i=1}^{n} X_{n,i}\|_2 = O(\sqrt{n})$. On the other hand, (7) implies that $\{X_{n,i}\}$ is L_2-mixingale with $\zeta_m \leq \bar{A} m^{-\beta(1/2-1/p)}$, and the application of de Jong's (1998, Theorem 7) mixingale LLN leads to

$$\|\sum_{i=1}^{n} X_{n,i}\|_2 = \begin{cases} O(n^{1-\beta(1/2-1/p)/2}) & \text{for } \beta \in (p/(p-2), 2p/(p-2)), \\ O(\sqrt{n \log n}) & \text{for } \beta \geq 2p/(p-2), \\ O(\sqrt{n}) & \text{for } \beta > 2p/(p-2), \end{cases} \tag{8}$$

whose proof is provided in Appendix A.4. Rates in (8) are inferior to $O(\sqrt{n})$ for $\beta \leq 2p/(p-2)$. de Jong's theorem requires $\beta > 2p/(p-2)$ to obtain the sharp rate $\sqrt{n}$, which is stronger than necessary, while we can check that the same rate is attained for $\beta > p/(p-2)$.

These two comparisons suggest the benefits of using our results. We do not claim that our convergence theorems dominate previous mixingale LLNs, but we argue that they are more likely to derive superior/sharper results if an array in question is mixing (rather than in the case when the mixingale LLNs were applied via (7)). We note that it is relatively easy to check mixing properties/rates of moving-average, autoregressive and Markov processes, as a number of sufficient conditions are available in the literature.

Examples of α-mixing arrays. Before concluding this subsection, we provide two examples that have slowly decaying mixing coefficients. Example 1 satisfies Conditions A1-A2, to which our new theorems are directly applicable. Example 2 is a near-unit-root (near-nonstationary) process. While this process does not satisfy (1) of Condition A1 in that $\alpha_n(m)$ depends directly on n (even when Δ is a constant, say $\Delta = 1$), the Bernstein-type inequality can still be employed to derive a LLN result. We take up this example to illustrate the usefulness of this technical device.

Example 1: A moving-average process with long lasting past shocks. Let $\{X_{n,i}\}$ be described by

$$X_{n,i} = X_i := \sum_{j=0}^{\infty} c_j \varepsilon_{i-j},$$

where $\{\varepsilon_i\}$ is a sequence of independent random variables each of which has the probability density f_i with

$$\int_{-\infty}^{\infty} |f_i(x) - f_i(x+y)| dx \leq C|y| \quad \text{for any } i \text{ (with some constant } C>0)$$

and satisfies $\sup_{i\in\mathbb{Z}} E[|\varepsilon_i|^{\delta}] < \infty$ for some $\delta > 0$, and the moving-average coefficients satisfy $\sum_{j=0}^{\infty} c_j z^j \neq 0$ for $|z| \leq 1$ $(c_j \in \mathbb{R})$.

Given these settings, we can derive the decay rate of $\alpha_n(m)$ in terms of $\{c_j\}$ by Gorodetskii' (1977) theorem. When effects of past shocks ε_{i-j} do not die out sufficiently fast, $\alpha_n(m)$ decays slowly (as $m \to \infty$). For example, if c_j decays only polynomially: $c_j = O(j^{-q})$ as $j \to \infty$ with "$q > 3/2$" and

"$\delta > 2/(q-1)$ or $\delta \geq 4$," then it holds that

$$\alpha_n(m) \leq \bar{\alpha}(m\Delta) = \begin{cases} O(m^{-[\delta(q-1)-2]/(1+\delta)}) & \text{for } \delta < 2q+1, \\ O(m^{-q+3/2}\sqrt{\log m}) & \text{for } \delta \geq 2q+1. \end{cases}$$

We note that the polynomial decay of c_j (with small q) represents a case in which the effects of past shocks $\{\varepsilon_i\}$ do not die out rapidly. If we have $\delta > 1$ and $E[\varepsilon_i] = 0$, the result (5) or (6) can be applied with $\delta = p$ and

$$g_n = \left\{ \sum_{j=0}^{\infty} |c_j| \sup_{i\in\mathbb{Z}} E[|\varepsilon_i|^{\delta}] \Big/ \sum_{k=0}^{\infty} |c_k| \right\}^{1/\delta} = O(1),$$

which can be derived by Jensen's inequality. If q and/or δ are not sufficiently large, usual CLTs cannot be used but our previous theorems may still be applied.

Example 2. Here, we consider a **near-unit-root process**, as in Stock (1991), Elliott, Rothenberg, and Stock (1996), and Phillips and Magdalinos (2007; PM, henceforth):

$$X_{n,i} = \theta_n X_{n,i-1} + u_i \quad \text{and} \quad \theta_n = 1 - \bar{c}/k_n, \tag{9}$$

where $\{u_i\}_{i\geq 1}$ is a sequence of independent random variables with $E[u_i] = 0$ and $\max_{i\geq 1} E\left[|u_i|^{\delta}\right] < \infty$ for $\delta > 0$; $X_0 = o_p(\sqrt{k_n})$ is independent of $\{u_i\}$ with $E[X_0] = 0$; k_n is a sequence increasing to ∞ such that $k_n = o(n)$ or $= n$; and $\bar{c} > 0$. This specification of $\{X_{n,i}\}$ follows that of PM (while we allow k_n to be n but additionally suppose that $E[X_0] = 0$ and $\bar{c} > 0$).

If $1/k_n$ is some positive constant (independent of n), $\{X_{n,i}\}$ is a standard autoregressive process of order 1. Its mixing coefficients decay exponentially fast, which follows from Gorodetskii's (1977) theorem. For the case $1/k_n = o(1)$, the decay rate of mixing coefficients can be derived in the same way (as long as $\bar{c} > 0$) but it depends on n. Specifically, the coefficient of $\{X_{n,i}\}$ satisfies

$$\alpha_n(m) = O(\exp\{\delta(1+\delta)^{-1} m(\log\theta_n)\}) \leq O(\exp\{-\delta(1+\delta)^{-1}\bar{c}m/k_n\}), \tag{10}$$

which can also be derived by Gorodetskii's (1977) theorem (as it leads to the upper bound of the α-mixing coefficient for each n), where we have let $\Delta = 1$ for simplicity (the inequality holds since $\log\theta_n \leq -\bar{c}/k_n$). We note that (9) behaves like a stable/ergodic autoregressive process in that its mixing coefficients decay exponentially as m grows for each (fixed) n, while it represents a near-nonstationary situation in which its autoregressive coefficient θ_n is (very) close to 1 and the convergence of the mixing coefficients does not occur quickly for large n (or k_n). The convergence rate in the LLN involves k_n, as follows:

Theorem 3. *Suppose that $\{X_{n,i}\}$ is specified by (9) with $\delta \geq 2$. Then, $\sum_{i=1}^{n} X_{n,i} = O_p(\rho_n)$ with $\rho_n = \sqrt{k_n n}$ as $n \to \infty$.*

The condition "$\delta \geq 2$" is imposed only for simplicity (it may be relaxed, but a smaller δ will lead to a less sharp rate). This theorem may complement PM's limit results for (9). It seems that the literature on processes such as (9) have not necessarily paid attention to the mixing property of the processes, which can still be used for establishing asymptotic results.

2.3 Convergence rates of sums under infill and long-span asymptotics

Here, we consider an array whose observation intervals shrink to zero as $n \to \infty$. The results obtained here are useful for analyzing convergence rates of estimators for continuous-time processes, as we see in the next section.

We set out an additional condition:

A3. $\Delta^{-1} \leq c_1 T^{c_2}$ for some $c_1, c_2 > 0$ (as $\Delta \to 0$ and $T \to \infty$ with $n \to \infty$).

This condition may slightly simplify rate expressions in Theorems 4 and 5 for the intermediate cases $\beta \geq 1$ and $\beta \geq p/(p-2)$, respectively. We can derive convergence results without A3. However, this bound of Δ^{-1} can be used to control the degree of dependence in arrays: if a divergence rate of Δ^{-1} is fast beyond this condition, the dependence between consecutive observations becomes too strong, which may hamper establishing simple and sharp rate expressions. We also note that there should be no practical restriction due to Condition A3, since c_1 and c_2 may be arbitrarily large.

We subsequently write the rate of ρ_n (given in Condition A2) in terms $T(:= T_n)$ and $\Delta(:= \Delta_n)$. While it may be more reasonable to write $\rho_{T,\Delta}$ instead of ρ_n (as pointed out by a referee), we use ρ_n to avoid writing the same condition using a different notation (the same remark applies to b_n, g_n, and δ_n, which are given in Condition A2, (4) and Condition A4 below).[5]

Now, analogously to the fixed-distance case, we present two additional theorems:

Theorem 4. *Suppose that Conditions A1-A2 hold. Then, for any sequences $\{\rho_n\}$ and $\{b_n\}$ satisfying (2) and*

$$\frac{b_n}{\rho_n} = \begin{cases} O\left(T^{-1/(1+\beta)}\Delta\right) & \text{for } \beta \in (0,1), \\ O(\Delta/\sqrt{T \log(T/\Delta)}) & \text{for } \beta \geq 1, \\ O\left(T^{-1/2}\Delta\right) & \text{for } \beta > 1, \end{cases} \tag{11}$$

it holds that $\sum_{i=1}^{n} X_{n,i} = O_p(\rho_n)$ as $T \to \infty$ and $\Delta \to 0$ with $n \to \infty$. If Condition A3 is additionally supposed, then $b_n/\rho_n = O(\Delta/\sqrt{T \log T})$ for $\beta \geq 1$ in (11).

Due to the infill assumption, the rate of ρ_n depends on the shrinking rate of Δ. Given that β is fixed, the smaller Δ leads to the faster divergence rate of ρ_n (or equivalently, the slower shrinking rate of the sample mean). This is because a small Δ increases the degree of time-series dependence in the array and makes consecutive observations highly correlated.

Analogously to (5), if $p > 1$, we can write ρ_n in Theorem 4 as follows:

$$\rho_n = \begin{cases} T^{(p+\beta)/p(1+\beta)}\Delta^{-1} g_n & \text{for } \beta \in (0,1), \\ T^{(p+1)/2p}[\log(T/\Delta)]^{(p-1)/2p}\Delta^{-1} g_n & \text{for } \beta \geq 1, \\ T^{(p+1)/2p}\Delta^{-1} g_n & \text{for } \beta > 1, \end{cases} \tag{12}$$

where $\rho_n = T^{(p+1)/2p}(\log T)^{(p-1)/2p}\Delta^{-1} g_n$ under Condition A3. This result may not fully exploit the existence of higher order moments. If $p > 2$, the rate of ρ_n is improved:

[5] We also note that the rates of $T(= T_n)$ and $\Delta(= \Delta_n)$ may be written in terms of n through the definition $n = T/\Delta$ and Condition A3. Then, we can regard $T_n \to \infty$ and $\Delta_n \to 0$ as $n \to \infty$.

Theorem 5. *Suppose that Conditions A1 holds with $p > 2$. Then, for a sequence $\{\rho_n\}$, such as*

$$\rho_n = \begin{cases} T^{(p+\beta)/p(1+\beta)}\Delta^{-1}g_n & \text{for } \beta \in (0, p/(p-2)), \\ \sqrt{T\log(T/\Delta)}\Delta^{-1}g_n & \text{for } \beta \geq p/(p-2), \\ T^{1/2}\Delta^{-1}g_n & \text{for } \beta > p/(p-2), \end{cases} \tag{13}$$

where g_n is given in (4), it holds that $\sum_{i=1}^{n} X_{n,i} = O_p(\rho_n)$, as $T \to \infty$ and $\Delta \to 0$ with $n \to \infty$. If Condition A3 is additionally supposed, then $\rho_n = \sqrt{T\log T}\Delta^{-1}g_n$ for $\beta \geq p/(p-2)$ in (13).

As can be seen in Theorems 4-5, the sample mean of the array is not convergent unless $T \to \infty$ (e.g., $n^{-1}\sum_{i=1}^{n} X_{n,i} = O_p(g_n/\sqrt{T})$ for $\beta > p/(p-2)$ in Theorem 5). Note that both the infill and long-span assumptions are often necessary in drift function estimation, as in the next section (see p. 243 of Bandi and Phillips, 2003; Section 5 of Kristensen, 2010).

Remark 6. The rate for $\beta > p/(p-2)$ is sharp and corresponds to the rate obtained in the CLT for continuously observed processes. To see this point, let $X_{n,i}^{a} := \int_{\Delta 2(i-1)}^{\Delta(2i-1)} Z_s ds$ and $X_{n,i}^{b} := \int_{\Delta(2i-1)}^{\Delta 2i} Z_s ds$, where $\{Z_s\}_{s\geq 0}$ is a zero-mean continuous time process with $\sup_{s\geq 0}\|Z_s\|_p = O(1)$ for $p > 2$, whose mixing coefficients satisfy $\alpha_Z(s) \leq As^{-\beta}$. We can then write

$$\int_0^T Z_s ds = \sum_{i=1}^{n}(X_{n,i}^{a} + X_{n,i}^{b}) + \int_n^T Z_s ds.$$

Applying Theorem 5, we obtain $\int_0^T Z_s ds = O_p(\sqrt{T})$, since $\sum_{i=1}^{n} X_{n,i}^{k} = O_p(T^{1/2}\Delta^{-1}\sup_{i\geq 1}\|X_{n,i}^{k}\|_p) = O_p(\sqrt{T})$ for $k = a, b$, $\sup_{i\geq 1}\|X_{n,i}^{k}\|_p = O(\Delta)$ and $\int_n^T Z_s ds = O_p(1)$. Note that the rate of $\int_0^T Z_s ds$ may also be derived using Theorems 1 or 2 analogously.

Before concluding this subsection, we present a simple example of a continuous-time process to which our convergence results are relevant.

Example 3: A diffusion process with weak mean-reversion drift. Let $\{Z_s\}_{s\geq 0}$ be an $\mathbb{R}$-valued, continuous-time diffusion process defined through (a weak solution to) the following stochastic differential equation:

$$dZ_s = \mu_Z(Z_s)\,ds + dW_s,$$

where $\mu_Z(\cdot)$ is the drift function; $\{W_s\}_{s\geq 0}$ is a standard Brownian motion; and the process is supposed to start from some constant $z_0 \in \mathbb{R}$, i.e., $Z_0 = z_0$.[6] If the drift function satisfies

$$\mu_Z(z) \leq -r/z \ \text{ for } z \geq M_0, \text{ and } \ \mu_Z(z) \geq -r/z \ \text{ for } z \leq -M_0, \tag{14}$$

for some $r > 1/2$ and $M_0 \geq 0$, then we can find the decay rate of α-mixing coefficients of the process $\{Z_s\}$ and its uniform moment bound:

$$\alpha_Z(t) \leq At^{-\beta} \ \text{ for any } \beta < (0, r-1/2);$$
$$\sup_{s\geq 0} E[|Z_s|^p] < \infty \ \text{ for any } p \in (0, 2r),$$

[6] We can also think of more general cases: (i) Z_0 follows some distribution, (ii) $\{Z_s\}$ is a general diffusion process with its non-constant diffusion function (say, $dZ_s = \mu_Z(Z_s)\,ds + \sigma_Z(Z_s)\,dW_s$), and/or (iii) it is of multi-dimension. Mixing rates and moment bounds for these cases are investigated in Veretennikov (1997).

where $A\,(>0)$ is some positive constant and $\alpha_Z(t)$ stands for the mixing coefficient for the continuous-time process $\{Z_s\}$ (see the definition in Footnote 4). These results follow from Section 5.3 of Chen, Hansen, and Carrasco (2012) and Lemma 1 of Veretennikov (1997). In this example, the mixing rate and the moment existence rely directly on the parameter r. The condition (14) on $\mu_Z(\cdot)$ represents a case in which the drift function's effect of the mean-reversion is very weak.[7] If $\{X_{n,i}\}$ is an array of (normalized) discrete-time observations from $\{Z_s\}$, say, $X_{n,i} = (Z_{\Delta i} - E[Z_{\Delta i}])$, then its mixing coefficients satisfy $\alpha_n(m) \leq A(m\Delta)^{-\beta}$ and the uniform moment bound $g_n = g_n(p)$ is well-defined for $p \in (0, 2r)$. Therefore, we can apply the rate results of (12) and (13). In particular, for a small r, we cannot obtain a sufficiently fast convergence rate.

3 Non-parametric drift function estimation of continuous-time processes

3.1 Continuous-time framework

We now apply the results from the previous section to the drift function estimation for continuous-time processes. We pay particular attention to processes with (relatively) strong time-series dependence and non-existence of higher-order moments, both of which may be prominent features in financial data. We consider the following type of continuous-time process:

$$dY_s = \mu(Z_{s-})\,ds + dU_s, \tag{15}$$

where $\{Y_s\}$ and $\{Z_s\}$ are real-valued càdlàg processes defined on a filtered probability space $(\Omega, \mathfrak{F}, \{\mathfrak{F}_s\}_{s\geq 0}, \Pr)$, and $\{U_s\}$ is a martingale process on the same space. We suppose that each process is adapted to the filtration $\{\mathfrak{F}_s\}$ and $\mu(\cdot)$ is a continuous function. $\mu(\cdot)$ is termed the drift (or instantaneous conditional mean) function. This general specification, called the martingale regression, has been proposed by Park (2009) (see also Kim and Park, 2016). It includes many interesting models used in the fields of economics and finance. A leading example is a univariate diffusion process with $Y_s = Z_s$ and $dU_s = \sigma(Z_s)\,dW_s$, where $\{W_s\}$ is a standard Brownian motion and $\sigma(\cdot)$ is a volatility function.[8] As an example with $Y_s \neq Z_s$, we note that (15) may be used to construct a continuous-time analog of a long run risk model, as in Bansal and Yaron (2004). For other examples, see our subsequent arguments on Conditions B4 and B5 as well as Park (2009).

Given a set of observations sampled at discrete-time points, $\{(Y_{i\Delta}, Z_{i\Delta}) : 1 \leq i \leq n+1\}$ (with $n = T/\Delta$), we consider the following non-parametric estimator of the drift function $\mu(\cdot)$:

$$\hat{\mu}(z) := \hat{\eta}(z)/\hat{\pi}(z),$$

[7] For instance, compare (14) with a linear-drift case (possessing a strong mean-reversion property): $\tilde{\mu}_Z(x) = \kappa(m - z)$ with $\kappa > 0$ and $m \in \mathbb{R}$. Given this drift specification, the process is pulled toward m: for a sufficiently large positive (resp. small negative) value of z, $\tilde{\mu}_Z(x)$ takes a very small negative (resp. large positive) value.

[8] We note that the specification (15) allows for general diffusion and stochastic volatility processes. This can be understood by writing dU_s in terms of a (formal) integral expression. For example, in the case of a stochastic volatility process with a spot volatility process $\{\sigma_s\}$, we have $U_t - U_0 = \int_0^t \sigma_s dW_s$, which is equivalently written as $dU_s = \sigma_s dW_s$ by convention.

where

$$\hat{\eta}(z) := \frac{1}{T}\sum\nolimits_{i=1}^{n} K_h(Z_{i\Delta} - z)\left[Y_{(i+1)\Delta} - Y_{i\Delta}\right];$$
$$\hat{\pi}(z) := \frac{\Delta}{T}\sum\nolimits_{i=1}^{n} K_h(Z_{i\Delta} - z);$$

$K_h(x) := K(x/h)/h$; K is a kernel function; and h is a smoothing parameter (bandwidth). By using the results obtained in the previous section, we investigate the convergence rate of $\hat{\mu}(z)$. To this end, we impose the following conditions:

B1. The support of $K(: \mathbb{R} \to \mathbb{R})$ is bounded (say, it is included in $[-c_K, c_K]$ for some $c_K \in (0,\infty)$), and satisfies $\int_{-\infty}^{\infty} K(x)\,dx = 1$; $\int_{-\infty}^{\infty} xK(x)\,dx = 0$; and $\sup_{x\in\mathbb{R}} |K(x)| \leq \bar{K}$ for some $\bar{K}\,(< \infty)$.

B2-i. $\{Z_s\}_{s\geq 0}$ is an α-mixing process whose mixing coefficients satisfy $\alpha_Z(t) \leq At^{-\beta}$ for some $A, \beta > 0$.

B2-ii. The drift function $\mu(z)$ is differentiable at z and each Z_s has the marginal density $\pi_s(\cdot)$; and there exists some $\bar{\epsilon}(> 0)$, such that $\sup_{|\epsilon|\leq\bar{\epsilon}} |\mu'(\epsilon + z)| < \infty$ and $\sup_{s\geq 0; |\epsilon|\leq\bar{\epsilon}} \pi_s(z + \epsilon) < \infty$.

Condition B1 is standard, except for the boundedness of the support. While some kernels (such as the normal one) are excluded, this condition allows us to work without imposing the boundedness of the moment (say $E[|\mu(Z_s)|^p] < \infty$). If some kernel with an unbounded support is employed, the existence of some (higher-order) moment is likely to be a required condition. Condition B2 allows for some sort of non-stationarity/heterogenous process, e.g., the process need not be initialized by the invariant distribution, while it excludes strongly non-stationary processes (such as null-recurrent processes). While many parametric models used in the econometric literature turn out to be geometrically α-mixing, we can easily find exceptions, as in Example 3 (see also Veretennikov, 1997, 1999). Chen, Hansen, and Carrasco (2010) also present a class of Markov diffusion processes with very slowly decaying mixing coefficients.[9] They observe that such processes look like long memory processes from the vantage point of sample statistics. Note also that diffusion processes with the so-called volatility-induced stationary (see Conley, Hansen, Luttmer, and Scheinkman, 1997), which may look like unit-root processes, often fail to possess higher-order moments, as argued and exemplified in Nicolau (2005). These kinds of processes are also allowed under Condition B2. Additionally, notice that Condition B2-i does not imply the Markov property of processes. While several mixing results available in the literature are derived using the Markov condition, we do not exclude non-Markov cases.

3.2 Effects of discretized observations

In this subsection, we illustrate that our convergence results under the infill assumption "$\Delta \to 0$" (Theorems 4-5) are particularly useful in the context of kernel-based estimation for a continuous-time

[9] Veretennikov (1997, 1999) and Chen, Hansen, and Carrasco (2010) present some conditions for processes to be polynomially β-mixing. Their conditions can be used to check Condition B2-i, since α-mixing coefficients are always smaller than β-mixing ones. In addition, in view of Bradley (2005, Section 4.2), where he observes that various mixing conditions can occur simultaneously at essentially the same decay rate, we note that to investigate a sharp α-mixing rate of a process, it is often sufficient to investigate its sharp β-mixing rate (e.g., if β-mixing coefficients of some process decay at the (exact) rate of t^{-c} with some $c > 0$, then its α-mixing coefficients are less likely to decay at a faster rate of t^{-d} with some $d > c$).

process $\{Z_s\}_{s\geq 0}$. As we can see below, one has to incur some sort of discretization bias in such estimation since only the availability of discretized observations $\{Z_{i\Delta}\}_{i=1}^n$ has been assumed (instead of that of a continuous trajectory). Roughly speaking, we can think of (at least) two ways to verify the estimator's convergence. Here, we show that one approach based on *fully discretized* processes may allow us to derive a sharper rate for the discretization biases than the other one based on *fully continuous-time* processes in the mixing/ergodic environment, and the former approach requires the rate results based on the infill assumption (as Theorems 4-5).

Now, consider the following decomposition of the nonparametric drift estimator:

$$\hat{\mu}(z) - \mu(z) = [\tilde{\xi}(z) + \tilde{\psi}(z)]/\hat{\pi}(z),$$

where

$$\begin{aligned}
\tilde{\xi}(z) &:= \frac{1}{T}\sum\nolimits_{i=1}^{n} K_h(Z_{i\Delta} - z)\int_{i\Delta}^{(i+1)\Delta}[\mu(Z_{s-}) - \mu(z)]\,ds; \\
\tilde{\psi}(z) &:= \frac{1}{T}\sum\nolimits_{i=1}^{n} K_h(Z_{i\Delta} - z)\left[U_{(i+1)\Delta} - U_{i\Delta}\right].
\end{aligned}$$

Note that we have the regressor process $\{Z_t\}$ evaluated at $t = i\Delta$ (a sample time) as well as at all $t \in [i\Delta, (i+1)\Delta]$ (not necessarily sample times) in each summand of $\tilde{\xi}(z)$.

Decomposition with a *fully discretized* process. For the first approach mentioned above, we further decompose $\tilde{\xi}(z)$ into

$$\tilde{\xi}(z) = \frac{1}{nh}\sum\nolimits_{i=1}^{n}\tilde{X}_{n,i} + \frac{1}{nh}\sum\nolimits_{i=1}^{n} E\left[K\left(\frac{Z_{i\Delta} - z}{h}\right)[\mu(Z_{i\Delta}) - \mu(z)]\right] + \frac{1}{Th}\sum\nolimits_{i=1}^{n}\tilde{\kappa}_{n,i}, \tag{16}$$

where

$$\begin{aligned}
\tilde{X}_{n,i} &:= K\left(\frac{Z_{i\Delta} - z}{h}\right)[\mu(Z_{i\Delta}) - \mu(z)] - E\left[K\left(\frac{Z_{i\Delta} - z}{h}\right)[\mu(Z_{i\Delta}) - \mu(z)]\right]; \\
\tilde{\kappa}_{n,i} &:= K\left(\frac{Z_{i\Delta} - z}{h}\right)\int_{i\Delta}^{(i+1)\Delta}[\mu(Z_{s-}) - \mu(Z_{i\Delta})]\,ds.
\end{aligned}$$

We call the first term $(1/nh)\sum_{i=1}^{n}\tilde{X}_{n,i}$ on the RHS of (16) as a *fully discretized* component, as $\{Z_s\}$ is only evaluated at sample times. The convergence rate of this term can be obtained using Theorem 4 or 5. The second term is the bias due to smoothing. The third term $(1/Th)\sum_{i=1}^{n}\tilde{\kappa}_{n,i}$ is the discretization bias that appears only when estimating continuous-time processes from discrete observations. In this discretization term, $\{Z_s\}$ also needs to be evaluated at non-sample points, but such an evaluation occurs only when it is put outside the kernel function.

Decomposition with a *fully continuous-time* process. Another way of decomposition other than (16) is also possible:

$$\tilde{\xi}(z) = \frac{1}{Th}\int_0^T \tilde{X}_s^c ds + \frac{1}{Th}\int_0^T E\left[K\left(\frac{Z_{s-} - z}{h}\right)[\mu(Z_{s-}) - \mu(z)]\right]ds - \frac{1}{Th}\sum\nolimits_{i=1}^{n}\tilde{\kappa}_{n,i}^c, \tag{17}$$

where

$$\tilde{X}_s^c := K\left(\frac{Z_{s-}-z}{h}\right)\left[\mu\left(Z_{s-}\right)-\mu\left(z\right)\right]-E\left[K\left(\frac{Z_{s-}-z}{h}\right)\left[\mu\left(Z_{s-}\right)-\mu\left(z\right)\right]\right];$$

$$\tilde{\kappa}_{n,i}^c := \int_{i\Delta}^{(i+1)\Delta}\left[K\left(\frac{Z_{s-}-z}{h}\right)-K\left(\frac{Z_{i\Delta}-z}{h}\right)\right]\mu\left(Z_{s-}\right)ds.$$

The first term $(1/Th)\int_0^T \tilde{X}_s^c ds$ on the RHS of (17) is defined through an integral (rather than a sum of discretized components), where $\{Z_s\}$ has to be evaluated at all $t\in[0,T]$. The second and third terms on the RHS of (17) are also smoothing and discretization biases, respectively, while their expressions are different from those in (16) due to the different nature of the decomposition.

Comparison of the two decompositions. We can usually show that the convergence rates of $(1/nh)\sum_{i=1}^n \tilde{X}_{n,i}$ and $(1/Th)\int_0^T \tilde{X}_s^c ds$ are the same and that those of the smoothing biases in (16) and (17) are also the same. However, the convergence rate of $(1/Th)\sum_{i=1}^n \tilde{\kappa}_{n,i}$ in (16) may be different from that of $(1/Th)\sum_{i=1}^n \tilde{\kappa}_{n,i}^c$ in (17). In general, we are able to show that the rate of $(1/Th)\sum_{i=1}^n \tilde{\kappa}_{n,i}$ is faster than that of $(1/Th)\sum_{i=1}^n \tilde{\kappa}_{n,i}^c$. This is the advantage of using the decomposition with *fully discretized* processes, whose convergence needs to be analyzed by limit results under the infill scheme, as in Theorem 4 or 5.

If we can find some uniform rate δ_n ($\to 0$ as $n\to\infty$) (or assume its existence), such that

$$\max_{1\le i\le n}\sup_{s\in[i\Delta,(i+1)\Delta]}\left|Z_s - Z_{i\Delta}\right| = O_p(\delta_n) \quad \text{as } T,n\to\infty \text{ and } \Delta\to 0,$$

which may hold for continuous diffusion processes (see our subsequent discussions on B4), then we are able to show that $(1/Th)\sum_{i=1}^n \tilde{\kappa}_{n,i} = O_p(\delta_n)$, as in Theorem 7. However, it is likely that $(1/Th)\sum_{i=1}^n \tilde{\kappa}_{n,i}^c$ can be shown to be only $O_p(\delta_n/h)$ or to shrink at a slower rate. This is because we need to evaluate the difference between Z_{s-} and $Z_{i\Delta}$ (for $\tilde{\kappa}_{n,i}^c$), which is on the inside of $K(\cdot)$, taking into account its interaction with $1/h(\to\infty)$. Bandi and Phillips (2003) and Aït-Sahalia and Park (2016) have considered decompositions with fully continuous-time processes, as in the second one (17). Bandi and Phillips assume that

$$\max_{1\le i\le n}\sup_{s\in[i\Delta,(i+1)\Delta]}\left|Z_s - Z_{i\Delta}\right| = O_{a.s.}(\delta_n) \quad \text{with } \delta_n = \sqrt{\Delta\log\left(1/\Delta\right)}, \tag{18}$$

for continuous diffusion processes (see Section A.3 for related results), and their convergence rates of nonparametric estimators include a term $\sqrt{\Delta\log\left(1/\Delta\right)}/h$, which is assumed to be $o\left(1\right)$.[10] Aït-Sahalia and Park (2016) impose the condition $\Delta/h^4 = o\left(1\right)$ (see their Assumption 4 for the drift estimation), which guarantees the negligibility of discretization biases (relative to smoothing biases). Since the obtained convergence rates of the discretization biases are slower, we need to impose more restrictive conditions on the shrinking rates of h and Δ in the approach with (17) than in that with (16).

For more details, look at

$$\frac{1}{Th}\sum_{i=1}^n \tilde{\kappa}_{n,i}^c = \frac{1}{Th}\sum_{i=1}^n \int_{i\Delta}^{(i+1)\Delta} K'\left(\frac{Z_{s-}-z}{h}+O_{a.s.}\left(\frac{Z_{s-}-Z_{i\Delta}}{h}\right)\right)\frac{Z_{s-}-Z_{i\Delta}}{h}\mu\left(Z_{s-}\right)ds, \tag{19}$$

[10] Bandi and Phillips (2003) imposed the condition that $\bar{L}_Z\left(T,z\right)\sqrt{\Delta\log\left(1/\Delta\right)}/h = o_{a.s.}\left(1\right)$ for their asymptotic distribution result, where $\bar{L}_Z\left(T,z\right)$ is the chronological local time of $\{Z_s\}$. This condition corresponds to $T\sqrt{\Delta\log\left(1/\Delta\right)}/h = o\left(1\right)$ in ergodic cases like ours.

where the equality follows from the Taylor expansion under the differentiability of K. Given (19), if (18) were assumed with $\sqrt{\Delta \log(1/\Delta)}/h \to 0$, we could obtain $(1/Th)\sum_{i=1}^{n} \tilde{\kappa}_{n,i}^{c} = O_p(\sqrt{\Delta \log(1/\Delta)}/h)$, since

$$\frac{1}{Th}\sum\nolimits_{i=1}^{n}\int_{i\Delta}^{(i+1)\Delta}\left|K'\left(\frac{Z_{s-}-z}{h}+O_{a.s.}\left(\frac{Z_{s-}-Z_{i\Delta}}{h}\right)\right)\right|\left|\mu\left(Z_{s-}\right)\right|ds=O_p\left(1\right), \tag{20}$$

where we will outline how to prove this result (20) in Appendix A.4. We also note that Aït-Sahalia and Park's (2016) results and derivations (e.g., Lemmas 10-12) may be used to compare the two types of decompositions.

Before concluding this subsection, note that for non-stationary/non-ergodic Markov processes, it is often more convenient to consider decompositions with fully continuous-time processes as (17) (rather than ones as (16)), which allow us to exploit integral forms, as $\int_0^T \tilde{X}_s^c ds$, and use some mathematical devices, such as the local time and occupation time formulae in Bandi and Phillips (2003). Aït-Sahalia and Park (2016), Jeong and Park (2014), and Kim and Park (2015) have made significant contributions to asymptotic theory in this line for null-recurrent Markov diffusion processes. The two approaches outlined here may be seen as complementary to each other.

3.3 Convergence results for the non-parametric drift estimator

We now present the convergence results for the estimator $\hat{\mu}(z)$, which is based on the decomposition in (16):

Theorem 6. *Suppose that (B1)-(B2) hold. Then, for any $\nu \in (0,1/2)$,*

$$\frac{1}{nh}\sum\nolimits_{i=1}^{n}\tilde{X}_{n,i}=O_p(\rho_n^{\mu}) \quad \text{with } \rho_n^{\mu}=\begin{cases} T^{-\beta/(1+\beta)}[T^{\beta/(1+\beta)}h]^{\nu} & \text{if } \beta\in(0,1], \\ T^{-1/2}h^{\nu} & \text{if } \beta>1, \end{cases} \tag{21}$$

as $T \to \infty$ and $\Delta \to 0$ with $n \to \infty$, where the bandwidth h is chosen so that $\rho_n^{\mu} \to 0$.

This is a direct application of Theorem 5. The parameter $\nu \in (0,1/2)$ may be chosen arbitrarily, which corresponds to the existence of any arbitrary order moment of $\tilde{X}_{n,i}$ under Condition B2-ii. For $\beta \in (0,1]$, we can have a shrinking rate arbitrarily close to $T^{-\beta/(1+\beta)}$ (by letting ν very small) if the bandwidth is selected as $h = O(T^{-b})$ for some $b \in (0,1)$, which is likely to be a standard choice (say, $b = 1/5$), balancing the effects of the variance and smoothing bias components, i.e., the term $\tilde{\psi}(z)$ and the second term on the RHS of (16), respectively.

To complete our analysis of $\hat{\mu}(z)$, we also present the convergence results for the other terms. To this end, we consider the following conditions:

B3. $\pi_s(\cdot)$ and $\mu(\cdot)$ are continuously twice differentiable satisfying

$$\sup_{s\geq 0}\int_{\mathbb{R}}\left|\pi_s''(z)\right|dz<\infty \quad \text{and} \quad \sup_{s\geq 0}\int_{\mathbb{R}}\left|l_s''(z)\right|dz<\infty,$$

where $l_s(z) := \mu(z)\pi_s(z)$.

B4. $\{Z_s\}$ and $\mu(\cdot)$ satisfy either of the following conditions:

$$\max_{1\le i\le n}\sup_{s\in(i\Delta,(i+1)\Delta]}|\mu(Z_{s-})-\mu(Z_{i\Delta})| = O_p(\delta_n) \quad \text{or} \tag{22}$$

$$\max_{1\le i\le n}\sup_{s\in(i\Delta,(i+1)\Delta]}\|\mu(Z_{s-})-\mu(Z_{i\Delta})\|_{1+\epsilon} = O(\delta_n) \quad \text{for some } \epsilon>0, \tag{23}$$

where $\{\delta_n\}$ is some positive sequence tending to 0 as $n\to\infty$, and $\|\cdot\|_{1+\epsilon} := \{E[|\cdot|^{1+\epsilon}]\}^{1/(1+\epsilon)}$.

B5. $\{U_s\}$ is described by

$$dU_t = \sigma_{s-}dW_s + \int_{\mathbb{R}\setminus\{0\}}\phi^U_{s-}(x)(J_U - v_U)(ds,dx), \tag{24}$$

where $\{W_s\}$ is a standard Brownian motion; $\{\sigma_s\}$ is an adapted càdlàg process; J_U is a Poisson random measure with intensity measure $\nu_U(ds,dx) = dsF_U(dx)$; F_U is a σ-finite measure; $\phi^U_s(x)(\omega)$ is a map on $\Omega\times[0,\infty)\times\mathbb{R}$ into $\mathbb{R}$ that is $\mathfrak{F}_s\times\mathcal{B}(\mathbb{R})$-measurable for all s and càdlàg in s; and $\int_0^t\int_{\mathbb{R}\setminus\{0\}}|\phi^U_{s-}(x)|^2F_U(dx)\,ds<\infty$ almost surely for any $t>0$. For some $\varepsilon>0$, it holds that

$$\sup_{s\ge 0}||\sigma_s^2 + \int_{\mathbb{R}\setminus\{0\}}\left|\phi^U_s(x)\right|^2F_U(dx)||_{1+\varepsilon}<\infty.$$

Condition B3 is quite standard. The condition (22) in B4 holds with $\delta_n = \Delta^\gamma$ for any $\gamma\in(0,d/c)\subset(0,1/2)$ if $\{Z_s\}$ satisfies

$$E[|Z_s - Z_t|^c]\le C|s-t|^{1+d}, \tag{25}$$

for some positive constants C, c, and d (independent of s and t). If $\{Z_s\}$ is a continuous diffusion process, such as

$$dZ_s = a_Z(Z_s)\,ds + \theta_Z(Z_s)\,dB_s, \tag{26}$$

where $\{B_s\}$ is a standard Brownian motion, we may be able to show that

$$\sup\nolimits_{0\le s<t<T;\ |s-t|\le\delta}|Z_s - Z_t| = O_p(\sqrt{\delta\log(1/\delta)}) \ \text{ as } \delta\to 0 \text{ and } T\to\infty,$$

under certain conditions (see discussions on the global modulus of continuity in Kanaya, 2016). We can then let $\delta_n = \sqrt{\Delta\log(1/\Delta)}$ in (22), which may be satisfied by Example 3. We provide some more discussions on the two cases (25) and (26) in Appendix A.3.

We can show that the condition (23) in B4 holds with $\delta_n = O\left(\Delta^{1/2}\right)$ if $\{Z_s\}$ is driven by a semimartingale of the following type:

$$dZ_s = a_s ds + \theta_{s-}dB_s + \int_{\mathbb{R}\setminus\{0\}}\phi^Z_{s-}(x)(J_Z - v_Z)(ds,dx), \tag{27}$$

and if $\|\mu'(Z_s)a_s\|_q$, $\|\mu''(Z_s)\theta_s^2\|_q$, $\|\mu'(Z_s)\theta_s\|_2$, $E[\int_{\mathbb{R}\setminus\{0\}}|\mu\left(Z_s+\phi^Z_s(x)\right)-\mu(Z_s)|^2F_Z(dx)]$, and $||\mu'(Z_s)\int_{\mathbb{R}\setminus\{0\}}\phi^Z_s(x)F_Z(dx)||_q$ are bounded uniformly over $s\ge 0$ for some $q\in(1,2]$.[11] A specification similar to (27) can be found in Todorov (2011), where similar moment conditions are also imposed.

[11] $\{B_s\}$ is a standard Brownian motion; $\{a_s\}$ is locally bounded and predictable; $\{\theta_s\}$ is adapted and càdlàg; and the other components of the last term on the RHS of (27) are defined analogously to those in (24).

The specification (24) of $\{U_s\}$ in Condition B5 is quite general, including processes described by Lévy-type stochastic differential equations, and it covers almost all models used in the economics/finance literature. While we assume for simplicity that both the Brownian and Poisson components are univariate, they may be multivariate. Finally, the moment conditions imposed in B4 and/or B5 may be too strong for some classes of processes. Even in such cases, we may be able to derive the convergence rate of the estimators by using the damping function approach considered in Kanaya (2016).

Theorem 7. *(i) Under Conditions B1 and B3,*

$$\frac{1}{nh}\sum\nolimits_{i=1}^{n} E\left[K\left(\frac{Z_{i\Delta}-z}{h}\right)\left[\mu\left(Z_{i\Delta}\right)-\mu\left(z\right)\right]\right]=O\left(h^{2}\right),\ \textit{as } h\to 0 \textit{ (with } n\to\infty\textit{)}.$$

(ii) Under Conditions B1, B2-ii, and B4, $(1/Th)\sum_{i=1}^{n}\tilde{\kappa}_{ni}=O_p\left(\delta_n\right)$ *as* $T\to\infty$ *and* $h,\Delta\to 0$ *(with* $n\to\infty$*).*
(iii) Under Conditions B1, B2-ii, and B5, $\tilde{\psi}\left(z\right)=O_p(1/\sqrt{Th})$ *as* $T\to\infty$ *and* $h,\Delta\to 0$ *(with* $n\to\infty$*).*

These convergence results do not rely on the mixing condition, where we note that the last one (iii) is derived by the martingale condition B5 and that the regressand process $\{Y_s\}$ is not necessarily mixing.

We can show that $\hat{\pi}\left(z\right)$ is $O_p\left(1\right)$ in the same way as in Theorem 6 under Conditions B1-B2. Therefore, given the results of Theorems 6-7, we have

$$\hat{\mu}\left(z\right)-\mu\left(z\right)=O_p(\rho_n^{\mu})+O_p(\delta_n)+O(h^2)+O_p(1/\sqrt{Th}). \tag{28}$$

It is often assumed that the discretization bias is negligible (relative to the last term on the RHS) by setting $\delta_n\sqrt{Th}\to 0$, which typically corresponds to "$\Delta^{\gamma}\sqrt{Th}\to 0$" or "$\sqrt{\Delta\log\left(1/\Delta\right)Th}\to 0$" (see the previous arguments on Condition B4).

Bandwidth rate. According to the rate result (21) of Theorem 6, if $\beta>1$, the term $O_p(\rho_n^{\mu})$ in (28) is negligible, relative to $O_p(1/\sqrt{Th})$ (always, with any choice of the bandwidth $h\to 0$), since $\rho_n^{\mu}\sqrt{Th}=h^{\nu+1/2}$. Accordingly, the asymptotic distribution of $\hat{\mu}\left(z\right)$ is determined by $\tilde{\psi}\left(z\right)$, the sum of the martingale differences (the last term on the RHS of (28)). However, when $\beta\in(0,1]$, the choice of the bandwidth may matter to the relative magnitude of $O_p(\rho_n^{\mu})$ to $O_p(1/\sqrt{Th})$. For example, if we let $h=O(T^{-1/5})(\neq o(T^{-1/5}))$, then

$$\rho_n^{\mu}\sqrt{Th}=T^{[1-(\beta-2\nu)]/2(1+\beta)}h^{(1+2\nu)/2}=O(T^{(2-\nu)/5-\beta(1-\nu)/(1+\beta)}), \tag{29}$$

We note that $\dfrac{2-\nu}{5}-\dfrac{\beta\left(1-\nu\right)}{1+\beta}\geq 0\Leftrightarrow\dfrac{2-\nu}{3-4\nu}\geq\beta$ and also

$$\sup\nolimits_{\nu\in(0,1/2)}\frac{2-\nu}{3-4\nu}=3/2,$$

which exceeds 1. Therefore, given this (standard) bandwidth $h=O(T^{-1/5})$, the order of ρ_n^{μ} is greater than $O_p(1/\sqrt{Th})$ for any $\beta\in(0,1]$. This situation may arise for the process in Example 3 if the mean-reversion effect of the process is weak (in particular, when the parameter r in (14) is equal to or less than $3/2$).

If $\beta \in (0,1]$ and we want to have the negligibility of ρ_n^μ (relative to $1/\sqrt{Th}$, in order to obtain the asymptotic normality), then we must have a faster shrinking rate of h. That is, if we consider the form of $h = O(T^{-b})$, $b \in (0,1)$ must be large with $\beta > \dfrac{(1-b)(1+2\nu)}{b(1+2\nu)+1}$ or equivalently $b > \dfrac{(1-\beta+2\nu)}{(1+2\nu)(1+\beta)}$ for some $\nu \in (0,1/2)$, which can be obtained by plugging $h = O\left(T^{-b}\right)$ into (29), implying that b must satisfy at least

$$b > \inf_{\nu\in(0,1/2)} \frac{1-\beta+2\nu}{1+2\nu+\beta} = \frac{1-\beta}{1+\beta}. \tag{30}$$

For β close to 0, one has to have b close to 1. If one fails to select b satisfying (30), the asymptotic property of $\hat{\mu}(z)$ may be determined by $(1/nh)\sum_{i=1}^n \tilde{X}_{n,i}$ and the asymptotic normality may not necessarily hold. It is uncertain whether one can construct some reasonable inference procedure in this case while only the consistency is guaranteed by Theorems 6-7, as long as $h = O(T^{-b})(\neq o(T^{-b}))$ for any $b \in (0,1)$.

4 Conclusion

We have derived the convergence rates of the sums of α-mixing arrays for the cases where their mixing coefficients decay slowly and/or the higher-order moments do not exist. Our results may fill the gap between the CLT and LLN under the mixing environment and complement previous convergence rate results for weakly dependent time series (e.g., mixingale LLNs, as in Andrews, 1988; Davidson and de Jong, 1997; de Jong, 1998). We have also showed that our techniques, based on the Bernstein-type inequality, may be applied to a near-unit-root process. The fact that such processes may also be analyzed in the mixing framework does not seem to have attracted attention in the literature.

This paper may also contribute to the high-frequency econometrics literature, where double (infill and long-span) asymptotics are often necessary to estimate economically interesting objects (such as risk-preference parameters) in the continuous-time framework. Limit theorems used in this literature usually rely on a Markov or (semi) martingale type assumption. For the Markov diffusion case, Aït-Sahalia and Park (2016), Jeong and Park (2014), and Kim and Park (2015) have recently provided extensive studies of limit theorems under double asymptotics. This paper's results may complement their theorems, as the former are based on the mixing assumption and are applicable to non-Markov processes (as well as non semi-martingales).

References

[1] Aït-Sahalia, Y. and J.Y. Park (2016) Bandwidth selection and asymptotic properties of local nonparametric estimators in possibly nonstationary continuous-time models, *Journal of Econometrics* 192, 119-138.

[2] Andrews, D.W.K. (1988) Law of large numbers for dependent non-identically distributed random variables, *Econometric Theory* 4, 458-467.

[3] Bandi, M.F. and P.C.B. Phillips (2003) Fully nonparametric estimation of scalar diffusion models, *Econometrica* 71, 241-283.

[4] Bansal, R. and A. Yaron (2004) Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 58, 1481-1509.

[5] Barndorff-Nielsen, O.E. and N. Shephard (2002) Econometric analysis of realised volatility and its use in estimating stochastic volatility models, *Journal of the Royal Statistical Society*, Series B, 63, 253-280.

[6] Borkovec, N.H. and C. Klüpperlberg (1998) Extremal behaviour of diffusion models in finance, *Extremes* 1, 47-80.

[7] Bosq, D (1998) *Nonparametric Statistics for Stochastic Processes*, 2nd ed., Springer-Verlag, New York.

[8] Bosq, D., F. Merlevède, and M. Peligrad (1999) Asymptotic normality for density kernel estimators in discrete and continuous time, *Journal of Multivariate Analysis* 68, 78-95.

[9] Bradley, R.C. (1985) On a central limit question under absolute regularity, *Annals of Probability* 13, 1314-1325.

[10] Bradley, R.C. (1988) On a Theorem of Gordin, *Stochastic* 24, 357-392.

[11] Bradley, R.C. (2005) Basic properties of strong mixing conditions. A survey and some open questions, *Probability Surveys* 2, 107-144.

[12] Chen, X., L.P. Hansen, and M. Carrasco (2010) Nonlinearity and temporal dependence, *Journal of Econometrics* 155, 155-169.

[13] Conley, T., L.P. Hansen, E. Luttmer, and J. Scheinkman (1997) Short-term interest rates as subordinated diffusions, *Review of Financial Studies* 10, 525-577.

[14] Dahlhaus, R., and S.S. Rao (2006) Statistical inference for time varying ARCH processes, *Annals of Statistics* 34, 1075-1114.

[15] Davidson, J. (1993) An L_1-convergence theorem for heterogenous mixingale arrays with trending moments, *Statistics and Probability Letters* 16, 301-304.

[16] Davidson, J. (1994) *Stochastic Limit Theory*, Oxford University Press, Oxford.

[17] Davidson, J. and R.M. de Jong (1997) Strong laws of large numbers for dependent heterogenous processes: a synthesis of recent and new results, *Econometric Reviews* 16-3, 251-279.

[18] de Jong, R.M. (1995) Laws of large numbers for dependent heterogeneous processes, *Econometric Theory* 11, 347-358.

[19] de Jong, R.M. (1996) A strong law of large numbers for triangular mixingale arrays, *Statistics and Probability Letters* 27, 1-9.

[20] de Jong, R.M. (1998) Weak laws of large numbers for dependent random variables, *Annales D'Economie et de Statistique,* 51, 209-225.

[21] Elliott, G., T.J. Rothenberg, and J.H. Stock (1996) Efficient tests fro an autoregressive unit root, *Econometrica* 64, 813-836.

[22] Fan, J. and Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer-Verlag, New York.

[23] Florens-Zmirou, D. (1989) Approximate discrete-time scheme for statistics of diffusion processes, *Statistics* 20, 547-557.

[24] Hansen, B.E. (1991) Strong laws for dependent heterogeneous processes, *Econometric Theory* 7, 213-221.

[25] Hansen, B.E. (1992) Strong laws for dependent heterogeneous processes, Erratum, *Econometric Theory* 8, 421-422.

[26] Ibragimov, I.A. and Y.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff, Groningen.

[27] Jeong, M. and J.Y. Park (2014) Asymptotic theory of maximum likelihood estimator for diffusion model, Working Paper, Indiana University.

[28] Kanaya, S. (2016) Uniform convergence of kernel-based nonparametric estimators for continuous and discrete time processes: A damping function approach, forthcoming in *Econometric Theory* (first published online: 14 July 2016, doi: 10.1017/S0266466616000219).

[29] Kanaya, S. and D. Kristensen (2015) Estimation of stochastic volatility models by nonparametric filtering, forthcoming in *Econometric Theory* (first published online 12 April 2015, doi: 10.1017/S0266466615000079).

[30] Karatzas, I. and S.E. Shereve (1991) *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

[31] Kim, J. and J.Y. Park (2016) Asymptotics for recurrent diffusions with application to high frequency regression, forthcoming in *Journal of Econometrics*.

[32] Kristensen, D. (2010) Nonparametric filtering of the realized spot volatility: A kernel-based approach, *Econometric Theory* 26, 60-93.

[33] Liebscher, E. (1996) Strong convergence of sums of α-mixing random variables with applications to density estimation, *Stochastic Processes and their Applications* 65, 69-80.

[34] Louhichi, S. (2000) Convergence rates in the strong law for associated random variables, *Probability and Mathematical Statistics* 20, 203-214.

[35] Louhichi, S. and P. Soulier (2000) Marcinkiewicz-Zygmund strong laws for infinite variance time series, *Statistical Inference for Stochastic Processes* 3, 31-40.

[36] McLeish, D. L. (1975) A maximal inequality and dependent strong laws, *Annals of Probability* 3, 829-839.

[37] Merlevède, F., M. Peligrad, and S. Utev (2006) Recent advances in invariance principles for stationary sequences, *Probability Surveys* 3, 1-36.

[38] Nicolau, J. (2005) Processes with volatility-induced stationarity: an application for interest rates, *Statistica Neerlandica* 59, 376-396.

[39] Park, J.Y. (2009) Inference on conditional mean models in continuous time, Working Paper, Indiana University.

[40] Phillips, P.C.B (1987) Time series regression with a unit root, *Econometrica* 55, 277-301.

[41] Phillips, P.C.B. and T. Magdalinos (2007) Limit theory for moderate deviations from a unit root, *Journal of Econometrics* 136, 115-130.

[42] Shao, Q.M. (1993) Complete convergence for α-mixing sequences, *Stochastic and Probability Letters* 16, 279-287.

[43] Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* 52, 1973-2002.

[44] Stock, J.H. (1991) Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series, *Journal of Monetary Economics* 28, 435-459.

[45] Todorov, V. (2011) Econometric analysis of jump-driven stochastic volatility models, *Journal of Econometrics* 160, 12-21.

[46] Veretennikov, A.Y. (1997) On polynomial mixing bounds for stochastic differential equations, *Stochastic Processes and their Applications* 70, 115-127.

[47] Veretennikov, A.Y. (1999) On polynomial mixing and convergence rate for stochastic difference and differential equations, *Theory of Probability and its Applications* 44, 361-374.

A Appendix

A.1 Proofs of primary convergence results

Proofs of Theorems 1-5. We follow similar steps in the proofs of Theorems 1-5. Here, we outline points that are common among them. Further details tailored to each theorem are provided subsequently. First, we split the sum of $\{X_{n,i}\}$ into two parts:

$$\textstyle\sum_{i=1}^{n} X_{n,i} = \sum_{i=1}^{n} Z_{n,i} + \sum_{i=1}^{n} \tilde{Z}_{n,i}, \tag{31}$$

where

$$\begin{aligned} Z_{n,i} &:= X_{n,i} 1\left(|X_{n,i}| \leq b_n\right) - E\left[X_{n,i} 1\left(|X_{n,i}| \leq b_n\right)\right]; \\ \tilde{Z}_{n,i} &:= X_{ni} 1\left(|X_{n,i}| > b_n\right) - E\left[X_{n,i} 1\left(|X_{n,i}| > b_n\right)\right]. \end{aligned}$$

We can complete the proof if both terms on the RHS of (31) are shown to be $O_p(\rho_n)$ and the rate of ρ_n is as given in each of Theorems 1-5.

For Theorems 1 and 4, we have $\sum_{i=1}^n \tilde{Z}_{n,i} = O_p(\rho_n)$ by Condition A2 and the assumption that $E[|X_{n,i}|] < \infty$. For Theorem 3, we derive the rate of $\sum_{i=1}^n \tilde{Z}_{n,i}$ (below). For Theorems 2 and 5, we use the Markov inequality: for any $a > 0$,

$$\begin{aligned}\Pr\left(\left|\sum_{i=1}^n \tilde{Z}_{n,i}\right| \geq a\rho_n\right) &\leq a^{-1}\rho_n^{-1}E[|\sum_{i=1}^n \tilde{Z}_{n,i}|] \\ &\leq a^{-1}\rho_n^{-1}2\sum_{i=1}^n E[|X_{n,i}|^p]/b_n^{p-1} \\ &\leq a^{-1}2ng_n^p/\rho_n b_n^{p-1}, \end{aligned} \tag{32}$$

where the second inequality follows from the triangle inequality and the last inequality follows from the definition of the uniform moment bound in (4). The majorant side of (32) can be made arbitrarily small for sufficiently large a; therefore, we obtain $\sum_{i=1}^n \tilde{Z}_{n,i} = O_p(\rho_n)$ if $ng_n^p/b_n^{p-1} = O(\rho_n)$. We subsequently show that $ng_n^p/b_n^{p-1} = O(\rho_n)$ with ρ_n specified in (6) or (13).

To derive the rate of $\sum_{i=1}^n Z_{n,i}$, we use the Bernstein-type inequality in the proofs of all five theorems. We define the covariance process of $\{Z_{n,i}\}$ as

$$D(n,m) := \sup_{0\leq j\leq n-1} E[(\sum_{i=j+1}^{\min\{j+m,m\}} Z_{n,i})^2] \quad (\text{for } m = 1,\ldots,n). \tag{33}$$

Note that $E[Z_{n,i}] = 0$, $|Z_{n,i}| \leq 2b_n$, and $\{Z_{n,i}\}$ has the same mixing coefficient as $\{X_{n,i}\}$. Then, by Theorem 2.1 of Liebscher (1996), for any arbitrary $a > 0$ and each positive integer m satisfying

$$m \leq n \quad \text{and} \quad 4m(2b_n) < a\rho_n, \tag{34}$$

it holds that

$$\Pr(|\sum_{i=1}^n Z_{n,i}| \geq a\rho_n) \leq 4\exp\left\{-\frac{(a\rho_n)^2}{64nD(n,m)/m + (8/3)(a\rho_n)m(2b_n)}\right\} + 4\frac{n}{m}\alpha_n(m). \tag{35}$$

We derive the bound of $D(n,m)$ and choose an appropriate pair of m and b_n in the proof of each theorem. We also illustrate that for a large enough, the majorant side can be arbitrarily small (as $n \to \infty$), which means $\sum_{i=1}^n Z_{n,i} = O_p(\rho_n)$.

Proof of Theorem 1. Given the previous arguments, we first derive the bound of $D(n,m)$ (whose proof is provided in Section A.4):

Lemma 1. *There exists some constant $\omega\,(> 0)$ such that*

$$D(n,m) \leq \begin{cases} \omega b_n^2 m^{2-\beta} & \text{for } \beta \in (0,1), \\ \omega b_n^2 m \log m & \text{for } \beta \geq 1, \\ \omega b_n^2 m & \text{for } \beta > 1. \end{cases} \tag{36}$$

Now, suppose that $\beta \in (0,1)$. Letting $\mathrm{int}[x]$ denote the integer part of x, we set $m = \mathrm{int}[\sqrt{a}n^{1/(1+\beta)}]$ and $b_n = \rho_n n^{-1/(1+\beta)}$, both of which satisfy the conditions in (34) for any $a\,(> 0)$ if n is large enough. Then, by (1) with $D(n,m) \leq \omega b_n^2 m^{2-\beta}$ in (36), the RHS of (35) is bounded by

$$4\exp\left\{-\frac{a^2}{64\omega a^{(1-\beta)} + (16/3)a^{3/2}}\right\} + 4An/\mathrm{int}\left[\sqrt{a}n\right],$$

and we can let $\Pr\left(\left|\sum_{i=1}^{n} Z_{n,i}\right| \geq a\rho_n\right)$ be arbitrarily small for a large enough. This implies the desired result for $\beta \in (0,1)$. If $\beta = 1$ (resp. $\beta > 1$), we set $b_n = \rho_n n^{-1}\sqrt{\log m}$ (resp. $b_n = \rho_n n^{-1/2}$). This b_n, together with $m = \mathrm{int}\left[\sqrt{a} n^{1/(1+\beta)}\right]$, satisfies (34) for any n large enough. Then, given the corresponding bound of $D(n,m)$ in (36), we can show that $\sum_{i=1}^{n} Z_{n,i} = O_p(\rho_n)$ by the same argument. The proof is now complete. □

Proof of Theorem 2. First, consider the case where $\beta(p-2)/p \in (0,1)$. Given any arbitrary $a\,(>0)$, we set $m = \mathrm{int}\left[\sqrt{a} n^{1/(1+\beta)}\right]$ and $b_n = \rho_n n^{-1/(1+\beta)}$ (these m and b_n satisfy (34) for n large enough). Then, we apply (35) to $\sum_{i=1}^{n} Z_{n,i}$ with the following covariance bound:

$$D(n,m) \leq \omega g_n^2 m^{2-\beta(p-2)/p}, \tag{37}$$

where $p > 2$; $\beta(p-2)/p \in (0,1)$; $\omega\,(>0)$ is some constant; and g_n is defined in (4). Given ρ_n in (6), we can show that $\Pr(|\sum_{i=1}^{n} Z_{ni}| \geq a\rho_n) \to 0$ as $a \to \infty$, as in the proof of Theorem 1. This, together with (32), implies that

$$\textstyle\sum_{i=1}^{n} X_{n,i} = O_p(\rho_n) + O_p(n g_n^p / b_n^{p-1}).$$

Therefore, we can complete the proof if we show that $n g_n^p / b_n^{p-1} = O(\rho_n)$. Since $b_n = \rho_n n^{-1/(1+\beta)}$ and $\rho_n = n^{(p+\beta)/p(1+\beta)} g_n$, it holds that $n g_n^p / b_n^{p-1} = O\left(n^{(p+\beta)/p(1+\beta)} g_n\right) = O(\rho_n)$. Now, we have obtained the desired result.

For the cases where $\beta(p-2)/p = 1$ and > 1, we compute the L_2-bound of $\sum_{i=1}^{n} X_{n,i}$ by using Davydov's inequality, which allows us to derive the same convergence rate as when using the Bernstein-type inequality but with a stronger notion of the L_2-convergence (rather than the convergence in probability). By Davydov's inequality (Corollary 1.1 of Bosq, 1998), we have

$$\begin{aligned}\gamma_n(l) &:= \sup_{1\leq k\leq n-l} \left|\mathrm{Cov}\left(X_{n,k}, X_{n,(k+l)}\right)\right| \\ &\leq 4p(2Al^{-\beta})^{(p-2)/p} g_n^2 = O(l^{-\beta(p-2)/p} g_n^2),\end{aligned} \tag{38}$$

uniformly over l. Then,

$$\begin{aligned}E[|\textstyle\sum_{i=1}^{n} X_{n,i}|^2] &\leq \textstyle\sum_{i=1}^{n} E[X_{n,i}^2] + 2n\sum_{l=1}^{n-1}\gamma_n(l) \\ &\leq n g_n^2 + O\left(n g_n^2\right) \times \textstyle\sum_{l=1}^{n-1} l^{-\beta(p-2)/p}.\end{aligned} \tag{39}$$

The bound of $\sum_{l=1}^{n-1} l^{-\beta(p-2)/p}$ can be computed as

$$\textstyle\sum_{l=1}^{n-1} l^{-\beta(p-2)/p} \leq \begin{cases} 1 + \int_1^n x^{-1}dx = 1 + \log n & \text{for } \beta = p/(p-2), \\ 1 + \int_1^\infty x^{-\beta(p-2)/p}dx \leq 1 + \frac{1}{\beta(p-2)/p-1} & \text{for } \beta > p/(p-2). \end{cases} \tag{40}$$

This, together with (39), implies

$$E[|\textstyle\sum_{i=1}^{n} X_{n,i}|^2] = \begin{cases} n g_n^2 + O\left(n g_n^2\right) \times [1 + \log n] = O\left(n(\log n) g_n^2\right) & \text{for } \beta = p/(p-2), \\ n g_n^2 + O\left(n g_n^2\right) \times [1 + \frac{1}{\beta(p-2)/p-1}] = O\left(n g_n^2\right) & \text{for } \beta > p/(p-2). \end{cases}$$

Therefore, we have shown the desired result: $E[|\sum_{i=1}^{n} X_{n,i}|^2] = O_p(\sqrt{n\log n}\, g_n)$ for $\beta = p/(p-2)$ and $= O_p(\sqrt{n} g_n)$ for $\beta > p/(p-2)$, completing the proof. □

Proof of Theorem 3. Since the mixing coefficient satisfies $\alpha_n(m) \leq O(\exp\{\delta(1+\delta)^{-1} m \log \theta_n\})$ as in (10), we can find some $\omega \in (0, \infty)$, such that

$$\begin{aligned} D_{n,m} &= O(\omega m b_n^2 [1 + \tfrac{1}{\log \theta_n} \exp\{\delta(1+\delta)^{-1} m \log \theta_n\}]) \\ &\leq \omega m b_n^2 \left[1 + k_n \exp\{-\tilde{c} m / k_n\}\right], \end{aligned}$$

where $\tilde{c} := \delta(1+\delta)^{-1}\bar{c}$; the first equality follows from arguments similar to those for (38)-(40); and the last inequality holds since $\frac{1}{\log \theta_n} = O(k_n)$ and $\log \theta_n \leq -\bar{c} k_n$. By the proof of Lemma 3.1 of Phillips and Magdalinos (2007), we have $\sum_{i=1}^n X_{n,i}^2 = O_p(n k_n)$. Now, we set $b_n = \rho_n / \sqrt{n k_n}$ and $\rho_n = \sqrt{n k_n}$, which together with $E[|X_{n,i}|] < \infty$ (for each i) imply that $\sum_{i=1}^n \tilde{Z}_{n,i} = O_p(\rho_n)$.

If $k_n = o(n)$, we let $m = \sqrt{a n k_n}$ (which satisfies (34) for large a). Then, the RHS of (35) is bounded by

$$4 \exp\left\{-\frac{a^2}{64 + (16/3)\, a^{3/2}}\right\} + O\left(a^{-1/2}\sqrt{n/k_n} \exp\{-\tilde{c} a^{1/2} \sqrt{n/k_n}\}\right),$$

which can be made arbitrarily small for sufficiently large a, as $n \to \infty$. For the case of $k_n = n$, by letting $m = n/(\log n)$ (which also satisfies (34) for any a if n is large enough), the RHS of (35) is bounded by

$$4 \exp\left\{-\frac{a^2}{64 + (16/3)\, a/(\log n)}\right\} + O\left((\log n) \exp\{-\tilde{c} \log n\}\right),$$

which is also arbitrarily small for large a, as $n \to \infty$. Now, we have shown that $\sum_{i=1}^n Z_{n,i} = O_p(\rho_n)$, completing the proof. □

Proof of Theorem 4. We follow the same strategy as in the proof of Theorem 1. Thus, we omit details and outline only the main points. For the Bernstein-type inequality (35), we set $m = \mathrm{int}[\sqrt{a} n^{1/(1+\beta)} \Delta^{-\beta/(1+\beta)}]$ (for each a; this m satisfies the first conditions in (34) as $T \to \infty$) throughout this proof. Furthermore, we use the following covariance bound:

Lemma 2. *There exists some constant $\omega\,(> 0)$ such that*

$$D(n,m) \leq \begin{cases} \omega b_n^2 \Delta^{-\beta} m^{2-\beta} & \text{for } \beta \in (0,1), \\ \omega b_n^2 \Delta^{-1} m \log m & \text{for } \beta \geq 1, \\ \omega b_n^2 \Delta^{-1} m & \text{for } \beta > 1. \end{cases} \tag{41}$$

Given this bound for each case, we set b_n/ρ_n to satisfy the rate given in (11). Then the second condition in (34) is satisfied for any a large enough (for example, if we let $b_n = \rho_n T^{-1/(1+\beta)} \Delta$ for $\beta \in (0,1)$ then "$a \geq 64$" is enough). Note that we use "$\log m = O(\log n)$" for $\beta \geq 1$, since $m = \mathrm{int}\left[\sqrt{a} n^{1/(1+\beta)} \Delta^{-\beta/(1+\beta)}\right]$ and $\Delta^{-1} = n/T = O(n)$. If Condition A3 is supposed, then we can write $\log m = O(\log T)$. Then, by the same arguments as in the proof of Theorem 1, we can show that $\sum_{i=1}^n Z_{ni} = O_p(\rho_n)$, completing the proof. □

Proof of Theorem 5. For the case where $\beta < p/(p-2)$, we apply (35) to $\sum_{i=1}^n Z_{n,i}$. We let $m = \mathrm{int}\left[\sqrt{a} n^{1/(1+\beta)} \Delta^{-\beta/(1+\beta)}\right]$ and $b_n = \rho_n n^{-1/(1+\beta)} \Delta^{\beta/(1+\beta)}$ with ρ_n in (13), and we use the covariance bound:

$$D(n,m) \leq \omega g_n^2 \Delta^{-\beta(p-2)/p} m^{2-\beta(p-2)/p}, \tag{42}$$

where $\omega\,(>0)$ is some constant. This (42) follows from the same argument as in (37), whose proof is omitted. Now, we can show that $\sum_{i=1}^{n} Z_{ni} = O_p(\rho_n)$ by the previous argument. Given these ρ_n and b_n, we have $ng_n^p/b_n^{p-1} = O(\rho_n)$, completing the proof for $\beta \in (0, p/(p-2))$.

For the cases where $\beta(p-2)/p = 1$ and > 1, we derive the L_2-bound of $\sum_{i=1}^{n} X_{n,i}$. As in (39), we have

$$E[|\textstyle\sum_{i=1}^{n} X_{n,i}|^2] \leq ng_n^2 + O\left(ng_n^2\right) \times \textstyle\sum_{l=1}^{n-1} (\Delta l)^{-\beta(p-2)/p}. \tag{43}$$

If $\beta(p-2)/p = 1$, we can show that $\sum_{l=1}^{n-1} (\Delta l)^{-\beta(p-2)/p} = O(\Delta^{-1}\log n)$, as we did in (40). Then, together with (43), we obtain

$$||\textstyle\sum_{i=1}^{n} X_{n,i}||_2 = O(\sqrt{n\Delta^{-1}\log n}\, g_n) = O(\sqrt{T\log n}\,\Delta^{-1} g_n),$$

where we note that $\log n = \log(T/\Delta) = O(\log T)$ under Condition A3. If $\beta(p-2)/p > 1$, then

$$\begin{aligned} E[|\textstyle\sum_{i=1}^{n} X_{n,i}|^2] &\leq \textstyle\sum_{i=1}^{n} E[X_{n,i}^2] + 2n[\textstyle\sum_{l=1}^{\phi}\gamma_n(l) + \textstyle\sum_{l=\phi+1}^{n-1}\gamma_n(l)] \\ &\leq ng_n^2 + 2ng_n^2[\phi + O(1) \times \textstyle\sum_{l=\phi+1}^{n-1} (\Delta l)^{-\beta(p-2)/p}], \end{aligned} \tag{44}$$

where $\gamma_n(l)$ is defined in (38). The last inequality holds since $\gamma_n(l) \leq g_n^2$ and $\gamma_n(l) = O(g_n^2 (\Delta l)^{-\beta(p-2)/p})$ uniformly over l (the latter can be derived in the same way as for (38) under $\Delta \to 0$). The second term in the square brackets is bounded as

$$\textstyle\sum_{\phi+1}^{n-1} (\Delta l)^{-\beta(p-2)/p} \leq \Delta^{-\beta(p-2)/p} \int_{\phi}^{\infty} x^{-\beta(p-2)/p} dx = \dfrac{\Delta^{-\beta(p-2)/p}\phi^{1-\beta(p-2)/p}}{\beta(p-2)/p - 1}.$$

Given this, we set $\phi = \mathrm{int}[\Delta^{-1}]$ and obtain

$$E[|\textstyle\sum_{i=1}^{n} X_{n,i}|^2] = O\left(ng_n^2\Delta^{-1}\right) = O\left(Tg_n^2\Delta^{-2}\right)$$

in (44), implying the desired result. □

A.2 Proofs for the nonparametric estimator's convergence

Proof of Theorem 6. We start by computing the uniform moment bound. For any $p \geq 1$,

$$\{E[|\tilde{X}_{n,i}|]\}^{1/p} \leq 2\{\textstyle\int_{-\infty}^{\infty} |K((q-z)/h)[\mu(q) - \mu(z)]|^p \pi_{i\Delta}(q)\, dq\}^{1/p}, \tag{45}$$

by the triangle and Hölder inequalities, where $\pi_{i\Delta}$ is the marginal density of $Z_{i\Delta}$. We also have

$$\begin{aligned} \max_{1\leq i\leq n} E[|\tilde{X}_{n,i}|^p] &= \max_{1\leq i\leq n} \textstyle\int_{-\infty}^{\infty} |K((q-z)/h)[\mu(q) - \mu(z)]|^p \pi_{i\Delta}(q)\, dq \\ &= \max_{1\leq i\leq n} h\textstyle\int_{-\infty}^{\infty} |K(r)|^p |\mu(rh+z) - \mu(z)|^p \pi_{i\Delta}(rh+z)\, dq \\ &= \max_{1\leq i\leq n} h^{p+1}\textstyle\int_{-\infty}^{\infty} |K(r)\, r|^p |\mu'(rh\lambda_{rh,z} + z)|\, \pi_{i\Delta}(rh+z)\, dr \\ &\leq h^{p+1}\textstyle\int_{-\infty}^{\infty} |K(r)\, r|^p dr \times \sup_{|\epsilon|\leq\bar{\epsilon}} |\mu'(\epsilon+z)| \times \sup_{s\geq 0;\ |\epsilon|\leq\bar{\epsilon}} \pi_s(\epsilon+z) \\ &= O\left(h^{p+1}\right), \end{aligned} \tag{46}$$

where the second and third equalities follow from changing variables and the Taylor expansion (with some $\lambda_{rh,z} \in [0,1]$, which may depend on rh and z); the inequality holds for h small enough, since the support of K is assumed to be bounded; and the last equality holds by Condition B2-ii.

Given (45)-(46), we obtain $g_n = \sup_{1\le i\le n} ||\tilde{X}_{n,i}||_p = O\left(h^{1+1/p}\right)$ for any $p > 1$. We then apply Theorem 5 with $p > 2$. For the case when $\beta \le 1$,

$$\begin{aligned}(1/nh)\textstyle\sum_{i=1}^n \tilde{X}_{ni} &= O_p\left((1/nh) \times T^{(p+\beta)/p(1+\beta)}\Delta^{-1} \times h^{1+1/p}\right)\\ &= O_p(T^{-\beta/(1+\beta)}\left[T^{\beta/(1+\beta)}h\right]^{1/p}).\end{aligned}$$

If $\beta > 1$, we can always have $\beta > p/(p-1)$ by setting $p\,(> 2)$ large enough, to obtain

$$(1/nh)\textstyle\sum_{i=1}^n \tilde{X}_{ni} = O_p\left((1/nh) \times T^{1/2}\Delta^{-1} \times h^{1+1/p}\right) = O_p\left(T^{-1/2}h^{1/p}\right).$$

Now, the results of the theorem follow since $1/p \in (0, 1/2)$ for $p > 2$. $\square$

Proof of Theorem 7 (i). The result can be proven by the standard arguments for the kernel method, where we use the Taylor expansion and an argument similar to that for (46) to show the negligibility of the remainder terms. We omit details for brevity. $\square$

Proof of Theorem 7 (ii). If the condition in (22) holds, we have

$$(1/Th)\textstyle\sum_{i=1}^n \tilde{\kappa}_{n,i} \le o_p(1) \times (1/nh)\sum_{i=1}^n \left|K\left(\frac{Z_{i\Delta}-z}{h}\right)\right|.$$

Note that $\{E\left[|K\left((Z_{i\Delta}-z)/h\right)|\right]\} = O(h)$ uniformly over i (for each $p \ge 1$ and each z), which follows from the uniform boundedness of $\pi_s(z)$. Consequently, we obtain the desired result. When (23) holds, we consider the following moment bound:

$$\begin{aligned}&E\left[|\tilde{\kappa}_{n,i}|\right] = \textstyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{u-z}{h}\right) v\tilde{\pi}_{i\Delta,i\Delta+1}(u,v)\,dudv\\
&= h\textstyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(w)\, v\tilde{\pi}_{i\Delta,i\Delta+1}(wh+z,v)\,dwdv\\
&\le h\left\{\textstyle\int_{-\infty}^{\infty}|K(w)|^{(1+\epsilon)/\epsilon}\pi_{i\Delta}(wh+z)\,dw\right\}^{\epsilon/(1+\epsilon)} ||\textstyle\int_{i\Delta}^{(i+1)\Delta}\left[\mu(Z_s)-\mu(Z_{i\Delta})\right]ds||_{1+\epsilon}\\
&\le h\left\{\sup_{s\ge 0; z\in\mathbb{R}}\pi_s(z) \times \bar{K}^{1/\epsilon}\textstyle\int_{-\infty}^{\infty}|K(w)|\,dw\right\}^{\epsilon/(1+\epsilon)}\Delta \max_{1\le i\le n}\sup_{s\in[i\Delta,(i+1)\Delta]}\|\mu(Z_s)-\mu(Z_{i\Delta})\|_{1+\epsilon}\\
&= O(h\Delta\delta_n),\end{aligned} \tag{47}$$

where $\tilde{\pi}_{i\Delta,i\Delta+1}(\cdot,\cdot)$ is the joint density of $Z_{i\Delta}$ and $\int_{i\Delta}^{(i+1)\Delta}[\mu(Z_s)-\mu(Z_{i\Delta})]ds$; $\pi_{i\Delta}(\cdot)$ is the marginal density of $Z_{i\Delta}$; the inequality holds by the Hölder inequality; and the last equality holds uniformly over i. Consequently, (47) leads to the desired result. $\square$

Proof of Theorem 7 (iii). First, look at

$$\begin{aligned}&E\left[K^2\left(\tfrac{Z_{i\Delta}-z}{h}\right)\left[U_{(i+1)\Delta}-U_{i\Delta}\right]^2\right]\\
&= \textstyle\int_{i\Delta}^{(i+1)\Delta} E\left[K^2\left(\frac{Z_{i\Delta}-z}{h}\right)\left[\sigma_{s-}^2 ds + \int_{\mathbb{R}\setminus\{0\}}\left|\phi_{s-}^U(x)\right|^2 F_U(dx)\right]\right] ds\\
&\le h\sup_{1\le i\le n}\left\{\textstyle\int_{-\infty}^{\infty}|K(w)|^{2(1+\varepsilon)/\varepsilon}\pi_{i\Delta}(wh+z)\,dw\right\}^{\varepsilon/(1+\varepsilon)}\Delta\sup_{s\ge 0}||\sigma_s^2 + \textstyle\int_{\mathbb{R}\setminus\{0\}}\left|\phi_s^U(x)\right|^2 F_U(dx)||_{1+\varepsilon}\\
&= O(h\Delta),\end{aligned}$$

uniformly over i, where the first equality holds by the Ito isometry and the Fubini theorem; the inequality holds by arguments similar to those for (47); and the last equality holds by the boundedness conditions in (B1), (B2) and (B5). Then, since $\{U_{(i+1)\Delta} - U_{i\Delta}\}$ is a martingale difference array,

$$\begin{aligned} E[|\tilde{\psi}(z)|^2] &= \left(1/T^2h^2\right)\sum_{i=1}^{n} E\left[K^2\left(\tfrac{Z_{i\Delta}-z}{h}\right)\left[U_{(i+1)\Delta} - U_{i\Delta}\right]^2\right] \\ &= \left(1/T^2h^2\right)\sum_{i=1}^{n} O(h\Delta) = O(1/Th), \end{aligned}$$

as desired. □

A.3 Discussions on eqs. (25) and (26), sufficient conditions for Condition B4

The condition (25) is called the Kolmogorov-Čentsov criterion (e.g., Theorem 2.8 in Chapter 2 of Karatzas and Shreve, 1991). If (25) holds, then there exists a continuous modification $\{\tilde{Z}_s\}$ of $\{Z_s\}$ that is almost surely Hölder continuous with any exponent $\gamma \in (0, d/c)$ and some $\vartheta \in (0, \infty)$ satisfying

$$\Pr\left(\omega \in \Omega \,\middle|\, \exists\bar{\Delta}(\omega) > 0 \;\text{ such that } \sup_{|t-s|\in\left(0,\bar{\Delta}(\omega)\right);\; s,t\in[0,\infty)} \frac{|\tilde{Z}_t(\omega) - \tilde{Z}_s(\omega)|}{|t-s|^{\gamma}} \leq \vartheta \right) = 1, \tag{48}$$

where $\bar{\Delta}$ is some positive-valued random variable. While the classical Kolmogorov-Čentsov theorem is a local result in that T must be fixed ($T = \bar{T} < \infty$), we can extend it to a global result where "$s, t \in [0, T]$" may be replaced with "$s, t \in [0, \infty)$," as in (48) (see arguments/proofs in Kanaya and Kristensen, 2015).

We consider the case where $\{Z_s\}$ follows (26). For example, if $\{Z_s\}$ is the OU process described by

$$dZ_s = \lambda_Z(Z_s - m_Z)\,ds + \bar{\theta}_Z dB_s,$$

where $\lambda_Z, \bar{\theta}_z \in (0, \infty)$ and $m_Z \in \mathbb{R}$, we can verify that

$$\sup\nolimits_{0\leq s<t<T;\; |s-t|\leq\delta} |Z_s - Z_t| = O_p(\delta\sqrt{\log T}) + O_p(\sqrt{\delta\log(1/\delta)}), \tag{49}$$

as $\delta \to 0$ and $T \to \infty$. Therefore, the RHS is $O_p(\sqrt{\delta\log(1/\delta)})$ if $\sqrt{\delta/\log(1/\delta)} = O(\sqrt{\log T})$, which is a mild restriction on δ (cf. Condition A3 with $\delta = \Delta$). More generally, if $\{Z_s\}$ is a general diffusion as in (26), then it holds that

$$\begin{aligned} &\sup\nolimits_{0\leq s<t<T;\; |s-t|\leq\delta} |Z_s - Z_t| \\ &= O_p(\delta\sup\nolimits_{0\leq s<T} |a_Z(Z_s)|) + O_p(\sqrt{\delta\log(1/\delta)}\max\left\{1, \sup\nolimits_{0\leq s<T} |\theta_Z(Z_s)|\right\}), \end{aligned} \tag{50}$$

as $\delta \to 0$ and $T \to \infty$, which is discussed and verified in Kanaya (2016), where the proof of (50) is based on a new result on the global modulus of continuity of Brownian motions (also developed in Kanaya, 2016), as well as so-called time-change arguments.

Given (50), we can check (49) since the OU process satisfies $\max\{1, \sup_{0\leq s<T} |\theta_Z(Z_s)|\} \leq 1 + \bar{\theta}_Z = O(1)$ and $\sup_{0\leq s<T} |a_Z(Z_s)| \leq |\lambda_Z m_z| + O_p(\sup_{0\leq s<T} |Z_s|) = O_p(\sqrt{\log T})$, where the latter follows from the result on extremal processes (see, e.g., Borkovec and Klüpperlberg, 1998; Jeong and Park, 2014, and

references therein). In the same way, (50) also allows us to verify that $\sup_{0\leq s<t<T;\ |s-t|\leq\delta}|Z_s - Z_t| = O_p(\sqrt{\delta \log(1/\delta)})$ for a diffusion process in Example 3 (for example, if its drift function is uniformly bounded on $\mathbb{R}$). Given these continuity results such as (48)-(50), as well as some further Hölder continuity condition of the drift function $\mu(\cdot)$ (or the use of the Ito lemma to a diffusion process (26)), we can check (22) of Condition B4.

A.4 Proofs of auxiliary results

Proof of eq. (8). By a careful investigation of the proof of Theorem 7 of de Jong (1998), we can check that replacing de Jong's condition (19) with

$$a_n^{-2}(\textstyle\sum_{t=1}^{n} c_{nt}^2)\sum_{m=0}^{n-1}\zeta_m = O(1) \tag{51}$$

leads to $\|\sum_{t=1}^{n} X_{nt}\|_2 = O(a_n)$, where ζ_m stands for the mixingale number, and a_n, c_{nt} and X_{nt} are used in the same sense as in de Jong's theorem. Given that $g_n(p) = O(1)$, we can let $c_{nt} = O(1)$. In this case, (51) is reduced to $a_n^{-2} n \sum_{m=0}^{n-1}\zeta_m = O(1)$. Since $\zeta_m \leq \bar{A} m^{-\beta(1/2-1/p)}$, we can derive the bound of $\sum_{m=0}^{n-1}\zeta_m$ in the same way as in the proof of Lemma 1 for the three cases in (8), where we note that $\beta(1/2 - 1/p) < 1 \Leftrightarrow \beta < 2p/(p-2)$. We can then check that the possible rates of a_n are as given on the RHS of (8). □

Proof of Eq. (20). We outline only the main points. Assuming that (18) holds, we have $Z_s = Z_{s-}$. Look at

$$\begin{aligned}
&(1/Th)\textstyle\sum_{i=1}^{n}\int_{i\Delta}^{(i+1)\Delta} \left|K'\left(\frac{Z_s-z}{h} + O_{a.s.}(\frac{Z_s-Z_{i\Delta}}{h})\right)\right| |\mu(Z_s)|\, ds \\
&= (1/Th)\textstyle\sum_{i=1}^{n}\int_{i\Delta}^{(i+1)\Delta} \left|K'\left(\frac{Z_s-z}{h} + O_{a.s.}(\frac{Z_s-Z_{i\Delta}}{h})\right)\right| |\mu(Z_s)|\, \mathbf{1}_{\{|Z_s-Z_{i\Delta}|/h\leq\eta\}} ds \\
&+ (1/Th)\textstyle\sum_{i=1}^{n}\int_{i\Delta}^{(i+1)\Delta} \left|K'\left(\frac{Z_s-z}{h} + O_{a.s.}(\frac{Z_s-Z_{i\Delta}}{h})\right)\right| |\mu(Z_s)|\, \mathbf{1}_{\{|Z_s-Z_{i\Delta}|/h>\eta\}} ds,
\end{aligned} \tag{52}$$

where $\eta\,(>0)$ is some positive constant. The second term on the RHS is negligible. This is because, for each $\omega \in \Omega^*$ with $\Pr(\Omega^*) = 1$, it holds that $\mathbf{1}_{\{|Z_s-Z_{i\Delta}|/h>\eta\}} = 0$ uniformly over $s \in [i\Delta, (i+1)\Delta]$ and $i \in \{1, \ldots, n\}$ for any $\sqrt{\Delta\log(1/\Delta)}/h$ small enough. To find the bound of the first term on the RHS, we look at

$$\left|K'\left(\tfrac{Z_s-z}{h} + O_{a.s.}\left(\tfrac{Z_s-Z_{i\Delta}}{h}\right)\right)\right| \mathbf{1}_{\{|Z_s-Z_{i\Delta}|/h\leq\eta\}} \leq K^*\left(\tfrac{Z_s-z}{h}\right),$$

where K^* is a dominant function such that $\forall|\epsilon| \leq \eta$ (with some $\epsilon > 0$), $|K'(x+\epsilon)| \leq K^*(x)$ for any x, and $\int_{-\infty}^{\infty} K^*(x)\,dx < \infty$. We can find such a dominant function for almost all standard kernels (in particular for kernels with bounded support; see the conditions on the kernel function and the proof of Theorem 1 in Kanaya and Kristensen, 2015). Therefore, the first term on the RHS of (52) is bounded by $\frac{1}{T}\int_0^T \frac{1}{h} K^*\left(\frac{Z_s-z}{h}\right)|\mu(Z_s)|\,ds$, which can be shown to be $O_p(1)$ by a standard argument for the kernel method. □

Proof of Lemma 1. We set $\Delta = 1$ without loss of generality. Let $\gamma_n(l) := \sup_{1\leq k\leq n-l}\left|\mathrm{Cov}\left(Z_{nk}, Z_{n(k+l)}\right)\right|$ for each $l(= 0, 1, \ldots, n-1)$. Since $|Z_{n,i}| \leq 2b_n$ and $\gamma_n(l) \leq \bar{\alpha}(l)|2b_n|^2 \leq 4Ab_n^2 l^{-\beta}$, where the latter follows from Billingsley's inequality (Corollary 1.1 of Bosq, 1998). Then,

$$\begin{aligned} E[(\textstyle\sum_{i=j+1}^{(j+m)\wedge n} Z_{ni})^2] &\leq \textstyle\sum_{i=j+1}^{(j+m)\wedge n} E[Z_{ni}^2] + 2m\textstyle\sum_{l=1}^{m-1}\gamma_n(l) \\ &\leq 4mb_n^2 + 2m\textstyle\sum_{l=1}^{m-1} 4Ab_n^2 l^{-\beta} \\ &= 4mb_n^2\left(1 + 2A\textstyle\sum_{l=1}^{m-1} l^{-\beta}\right). \end{aligned} \tag{53}$$

The proof is completed if we compute the bound of $\sum_{l=1}^{m-1} l^{-\beta}$ for each case:

$$\begin{aligned} \textstyle\sum_{l=1}^{m-1} l^{-\beta} &\leq \textstyle\int_0^m x^{-\beta}dx = \frac{m^{1-\beta}}{1-\beta} \quad \text{for } \beta \in (0,1), \\ \textstyle\sum_{l=1}^{m-1} l^{-\beta} &= 1 + \textstyle\sum_{l=2}^{m-1} l^{-\beta} \\ &\leq \begin{cases} 1 + \int_1^m x^{-1}dx = 1 + \log m & \text{for } \beta = 1, \\ 1 + \int_1^\infty x^{-\beta}dx = 1 + 1/(\beta - 1) & \text{for } \beta > 1. \end{cases} \end{aligned}$$

□

Proof of Inequality (37). This proof proceeds in the same way as that of Lemma 1 for $\beta \in (0,1)$, and the details are omitted. It differs only in using the covariance bound $\gamma_n(l) \leq 4p\left(2Al^{-\beta}\right)^{(p-2)/p} g_n^2$ in (53), where this bound follows from Davydov's inequality (Corollary 1.1 of Bosq, 1998). □

Proof of Lemma 2. We use the same notion as in the proof of Lemma 1. By Billingsley's inequality $\gamma_n(l) \leq \alpha(l)|2b_n|^2 \leq 4Ab_n^2(l\Delta)^{-\beta}$. For $\beta \leq 1$, we consider the following bound:

$$\begin{aligned} E[(\textstyle\sum_{i=j+1}^{(j+m)\wedge n} Z_{ni})^2] &\leq \textstyle\sum_{i=j+1}^{(j+m)\wedge n} E[Z_{ni}^2] + 2m\textstyle\sum_{l=1}^{m-1}\gamma_n(l) \\ &\leq 4mb_n^2\left(1 + 2A\textstyle\sum_{l=1}^{m-1}(l\Delta)^{-\beta}\right). \end{aligned} \tag{54}$$

The bound of $\sum_{l=1}^{m-1} l^{-\beta}$ can be computed as follows:

$$\textstyle\sum_{l=1}^{m-1}(l\Delta)^{-\beta} \leq \begin{cases} \Delta^{-\beta}\int_0^m x^{-\beta}dx \leq \Delta^{-\beta}m^{1-\beta}/(1-\beta) & \text{for } \beta \in (0,1), \\ \Delta^{-1}\left(1 + \int_1^m x^{-1}dx\right) = \Delta^{-1}(1 + \log m) & \text{for } \beta = 1. \end{cases}$$

These, together with (54), imply the desired results for $\beta \leq 1$. If $\beta > 1$, then for any integer $\phi\,(\geq 1)$,

$$\begin{aligned} E[(\textstyle\sum_{i=j+1}^{(j+m)\wedge n} Z_{ni})^2] &\leq \textstyle\sum_{i=j+1}^{(j+m)\wedge n} E[Z_{ni}^2] + 2m\left\{\textstyle\sum_{l=1}^{\phi}\gamma_n(l) + \textstyle\sum_{\phi+1}^{m-1}\gamma_n(l)\right\} \\ &\leq 4mb_n^2 + 8mb_n^2\left\{\phi + A\textstyle\sum_{\phi+1}^{m-1}(l\Delta)^{-\beta}\right\}, \end{aligned} \tag{55}$$

where the second term in the braces is bounded as

$$\begin{aligned} \textstyle\sum_{\phi+1}^{m-1}(l\Delta)^{-\beta} &\leq \Delta^{-\beta}\textstyle\int_\phi^\infty x^{-\beta}dx \\ &\leq \Delta^{-\beta}\phi^{1-\beta}/(\beta-1) \quad (\text{since } \beta > 1). \end{aligned}$$

Given this, we set $\phi = \operatorname{int}[\Delta^{-1}]$ and obtain the upper bound of the RHS of (55) as $O(mb_n^2\Delta^{-1})$, completing the proof. □