REGRESSION ASYMPTOTICS
USING MARTINGALE CONVERGENCE METHODS

By

Rustam Ibragimov and Peter C.B. Phillips

July 2004

COWLES FOUNDATION DISCUSSION PAPER NO. 1473

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

REGRESSION ASYMPTOTICS USING MARTINGALE CONVERGENCE METHODS[1]

Rustam Ibragimov

Department of Economics, Yale University

Peter C. B. Phillips

Cowles Foundation for Research in Economics, Yale University

University of Auckland and University of York

ABSTRACT

Weak convergence of partial sums and multilinear forms in independent random variables and linear processes to stochastic integrals now plays a major role in nonstationary time series and has been central to the development of unit root econometrics. The present paper develops a new and conceptually simple method for obtaining such forms of convergence. The method relies on the fact that the econometric quantities of interest involve discrete time martingales or semimartingales and shows how in the limit these quantities become continuous martingales and semimartingales. The limit theory itself uses very general convergence results for semimartingales that were obtained in work by Jacod and Shiryaev (2003). The theory that is developed here is applicable in a wide range of econometric models and many examples are given.

One notable outcome of the new approach is that it provides a unified treatment of the asymptotics for stationary autoregression and autoregression with roots at or near unity, as both these cases are subsumed within the martingale convergence approach and different rates of convergence are accommodated in a natural way. The approach is also useful in developing asymptotics for certain nonlinear functions of integrated processes, which are now receiving attention in econometric applications, and some new results in this area are presented. The paper is partly of pedagogical interest and the conceptual simplicity of the methods is appealing. Since this is the first time the methods have been used in econometrics, the exposition is presented in some detail with illustrations of new derivations of some well-known existing results, as well as some new asymptotic results and the unification of the limit theory for autoregression.

Key words and phrases: semimartingale, martingale, convergence, stochastic integrals, bilinear forms, multilinear forms, $U-$statistics, unit root, stationarity, Brownian motion, invariance principle, unification.

JEL Classification: C13, C14, C32

July, 2004

[1] Rustam Ibragimov gratefully acknowledges financial support from a Yale University Dissertation Fellowship and a Cowles Foundation Prize. Peter C. B. Phillips thanks the NSF for research support under grant SES 04-142254. Correspondence to: peter.phillips@yale.edu

1. Introduction

Much of the modern literature on asymptotic theory in statistics and econometrics involves the weak convergence of multilinear forms and $U-$statistics in independent random variables, martingale-differences and weakly dependent innovations to stochastic integrals (see, among others, Dynkin and Mandelbaum, 1983, Mandelbaum and Taqqu, 1984, Phillips, 1987a & b, Avram, 1988, and Borodin and I. A. Ibragimov, 1995). In econometrics, the interest in this limit theory is frequently motivated by its many applications in regression asymptotics for processes with autoregressive roots at or near unity (inter alia, see Phillips, 1987a & b, Phillips and Perron, 1988, Park and Phillips, 1988, 1989, Phillips and Magdalinos, 2004, and the references therein). Recent attention (Park and Phillips, 1999, 2001, De Jong, 2002, Jeganathan, 2003a, b, Pöetscher 2004, Saikkonen and Choi, 2004) has also been given to the limit behavior of certain types of nonlinear functions of integrated processes. Results of this type have interesting econometric applications that include transition behavior between regimes and market intervention policy (Hu and Phillips, 2004), where nonlinearities of nonstationary economic time series arise in a natural way.

Traditionally, functional limit theorems for multilinear forms have been derived by using their representation as polynomials in sample moments (via summation by parts arguments or, more generally, Newton polynomials relating sums of powers to the sums of products) and then applying standard weak convergence results for sums of independent or weakly dependent random variables or martingales. Avram (1988), for example, makes extensive use of this approach. Thus, in the case of a martingale-difference sequence (ϵ_t), weak convergence of the partial sum process $n^{-1/2}\sum_{i=1}^{[nr]}\epsilon_t$ to a Brownian motion limit process $W = (W(r), r \geq 0)$ by Donsker's theorem implies that the bilinear form

$$\frac{1}{n}\sum_{t=1}^{[nr]}\Big(\sum_{i=1}^{t-1}\epsilon_i\Big)\epsilon_t \tag{1.1}$$

converges to the stochastic integral $\int_0^r W(v)dW(v)$. This approach has a number of advantages and has been extensively used in econometric work since Phillips (1987a).

The approach also has drawbacks. One is that the approach is problem specific in certain ways. For instance, it cannot be directly used in the case of statistics like $\sum_{t=1}^{n} y_{t-1}u_t$, where $y_t = \alpha_n y_{t-1} + u_t$, $t = 1, ..., n$, and $\alpha_n \to 1$ as $n \to \infty$, that are central to the study of local deviations from a unit root in time series regression. Of course, there are ways of making the usual functional limit theory work (Phillips, 1987b; Chan and Wei, 1987 & 1988) and even extending it to situations where the deviations are moderately distant from unity (Phillips and Magdalinos, 2004). In addition, the method cannot be directly applied in the case of sample covariance functions of random walks and innovations, like $V_n = n^{-1/2}\sum_{t=1}^{n} f\big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}\epsilon_i\big)\epsilon_t$, where f is a certain nonlinear function. Such sample covariances commonly arise in econometric models where nonlinear functions are introduced to smooth transitions from one regime to another (e.g., Saikkonen and Choi, 2004). To deal with such complications, one currently has to appeal to stochastic Taylor expansions and polynomial approximations to V_n.

At a more fundamental level, the standard approach sheds little insight into the underlying nature of limit results such as $n^{-1}\sum_{t=1}^{[nr]}\big(\sum_{i=1}^{t-1}\epsilon_i\big)\epsilon_t \to \int_0^r W(s)dW(s)$ or $\int_0^r W(s)dW(s) + r\lambda$ for some constant λ in the case of weakly dependent ϵ_t. Such results are, in fact, the natural outcome of convergence of a sequence of (semi)martingales to a continuous (semi)martingale. As such, they may be treated directly in this way using powerful methods of reducing the study of semimartingale convergence to the study of convergence of its predictable characteristics. Jacod and Shiryaev (2003, hereafter JS) pioneered developments in stochastic process limit theory along these lines (see also He, Wang and Yan, 1992, hereafter HWY), but the method has so far not been used in the theory of weak convergence to stochastic integrals, nor has it yet been used in econometrics.

The asymptotic results for semimartingales obtained by JS have great generality. However, these results appear to have had little impact so far in statistics and none that we are aware of in econometrics. In part, this may be due to the fact that the book is difficult to read, contains many complex conceptualizations, and has a highly original and demanding notational system. The methods were recently applied by Coffman, Puhalskii and Reiman (1998) to study asymptotic properties of classical polling models that arise in performance studies of computer services. In this interesting paper, Coffman, Puhalskii and Reiman showed, using the JS semimartingale convergence results, that unfinished work in a queuing system under heavy traffic tends to a Bessel type diffusion.

One goal of the present paper is pedagogical - to show how the JS approach may be used to develop quite general asymptotic distribution results in time series econometrics and to provide a unifying principle for studying convergence to limit processes and stochastic integrals by means of semimartingale methods. The main advantage of this treatment is its generality and range of applicability. In particular, the approach unifies the proof of weak convergence of partial sums to Brownian motion with that of the weak convergence of sample covariances to stochastic integrals of Wiener processes. Beyond this, the methods can be used to develop asymptotics for time series regression with roots near unity and to study weak convergence of nonlinear functionals of integrated processes. In all these cases, the limit theory is reduced to a special case of the weak convergence of semimartingales.

For the case of a first order autoregression with martingale-difference errors, we show that an identical construction delivers a central limit theorem in the stationary case and weak convergence to a stochastic integral in the unit root case, thereby effectively unifying the limit theory for autoregressive estimation. In fact, the approach enables a unified treatment of the stationary, explosive, unit root and local to unity cases. In all these cases, normalized versions of the estimation error are represented in martingale form as a ratio $X_n(r)/[X_n]_r^{1/2}$, where $X_n(r)$ is a martingale with quadratic variation $[X_n]_r$, and the limit theory is delivered by martingale convergence in the form $X_n(r)/[X_n]_r^{1/2} \to_d X(r)/[X]_r^{1/2}$, where $X(r)$ is the limiting martingale process. To our knowledge, no other approach to the limit theory is able to accomplish this. As we will show, the martingale approach allows in a natural way for the differences in the rates of convergence that arise in the limit theory for autoregression. In contrast, conventional approaches require separate treatments for the stationary and nonstationary cases, as is very well-known.

In addition, the present paper contributes to the asymptotic theory of stochastic processes and time series in several other ways. First, applications of the general martingale convergence results to statistics considered in this work overcome some technical problems that have existed heretofore in the literature. For instance, the global strong majoration condition in JS that naturally appears in the study of weak convergence to a Brownian motion is not satisfied in the case of weak convergence to stochastic integrals. This failure may explain why the martingale convergence methods of JS have not so far been applied to such problems. The present paper demonstrates how this difficulty can be overcome by means of localized versions of general semimartingale results in JS that involve only a local majoration argument. These new arguments appear in the proofs of Theorems 5.1, 5.2 and 6.1.

Second, we provide general sufficient conditions for the assumptions of JS semimartingale convergence theorems to be satisfied for multivariate diffusion processes, including the case of stochastic integrals considered in this paper (see Section 11 and, in particular, Corollary 11.1). These results provide the key to the analysis of convergence to stochastics integrals and, especially, to the study of the asymptotics of functionals of martingales and linear processes in Theorem 6.1. Third, the general approach developed in this paper can be applied in a number of other fields of statistics and econometrics, where convergence to Gaussian processes and stochastic integrals arise. These areas include, for instance, the study of convergence of general multilinear forms and $U-$statistics to multiple stochastic integrals as well as the analysis of asymptotics for empirical copula processes, both of which are experiencing growing interest in econometric research.

The paper is organized as follows. Section 2 introduces the main definitions and notations used throughout the paper. Section 3 discusses the general JS results for convergence of semimartingales in terms of their predictable characteristics. Section 4 contains applications of the approach to partial sums and sample covariances of independent random variables and linear processes. Sections 5 and 6 present the paper's first group of main results, giving applications of semimartingale limit theorems to weak convergence to stochastic integrals, including some general classes of nonlinear functions of integrated processes. Section 7 applies the results to stationary autoregression and unit root regression. Section 8 provides extensions to multivariate cases, including new proofs of weak convergence to multivariate stochastic integrals. This section gives results on weak convergence of discontinuous martingales (arising from discrete time martingales) to continuous martingales and completes the unification of the limit theory for autoregression. Section 9 provides an explicit unified formulation of the limit theory for first order autoregression including the case of explosive autoregression which can also be handled by martingale methods. Section 10 concludes and mentions some further applications of the new techniques.

Sections 11-13 are appendices that contain definitions and technical results needed for the arguments in the body of the paper. These appendices are intended to provide enough background material to make the body of the paper accessible to econometric readers and to make the paper a self-contained resource for econometricians. In particular, Section 11 presents sufficient conditions for semimartingale convergence theorems to hold in the case where the limit semimartingale is a diffusion or a stochastic integral. Section 12 provides results on Skorohod embedding of martingales into a Brownian motion and rates of convergence that are needed in the asymptotic arguments. Section 13 contains some auxiliary lemmas needed for the proof of the main results.

2. Definitions

Throughout the paper we use standard concepts and definitions from stochastic process theory. To aid the presentation of the results of the paper, we state here some fundamental notions of semimartingale theory. In what follows, the processes are defined on a probability space $(\Omega, \Im, P)$ that is equipped with a filtration $\mathbb{F} = (\Im_s, s \geq 0)$ of sub-$\sigma-$fields of $\Im$. The definitions formulated below are based on the treatment in JS and HWY to make reference to those works more convenient, but they are adapted to the continuous process case that is studied in this paper.

Denote $\mathbf{R}_+ = [0, \infty)$, $\mathbf{N} = \{1, 2, ..., \}$ and $\mathbf{Z} = \{..., -2, -1, 0, 1, 2, ...\}$. Throughout the paper, $I(\cdot)$ stands for the indicator function.

Definition 2.1 *(**Increasing processes,** Definition I.3.1 of JS; Definition III.3.41 of HWY). A real-valued process $X = (X(s), s \geq 0)$ with $X(0) = 0$ called an **increasing** process if all its trajectories are non-negative right-continuous increasing functions.*

Definition 2.2 *(**Processes with finite variation**, Definition I.3.1 and Proposition I.3.3 in JS; Definition III.3.41 in HWY). A real-valued process $X = (X(s), s \geq 0)$ is said to be of finite variation if it is the difference of two increasing processes $Y = (Y(s), s \geq 0)$ and $Z = (Z(s), s \geq 0)$, viz., $X(s) = Y(s) - Z(s)$, $s \geq 0$. The process $Var(X) = (Var(X)(s), s \geq 0)$, where $Var(X)(s) = Y(s) + Z(s)$, $s \geq 0$, is called the variation process of X.*

Definition 2.3 *(**Strong majoration**, Definition VI.3.34 in JS). Let $X = (X(s), s \geq 0)$ and $Y = (Y(s), s \geq 0)$ be two real-valued increasing processes. It is said that X strongly majorizes Y if the process $X-Y = (X(s)-Y(s), s \geq 0)$ is itself increasing.*

Definition 2.4 *(**Semimartingales**, Definition I.4.21 in JS; Definition VIII.8.1 in HWY). An $\mathbf{R}^d-$valued process $X = (X(s), s \geq 0)$, $X(s) = (X^1(s), ..., X^d(s)) \in \mathbf{R}^d_+$, is called a $d-$dimensional semimartingale with respect to $\mathbb{F}$ (or a $d-$dimensional $\mathbb{F}-$semimartingale for short) if, for all $s \geq 0$ and all $j = 1, ..., d$,*

$$X^j(s) = X^j(0) + M^j(s) + B^j(s), \tag{2.2}$$

where $X^j(0)$, $j = 1, ..., d$, are finite-valued and $\Im_0-$measurable random variables, $M^j = (M^j(s), s \geq 0)$, $j = 1, ..., d$, are (real-valued) local $\mathbb{F}-$martingales with $M^j(0) = 0$, $j = 1, ..., d$, and $B^j = (B^j(s), s \geq 0)$, $j = 1, ..., d$, are (real-valued) $\mathbb{F}-$adapted processes with finite variation.

Definition 2.5 *(**Quadratic variation**, Section I.4e in JS; Section VI.4 in HWY). Let $M = (M(s), s \geq 0)$ be a continuous square integrable martingale. The **quadratic variation** of M, denoted $[M, M]$, is the unique continuous process $[M, M] = ([M, M](s), s \geq 0)$, for which $M^2 - [M, M]$ is a uniformly integrable martingale which is null at $s = 0$ (existence and uniqueness of $[M, M]$ holds by Doob-Meyer decomposition theorem, see Theorem V.5.48 and Section VI.4 in HWY).*

3. Predictable characteristics and convergence of continuous semimartingales

Let $X = (X(s), s \geq 0)$, where $X(s) = (X^1(s), ..., X^d(s)) \in \mathbf{R}^d$, be a continuous $d-$dimensional $\mathbb{F}-$semimartingale on $(\Omega, \Im, P)$. Then X admits a unique decomposition (2.2); furthermore, the processes $B^j = (B^j(s), s \geq 0)$, $j = 1, ..., d$, and $M^j = (M^j(s), s \geq 0)$, $j = 1, ..., d$, appearing in (2.2) are continuous (see Lemma I.4.24 in JS).

Definition 3.1 ***(Predictable characteristics of continuous semimartingales**, Definition II.2.6 in JS). The $\mathbf{R}^d-$valued process $B = (B(s), s \geq 0)$, where $B(s) = (B^1(s), ..., B^d(s))$, $s \geq 0$, is called the **first predictable characteristic** of X. The $\mathbf{R}^{d\times d}-$valued process $C = (C(s), s \geq 0)$, where $C(s) = (C^{ij}(s))_{1\leq i,j\leq d} \in \mathbf{R}^{d\times d}$, $C^{ij}(s) = [X^i, X^j](s)$, $s \geq 0$, $i, j = 1, ..., d$, is called the **second predictable characteristic** of X.*

In the terminology of JS (see Section II.2a in JS), $X = (X(s), s \geq 0)$ is a semimartingale with the triplet of predictable characteristics (B, C, ν), where the third predictable characteristic of X (the predictable measure of jumps) is zero in the present case, i.e., $\nu = 0$. Furthermore, since X is continuous, the triplet does not depend on a truncation function.

Definition 3.2 ***(Martingale problem**, Section III.2 in JS). Let $X = (X(s), s \geq 0)$, $X(s) = (X^1(s), ..., X^d(s)) \in \mathbf{R}^d$ be a $d-$dimensional continuous process and let $\mathcal{H}$ denote the $\sigma-$field generated by $X(0)$ and $\mathcal{L}_0$ denote the distribution of $X(0)$. A solution to the **martingale problem** associated with $(\mathcal{H}, X)$ and $(\mathcal{L}_0, B, C, \nu)$, where $\nu = 0$, is a probability measure P on $(\Omega, \Im)$ such that X is a $d-$dimensional $\mathbb{F}-$semimartingale on $(\Omega, \Im, P)$ with the first and second predictable characteristics B and C.*

Throughout the rest of the paper, we assume that $(\Omega, \Im)$ is the Skorohod space $(\mathbb{D}(\mathbf{R}_+^d), \mathcal{D}(\mathbf{R}_+^d))$, where $\mathbf{R}_+ = [0, \infty)$. A limit process $X = (X(s), s \geq 0)$ appearing in the asymptotic results is the canonical process $X(s, \alpha) = \alpha(s)$ for the element $\alpha = (\alpha(s), s \geq 0)$ of $\mathbb{D}(\mathbf{R}_+^d)$ (see Section VI.1 and Hypothesis IX.2.6 in JS) and $\mathbb{F}$ is the filtration generated by X. In what follows, $\rightarrow_d$ denotes convergence in distribution in an appropriate metric space and $\rightarrow_P$ stands for convergence in probability. The symbol $=_d$ means distributional equivalence. For a sequence of random variables ξ_n and constants a_n, we write $\xi_n = O_P(1)$ if the sequence ξ_n is bounded in probability and write $\xi_n = o_{a.s.}(a_n)$ if $\xi_n/a_n \rightarrow_{a.s.} 0$. Further, $W = (W(s), s \geq 0)$ denotes standard (one-dimensional) Brownian motion on $\mathbb{D}(\mathbf{R}_+)$. All processes considered in the paper are assumed to be continuous and locally square integrable, if not stated otherwise. Throughout the paper, K and L denote generic constants, not necessarily taking the same values from one place to another.

Let $X_n = (X_n(s), s \geq 0)$, $X_n(s) = (X_n^1(s), ..., X_n^d(s)) \in \mathbf{R}^d$, $n \geq 1$, be a sequence of $d-$dimensional continuous semimartingales on $(\Omega, \Im, P)$. For $a \geq 0$ and an element $\alpha = (\alpha(s), s \geq 0)$ of the Skorohod space $\mathbb{D}(\mathbf{R}_+^d)$, define, as in IX.3.38 of JS,

$$S^a(\alpha) = \inf(s : |\alpha(s)| \geq a \text{ or } |\alpha(s-)| \geq a),$$
(3.1)
$$S_n^a = \inf(s : |X_n(s)| \geq a),$$

where $\alpha(s-)$ denotes the left-hand limit of α at s. For $r \geq 0$ and $\alpha \in \mathbb{D}(\mathbf{R}_+^d)$, denote

$$\overline{\alpha}_{(r)}(x) = \alpha(x - r), \tag{3.2}$$

$x \in \mathbf{R}^d$. For $r \geq 0$, introduce the processes $\overline{B}_{(r)} = (\overline{B}_{(r)}(s), s \geq 0)$ and $\overline{C}_{(r)} = (\overline{C}_{(r)}(s), s \geq 0)$, where

$$\overline{B}_{(r)}(s, \alpha) = B(s + r, \overline{\alpha}_{(r)}) - B(r, \overline{\alpha}_{(r)}), \tag{3.3}$$

$$\overline{C}_{(r)}(s, \alpha) = C(s + r, \overline{\alpha}_{(r)}) - C(r, \overline{\alpha}_{(r)}), \tag{3.4}$$

$\alpha \in \mathbb{D}(\mathbf{R}_+^d)$, $s \geq 0$.

The following theorem gives sufficient conditions for the weak convergence of a sequence of continuous locally square integrable semimartingales. This theorem, together with Theorem 3.2 below, provides the basis for the study of asymptotic properties of functionals of partial sums in subsequent sections.

Throughout the rest of the section, $B_n = (B_n(s), s \geq 0)$ and $C_n = (C_n(s), s \geq 0)$, where $B_n(s) = (B_n^1(s), ..., B_n^d(s))$ and $C_n(s) = (C_n^{ij}(s))_{1\leq i,j\leq d}$, $s \geq 0$, denote the first and the second predictable characteristics of X_n, respectively.

In what follows in our initial applications of the martingale convergence argument, both X_n and X are continuous. Then, in the corresponding results in JS, the third predictable characteristics of X_n and X are zero (i.e., $\nu_n = \nu = 0$), the first characteristics without truncation of X_n and X are the same as B_n and B (i.e., $B'_n = B_n$, $B' = B$), and the modified characteristics without truncation of X_n and X are the same as C_n and C (i.e., $\tilde{C}\,{:}'_n = C_n$, $\tilde{C}' = C$). The final section of the paper will consider the case where X_n has discontinuities and X is continuous. This extension is particularly valuable in providing a martingale convergence proof of weak convergence of sample covariances to a multivariate stochastic integral.

Theorem 3.1 *(see Theorem IX.3.48, Remark IX.3.40, Theorem III.2.40 and Lemma IX.4.4 in JS and also the proof of Theorem 2.1 in Coffman, Puhalskii and Reiman, 1998). Suppose that the following conditions hold:*

(A1) ***The local strong majoration hypothesis:*** *For all $a \geq 0$, there is an increasing, deterministic function $F(a) = (F(s,a),\ s \geq 0)$ such that the stopped real-valued processes $(\sum_{j=1}^{d} Var(B^j)(s \wedge S_a, \alpha), s \geq 0)$ and $(C^{jj}(s \wedge S_a, \alpha), s \geq 0)$, $j = 1, ..., d$, are strongly majorized by $F(a)$ for all $\alpha \in \mathbb{D}(\mathbf{R}^d_+)$ (see Definitions 2.2 and 2.3).*

(A2) ***Uniqueness hypothesis:*** *Let $\mathcal{H}$ denote the $\sigma-$field generated by $X(0)$ and let $\mathcal{L}_0$ denote the distribution of $X(0)$. For each $z \in \mathbf{R}^d$ and $r \geq 0$, the martingale problem associated with $(\mathcal{H}, X)$ and $(\mathcal{L}_0, \overline{B}_{(r)}, \overline{C}_{(r)}, \nu)$, where $X(0) = z$ a.s. and $\nu = 0$, has a unique solution $P_{z,r}$ (see Definition 3.2).*

(A3) ***Measurability hypothesis:*** *The mapping $(z, r) \in \mathbf{R}^d \times \mathbf{R}_+ \to P_{z,r}(A)$ is Borel for all $A \in \Im$.*

(A4) ***The continuity condition:*** *The mappings $\alpha \to B(s, \alpha)$ and $\alpha \to C(s, \alpha)$ are continuous for the Skorohod topology on $\mathbb{D}(\mathbf{R}^d_+)$ for all $s > 0$.*

(A5) $X_n(0) \to_d X(0)$.

(A6) $[\mathbf{sup} - \boldsymbol{\beta}_{\mathbf{loc}}]$ $\quad \sup_{0<s\leq N} |B_n(s \wedge S^a_n) - B(s \wedge S^a, X_n)| \to_P 0$ *for all* $N \in \mathbf{N}$ *and all* $a > 0$.

$[\boldsymbol{\gamma}_{\mathbf{loc}} - \mathbf{R}_+]$ $\quad C_n(s \wedge S^a_n) - C(s \wedge S^a, X_n) \to_P 0$ *for all* $s > 0$ *and* $a > 0$.

Then $X_n \to_d X$.

A sufficient condition for (A6) is the following:

(A6') $[\mathbf{sup} - \boldsymbol{\beta}]$ $\quad \sup_{0<s\leq N} |B_n(s) - B(s, X_n)| \to_P 0$ *for all* $N \in \mathbf{N}$;

$[\mathbf{sup} - \boldsymbol{\gamma}]$ $\quad \sup_{0<s\leq N} |C_n(s) - C(s, X_n)| \to_P 0$ *for all* $N \in \mathbf{N}$.

In the case when the limit semimartingale X satisfies the condition of global strong majoration (see condition (B1) below), conditions (A2)-(A4) and (A6') of Theorem 3.1 simplify and the following result applies.

Theorem 3.2 *(Theorem IX.3.21 in JS). Suppose that the following conditions hold:*

(B1) ***The global strong majoration hypothesis:*** *There is an increasing, deterministic function $F = (F(s), s \geq 0)$ such that the real-valued processes $(\sum_{j=1}^{d} Var(B^j)(s, \alpha), s \geq 0)$ and $(\sum_{j=1}^{d} C^{jj}(s, \alpha), s \geq 0)$, $j = 1, ..., d$, are strongly majorized by F for all $\alpha \in \mathbb{D}(\mathbf{R}^d_+)$ (see Definitions 2.2 and 2.3).*

(B2) ***Uniqueness hypothesis:*** *Let $\mathcal{H}$ denote the $\sigma-$field generated by $X(0)$ and let $\mathcal{L}_0$ denote the distribution of $X(0)$. The martingale problem associated with $(\mathcal{H}, X)$ and $(\mathcal{L}_0, B, C, \nu)$, where $\nu = 0$, has a unique solution P.*

(B3) ***The continuity condition:*** *The mappings $\alpha \to B(s, \alpha)$ and $\alpha \to C(s, \alpha)$ are continuous for the Skorohod topology on $\mathbb{D}(\mathbf{R}^d_+)$ for all $s > 0$.*

(B4) $X_n(0) \to_d X(0)$.

(B5) $[\mathbf{sup} - \boldsymbol{\beta}]$ $\quad \sup_{0<s\leq N} |B_n(s) - B(s, X_n)| \to_P 0$ *for all* $N \in \mathbf{N}$;

$[\boldsymbol{\gamma} - \mathbf{R}_+]$ $\quad C_n(s) - C(s, X_n) \to_P 0$ *for all* $s > 0$.

Then $X_n \to_d X$.

The essence of Theorems 3.1 and 3.2 is that convergence of a sequence of semimartingales holds if their predictable characteristics and the initial distributions tend to those of the limit semimartingale (conditions (A5), (A6), (A6'), (B4) and (B5)), the predictable characteristics of the limit process grow in a regular way (conditions (A1) and (B1)) and the process is the only continuous semimartingale with characteristics B and C and the given initial distribution (conditions (A2), (A3), (B2)). Technically, conditions (A1), (A5), (A6), (A6') and (B1), (B4) and (B5) guarantee that the sequence (X_n) is tight and, under conditions (A2)-(A4), (A6), (B2), (B3) and (B5), the limit is identified (see Ch. IX in JS).

4. Invariance principles (IP) for partial sums, sample variances and sample covariances

Let $(\epsilon_t)_{t\in\mathbf{Z}}$ be a sequence of random variables and let $(\Im_t)_{t\in Z}$ be a natural filtration for (ϵ_t) (that is, $\Im_t$ is the $\sigma-$field generated by $\{\epsilon_k, k \le t\}$). The following conditions will be convenient at various points in the remainder of the paper.

Assumption D1: *$(\epsilon_t, \Im_t)$ is a martingale-difference sequence with $E(\epsilon_t^2|\Im_{t-1}) = \sigma_\epsilon^2 \in \mathbf{R}_+$ for all t and $\sup_{t\in\mathbf{Z}} E(|\epsilon_t|^p|\Im_{t-1}) < \infty$ a.s. for some $p > 2$.*

Assumption D2: *(ϵ_t) are mean-zero i.i.d. random variables with $E\epsilon_0^2 = \sigma_\epsilon^2 \in \mathbf{R}_+$ and $E|\epsilon_0|^p < \infty$ for some $p > 2$.*

The following theorem illustrates the simplest possible use of martingale convergence machinery in conjunction with the Skorohod embedding (see Appendix A2) in proving martingale limit results. Here, a sequence of discrete time martingales is embedded in a sequence of continuous martingales to which we may apply martingale convergence results for continuous martingales, giving an invariance principle for martingales with non-random conditional variances. As is conventional, the proof requires that the probability space on which the random sequences are defined has been appropriately enlarged so that Lemma 12.1 in Appendix A2 holds. In the proof of the main results of the paper, $(T_k)_{k\ge 0}$ denote the stopping times defined in Lemma 12.1.

Later in the paper in section 8, we show how to use martingale convergence results of discontinuous martingales (semimartingales) to continuous martingales (semimartingales) which avoid the use of the Skorohod embedding. In doing so, these results are particularly useful in multivariate extensions.

Theorem 4.1 ***(IP for martingales).*** *Under assumption (D1),*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} \epsilon_t \to_d \sigma_\epsilon W(r). \tag{4.1}$$

Proof. From Lemma 12.1 it follows that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} \epsilon_t =_d W\Big(\frac{T_{[nr]}}{n}\Big). \tag{4.2}$$

By (12.3) and Lemma 13.3 in the Appendix,

$$T_{[nr]}/n \to_P \sigma_\epsilon^2 r. \tag{4.3}$$

Therefore, from Lemma 13.2 it follows that $W(T_{[nr]}/n) \to_d W(\sigma_\epsilon^2 r)$. This and (4.2) imply (4.1). ■

The following theorem is the analogue of Theorem 4.1 for linear processes.

Theorem 4.2 ***(IP for linear processes).*** *Suppose that* $(u_t)_{t\in\mathbf{N}}$ *is the linear process* $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}$, $C(L) = \sum_{j=0}^{\infty} c_j L^j$, *where* $\sum_{j=1}^{\infty} j|c_j| < \infty$, $C(1) \neq 0$, *and* $(\epsilon_t)_{t\in\mathbf{Z}}$ *satisfy assumption (D1) with* $p \geq 4$. *Then*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t \to_d \omega W(r), \tag{4.4}$$

where $\omega^2 = \sigma_\epsilon^2 C^2(1)$.

Proof. Using the Phillips-Solo (1992) device and Lemma D in Phillips (1999) we get

$$u_t = C(1)\epsilon_t + \tilde{\epsilon}_{t-1} - \tilde{\epsilon}_t, \tag{4.5}$$

where $\tilde{\epsilon}_t = \tilde{C}(L)\epsilon_t = \sum_{j=0}^{\infty} \tilde{c}_j \epsilon_{t-j}$, $\tilde{c}_j = \sum_{i=j+1}^{\infty} c_i$ and $\sum_{j=0}^{\infty} |\tilde{c}_j| < \infty$. Consequently,

$$\sum_{t=1}^{k} u_t = C(1) \sum_{t=1}^{k} \epsilon_t + \tilde{\epsilon}_0 - \tilde{\epsilon}_k, \tag{4.6}$$

and, for all $N \in \mathbf{N}$,

$$\sup_{0\leq r\leq N} \Big| \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t - C(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \Big| \leq \frac{\tilde{\epsilon}_0}{\sqrt{n}} + \sup_{0\leq r\leq N} \Big| \frac{\tilde{\epsilon}_{[nr]}}{\sqrt{n}} \Big| \leq 2 \max_{0\leq k\leq nN} \Big| \frac{\tilde{\epsilon}_k}{\sqrt{n}} \Big|. \tag{4.7}$$

By Lemmas 13.4 and 13.6,

$$\max_{0\leq k\leq nN} \Big| \frac{\tilde{\epsilon}_k}{\sqrt{n}} \Big| \to_P 0. \tag{4.8}$$

By Lemma 13.3, from relations (4.7) and (4.8) it follows that, for the Skorohod metric ρ on $\mathbb{D}(\mathbf{R}_+)$,

$$\rho\Big(\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t, C(1) \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \Big) \to_P 0.$$

By Lemma 13.1, this and Theorem 4.1 imply the desired result. ■

The following theorem gives a corresponding IP for sample covariances of martingale-difference sequences.

Theorem 4.3 ***(IP for sample covariances of martingale-difference sequences).*** *Let* $(\epsilon_t)_{t\in\mathbf{Z}}$ *satisfy assumption (D1) with* $p > 4$. *Then, for all* $m \geq 1$,

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \epsilon_{t+m} \to_d \sigma_\epsilon^2 W(r). \tag{4.9}$$

Throughout the rest of the paper, we will use the symbol $\mathcal{I}$ to denote different quantities in the proofs and η_t will denote auxiliary sequences of random variables arising in the arguments; these quantities and sequences are not necessarily the same from one place to another.

Proof. Construct the sequence of processes

$$M_n(s) = \sum_{i=1}^{k-1} \epsilon_i \Big(W\Big(\frac{T_{i+m}}{n}\Big) - W\Big(\frac{T_{i+m-1}}{n}\Big)\Big) + \epsilon_k \Big(W(s) - W\Big(\frac{T_{k+m-1}}{n}\Big)\Big)$$

for $\frac{T_{k+m-1}}{n} < s \leq \frac{T_{k+m}}{n}$, $k = 1, 2, ...$ Note that M_n is a continuous martingale with

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \epsilon_{t+m} =_d M_n\Big(\frac{T_{[nr]+m-1}}{n}\Big) \tag{4.10}$$

by Lemma 12.1. Using Theorem 3.2, we show that $M_n \to_d \sigma_\epsilon W$.

The first characteristics of M_n and $\sigma_\epsilon W$ are identically zero: $B_n(s) = B(s) = 0$, $s \geq 0$. The second characteristic of $\sigma_\epsilon W$ is $C(\sigma_\epsilon W)$, where, for an element $\alpha = (\alpha(s), s \geq 0)$, of the Skorohod space $\mathbb{D}(\mathbf{R}_+)$, $C(s,\alpha) = [\sigma_\epsilon W, \sigma_\epsilon W](s,\alpha) = \sigma_\epsilon^2 s$. The second characteristic of M_n is the process $C_n = (C_n(s), s \geq 0)$, where

$$C_n(s) = [M_n, M_n](s) = \sum_{i=1}^{k-1} \epsilon_i^2 \Big(\frac{T_{i+m}}{n} - \frac{T_{i+m-1}}{n} \Big) + \epsilon_k^2 \Big(s - \frac{T_{k+m-1}}{n} \Big)$$

for $\frac{T_{k+m-1}}{n} < s \leq \frac{T_{k+m}}{n}$, $k = 1, 2, \ldots$.

Condition (B1) of Theorem 3.2 is obviously satisfied with $F(s) = \sigma_\epsilon^2 s$. Condition (B2) of Theorem 3.2 is evidently satisfied by Theorem 11.2 (or by Remark 11.3). Conditions (B3) and (B4) of Theorem 3.2 and $[sup - \beta]$ in (B5) are trivially met.

Next, we have, for $\frac{T_{k+m-1}}{n} < s \leq \frac{T_{k+m}}{n}$, $k = 1, 2, \ldots$,

$$|C_n(s) - C(s, M_n)| = |C_n(s) - \sigma_\epsilon^2 s| = \Big| \sum_{i=1}^{k-1} (\epsilon_i^2 - \sigma_\epsilon^2) \Big(\frac{T_{i+m}}{n} - \frac{T_{i+m-1}}{n} \Big) + (\epsilon_k^2 - \sigma_\epsilon^2) \Big(s - \frac{T_{k+m-1}}{n} \Big) \Big|. \tag{4.11}$$

Since, by (12.2), for $N \in \mathbf{N}$,

$$\max_{k \geq 1} \{ k : T_{k-1}/n < N \} \leq KNn \ \ a.s. \tag{4.12}$$

for some constant $K \in \mathbf{N}$, condition $[\gamma - \mathbf{R}_+]$ in (B5) holds if

$$\mathcal{I}_{1n} = \max_{1 \leq k \leq KNn} \Big| \sum_{i=1}^{k-1} (\epsilon_i^2 - \sigma_\epsilon^2) \Big(\frac{T_{i+m}}{n} - \frac{T_{i+m-1}}{n} \Big) \Big| \to_P 0 \tag{4.13}$$

and

$$\mathcal{I}_{2n} = \max_{1 \leq k \leq KNn} \Big| (\epsilon_k^2 - \sigma_\epsilon^2) \Big(\frac{T_{k+m}}{n} - \frac{T_{k+m-1}}{n} \Big) \Big| \to_P 0. \tag{4.14}$$

Evidently,

$$\mathcal{I}_{1n} \leq \max_{1 \leq k \leq KNn} \frac{\sigma_\epsilon^2}{n} \Big| \sum_{i=1}^{k-1} (\epsilon_i^2 - \sigma_\epsilon^2) \Big| + \max_{1 \leq k \leq KNn} \Big| \sum_{i=1}^{k-1} (\epsilon_i^2 - \sigma_\epsilon^2) \Big(\frac{T_{i+m}}{n} - \frac{T_{i+m-1}}{n} - \frac{\sigma_\epsilon^2}{n} \Big) \Big| = \mathcal{I}_{1n}^{(1)} + \mathcal{I}_{1n}^{(2)}.$$

By the assumptions of the theorem and Lemma 12.1, $\eta_t^{(1)} = \epsilon_t^2 - \sigma_\epsilon^2$ and $\eta_t^{(2)} = (\epsilon_t^2 - \sigma_\epsilon^2)(T_{t+m} - T_{t+m-1} - \sigma_\epsilon^2)$, $t \geq 0$, are martingale-difference sequences with $E\big(\eta_0^{(1)}\big)^2 = E(\epsilon_0^2 - \sigma_\epsilon^2)^2 < \infty$ and $\sup_t E\big(\eta_t^{(2)}\big)^2 \leq LE(\epsilon_0^2 - \sigma_\epsilon^2)^2 \sup_t E(\epsilon_t^4 | \Im_{t-1}) < \infty$ for some constant L and all t. Therefore, from Lemma 13.5, we have $|\mathcal{I}_{1n}^{(1)}| \to_P 0$ and $|\mathcal{I}_{1n}^{(2)}| \to_P 0$ and thus (4.13) holds.

By (12.2),

$$\max_{1 \leq k \leq KNn} \Big| \frac{T_{k+m}}{n} - \frac{T_{k+m-1}}{n} \Big| = o(n^{q-1}), \tag{4.15}$$

for any $q \in \max(1/2, 2/p) = 1/2$. Since, under the assumptions of the theorem, $\max_{1 \leq k \leq KNn} n^{-2/p} |\epsilon_k^2 - \sigma_\epsilon^2| \to_P 0$ by Lemma 13.4, using (4.15) with $q \in (1/2, 1 - 2/q)$ (such a choice is possible since $p > 4$), we get (4.14) and thus $[\gamma - \mathbf{R}_+]$.

Consequently, all the conditions of Theorem 3.2 are satisfied and we have that $M_n \to_d \sigma_\epsilon W$. This, together with (4.3) and (4.10) implies, by Lemma 13.2, that $\frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \epsilon_t \epsilon_{t+m} \to_d \sigma_\epsilon W(\sigma_\epsilon^2 r)$, that is, (4.9) holds. ■

Remark 4.1 *Similar to the proof of Theorem 4.3, it is not difficult to obtain the following generalization of (4.9) to the case of martingale transforms. Let* $(\epsilon_t)_{t\in\mathbf{Z}}$ *satisfy assumption (D1) with* $p > 4$ *and let* $(Y_t)_{t\in\mathbf{N}_0}$ *be identically distributed mean-zero* $\Im_t$*-measurable random variables with* $EY_t^2 = \sigma_y^2$ *for all* t *and* $\sup_t E|Y_t|^p < \infty$ *for some* $p > 4$*. If* $\max_{1\le k\le n} \frac{1}{n}\sum_{t=1}^{n}(Y_t^2 - \sigma_y^2) \to_P 0$*, then, for all* $m \geq 1$*,* $\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]} Y_t\epsilon_{t+m} \to_d \tilde{\sigma}W(r)$*, where* $\tilde{\sigma} = \sigma_y\sigma_\epsilon$*. To establish the latter relation, the argument in the proof of Theorem 4.3 is repeated verbatim for the sequence of martingales*

$$M_n(s) = \sum_{i=1}^{k-1} Y_i\Big(W\Big(\frac{T_{i+m}}{n}\Big) - W\Big(\frac{T_{i+m-1}}{n}\Big)\Big) + Y_k\Big(W(s) - W\Big(\frac{T_{k+m-1}}{n}\Big)\Big), \tag{4.16}$$

$\frac{T_{k+m-1}}{n} < s \leq \frac{T_{k+m}}{n}$, $k = 1, 2, ...,$ *and the limit martingale* $\sigma_y W(s)$*.*

Theorem 4.4 ***(IP for sample covariances of linear processes)****. Suppose that* u_t *is the linear process* $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j\epsilon_{t-j}$*,* $C(L) = \sum_{j=0}^{\infty} c_jL^j$*, where* $\sum_{j=1}^{\infty} jc_j^2 < \infty$*,* $C(1) \neq 0$*, and* $(\epsilon_t)_{t\in\mathbf{Z}}$ *satisfy assumption (D2) with* $p \geq 4$*. Then, for all* $m \geq 1$*,*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}(u_tu_{t+m} - \gamma_m) \to_d v(m)W(r), \tag{4.17}$$

where[2] $\gamma_m = g_m(1)\sigma_\epsilon^2$*,* $v(m) = \Big(g_m^2(1)E(\epsilon_0^2 - \sigma_\epsilon^2)^2 + \sum_{r=1}^{\infty}(g_{m+r}(1) + g_{m-r}(1))^2\sigma_\epsilon^4\Big)^{1/2}$*,* $g_j(1) = \sum_{k=0}^{\infty} c_kc_{k+j}$*,* $j \in \mathbf{Z}$*, and it is assumed that* $c_j = 0$ *for* $j < 0$*.*

Proof. Treating c_j as zero for $j < 0$, define the lag polynomials $g_j(L)$, $j \in \mathbf{Z}$, by $g_j(L) = \sum_{k=0}^{\infty} c_kc_{k+j}L^k = \sum_{k=0}^{\infty} g_{jk}L^k$. Further, let $\tilde{g}_j(L) = \sum_{k=0}^{\infty}\tilde{g}_{jk}L^k$, where $\tilde{g}_{jk} = \sum_{s=k+1}^{\infty} g_{js} = \sum_{s=k+1}^{\infty} c_sc_{s+j}$. As in Remark 3.9 of Phillips and Solo (1992), we have

$$u_tu_{t+m} = g_m(L)\epsilon_t^2 + \sum_{r=1}^{\infty} g_{m+r}(L)\epsilon_{t-r}\epsilon_t + \sum_{r=1}^{m} g_{m-r}(L)\epsilon_t\epsilon_{t+r} +$$

$$\sum_{r=m+1}^{\infty} g_{r-m}(L)\epsilon_{t+m-r}\epsilon_{t+m} = g_m(1)\epsilon_t^2 + \sum_{r=1}^{\infty} g_{m+r}(1)\epsilon_{t-r}\epsilon_t + \sum_{r=1}^{m} g_{m-r}(1)\epsilon_t\epsilon_{t+r} +$$

$$\sum_{r=m+1}^{\infty} g_{r-m}(1)\epsilon_{t+m-r}\epsilon_{t+m} - (1-L)\tilde{u}_{at} - (1-L)\tilde{u}_{bt}, \tag{4.18}$$

where $\tilde{u}_{at} = \tilde{g}_m(L)\epsilon_t^2$ and $\tilde{u}_{bt} = +\sum_{r=1}^{\infty}\tilde{g}_{m+r}(L)\epsilon_{t-r}\epsilon_t + \sum_{r=1}^{m}\tilde{g}_{m-r}(L)\epsilon_t\epsilon_{t+r} + \sum_{r=m+1}^{\infty}\tilde{g}_{r-m}(L)\epsilon_{t+m-r}\epsilon_{t+m}$ (the validity of decomposition (4.18) follows from Lemma 3.6 in Phillips and Solo, 1992). Thus,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}(u_tu_{t+m} - \gamma_m) = \frac{1}{\sqrt{n}}g_m(1)\sum_{t=1}^{[nr]}(\epsilon_t^2 - \sigma_\epsilon^2) + \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\Big(\sum_{r=1}^{\infty} g_{m+r}(1)\epsilon_{t-r}\Big)\epsilon_t +$$

$$\sum_{r=1}^{m}\frac{1}{\sqrt{n}}\Big(\sum_{t=1}^{[nr]} g_{m-r}(1)\epsilon_t\epsilon_{t+r}\Big) + \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\Big(\sum_{r=m+1}^{\infty} g_{r-m}(1)\epsilon_{t+m-r}\Big)\epsilon_{t+m} -$$

$$\frac{1}{\sqrt{n}}(\tilde{u}_{a0} - \tilde{u}_{a,[nr]}) - \frac{1}{\sqrt{n}}(\tilde{u}_{b0} - \tilde{u}_{b,[nr]}). \tag{4.19}$$

Using Theorem 4.1 (applied to $\epsilon_t^2 - \sigma_\epsilon^2$) and Remark 4.1, it is not difficult to show that

$$\frac{1}{\sqrt{n}}g_m(1)\sum_{t=1}^{[nr]}(\epsilon_t^2 - \sigma_\epsilon^2) + \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\Big(\sum_{r=1}^{\infty} g_{m+r}(1)\epsilon_{t-r}\Big)\epsilon_t +$$

$$\sum_{r=1}^{m}\frac{1}{\sqrt{n}}\Big(\sum_{t=1}^{[nr]} g_{m-r}(1)\epsilon_t\epsilon_{t+r}\Big) + \frac{1}{\sqrt{n}}\sum_{t=1}^{[nr]}\Big(\sum_{r=m+1}^{\infty} g_{r-m}(1)\epsilon_{t+m-r}\Big)\epsilon_{t+m} \to_d v(m)W(r).$$

[2] $g_j(1)$ are the values of the lag polynomials defined in the proof.

By (4.19) and Lemmas 13.1 and 13.3, it remains to prove that, for all $N > 0$,

$$\sup_{0\le r\le N}\Big|\frac{1}{\sqrt{n}}(\tilde{u}_{a0}-\tilde{u}_{a,[nr]})+\frac{1}{\sqrt{n}}(\tilde{u}_{b0}-\tilde{u}_{b,[nr]})\Big| \to_P 0. \tag{4.20}$$

But this holds since, by Lemma 13.8, $Eu_{a0}^2 < \infty$ and $Eu_{b0}^2 < \infty$, and, thus, according to Lemma 13.4,

$$\max_{0\le k\le nN} n^{-1/2}|\tilde{u}_{a,k}| \to_P 0$$

and $\max_{0\le k\le nN} n^{-1/2}|\tilde{u}_{b,k}| \to_P 0$. ■

5. Convergence to stochastic integrals

The following result provides the conventional weak convergence limit theory for the sample covariance of ϵ_t and its partial sums to a stochastic integral that arises in a unit root autoregression. While other proofs of this result are available (using partial summation, for example), the following derivation shows that the result may be obtained directly by a martingale convergence argument.

Theorem 5.1 *Under assumption (D1) with $p > 4$,*

$$\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\epsilon_i\right)\epsilon_t \to_d \sigma_\epsilon^2\int_0^r W(v)dW(v). \tag{5.1}$$

Proof. Consider the process $M_n = (M_n(s), s\ge 0)$, where

$$M_n(s)=\sum_{i=1}^{k-1}W\left(\frac{T_{i-1}}{n}\right)\left(W\left(\frac{T_i}{n}\right)-W\left(\frac{T_{i-1}}{n}\right)\right)+W\left(\frac{T_{k-1}}{n}\right)\left(W(s)-W\left(\frac{T_{k-1}}{n}\right)\right) \tag{5.2}$$

for $\frac{T_{k-1}}{n} < s \le \frac{T_k}{n}$, $k = 1, 2, ...$ By Lemma 12.1, we have the following martingale representation for the left-hand side of (5.1):

$$\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\epsilon_i\right)\epsilon_t =_d M_n\Big(\frac{T_{[nr]}}{n}\Big). \tag{5.3}$$

Further, for $n\ge 1$, let $X_n = (X_n(s), s\ge 0)$ and $X = (X(s), s\ge 0)$ be the continuous vector martingales with $X_n(s) = (M_n(s), W(s))$ and $X(s) = (\int_0^s W(v)dW(v), W(s))$.

The first characteristic of X_n is identically zero: $B_n(s) = (0,0)\in \mathbf{R}^2$, $s\ge 0$. The second characteristic of X_n is the process $C_n = (C_n(s), s\ge 0)$ with

$$C_n(s)=[X_n,X_n](s)=\begin{pmatrix} C_n^{11}(s) & C_n^{12}(s)\\ C_n^{21}(s) & C_n^{22}(s)\end{pmatrix}, \tag{5.4}$$

where

$$C_n^{11}(s)=\sum_{i=1}^{k-1}W^2\left(\frac{T_{i-1}}{n}\right)\left(\frac{T_i}{n}-\frac{T_{i-1}}{n}\right)+W^2\left(\frac{T_{k-1}}{n}\right)\left(s-\frac{T_{k-1}}{n}\right), \tag{5.5}$$

$$C_n^{12}(s)=C_n^{21}(s)=\sum_{i=1}^{k-1}W\left(\frac{T_{i-1}}{n}\right)\left(\frac{T_i}{n}-\frac{T_{i-1}}{n}\right)+W\left(\frac{T_{k-1}}{n}\right)\left(s-\frac{T_{k-1}}{n}\right) \tag{5.6}$$

for $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$, $k = 1, 2, ...$, and

$$C_n^{22}(s) = s. \tag{5.7}$$

The process X is a solution to the stochastic differential equation (11.6) with $g_1(x) = x$, $x \in \mathbf{R}$, and $g_2(x) = 0$, $x \in \mathbf{R}$. Its first and second predictable characteristics are, respectively, $B(X)$ and $C(X)$, where B and C are defined in (11.7) with the above functions $g_i(x)$, $i = 1, 2$ (so that the first predictable characteristic of X is identically zero, as that of X_n: $B(s) = (0, 0) \in \mathbf{R}^2$).

Obviously, the strong majoration condition (B1) of Theorem 3.2 is not satisfied for the limit semimartingale X. Therefore, in contrast to the proof of Theorem 4.3, we apply Theorem 3.1 instead of Theorem 3.2 to show that $X_n \to_d X$.

Let $a \geq 0$ and let $0 \leq r < s$. If $\alpha = (\alpha(s), s \geq 0)$, $\alpha(s) = (\alpha_1(s), \alpha_2(s))$, is an element of the Skorohod space $\mathbb{D}(\mathbf{R}_+^2)$, then for the stopping time $S^a(\alpha)$ defined in (3.1) and all $v \in (r \wedge S^a(\alpha), s \wedge S^a(\alpha))$, we have $\alpha_2^2(v) \leq |\alpha(v)|^2 < a^2$. Consequently, for $C^{ij}(s, \alpha)$, $i, j = 1, 2$, defined in (11.7) with $g_1(x) = x$, one has

$$C^{11}(s \wedge S^a(\alpha), \alpha) - C^{11}(r \wedge S^a(\alpha), \alpha) = \int_{r \wedge S^a(\alpha)}^{s \wedge S^a(\alpha)} \alpha_2^2(v) dv \leq a^2(s - r), \tag{5.8}$$

$$C^{22}(s \wedge S^a(\alpha), \alpha) - C^{22}(r \wedge S^a(\alpha), \alpha) = s \wedge S^a(\alpha) - r \wedge S^a(\alpha) \leq (s - r). \tag{5.9}$$

Thus, condition (A1) of Theorem 3.1 is satisfied with $F(s, a) = \max(1, a^2)s$.

Since the functions $g_1(x) = x$ and $g_2(x) = 0$ are obviously Lipschitz continuous and satisfy growth condition (11.8), Corollaries 11.1 and 11.2 ensure that the uniqueness and measurability hypotheses (A2) and (A3) and the continuity condition (A4) of Theorem 3.1 are satisfied. Condition (A5) of Theorem 3.1 is trivially satisfied since $X_n(0) = X(0) = 0$. Condition $[sup - \beta]$ (and thus $[sup - \beta_{loc}]$) is satisfied since $B_n(s) = 0$, $s \geq 0$.

From the definition of $C(s, \alpha)$ in (11.7) with $g_1(x) = x$, we have

$$C(s, X_n) = \begin{pmatrix} \int_0^s W^2(v) dv & \int_0^s W(v) dv \\ \int_0^s W(v) dv & s \end{pmatrix} = \begin{pmatrix} \tilde{C}^{11}(s) & \tilde{C}^{12}(s) \\ \tilde{C}^{21}(s) & \tilde{C}^{22}(s) \end{pmatrix}, \tag{5.10}$$

where $\tilde{C}^{11}(s) = \int_0^s W^2(v) dv$, $\tilde{C}^{12}(s) = \tilde{C}^{21}(s) = \int_0^s W(v) dv$ and $\tilde{C}^{22}(s) = s$. Then, by (5.5) and (5.6), for $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$, $k = 1, 2, ...$, we have

$$\begin{aligned} |C_n^{11}(s) - \tilde{C}^{11}(s)| &= \Big| \sum_{i=1}^{k-1} \int_{\frac{T_{i-1}}{n}}^{\frac{T_i}{n}} \Big(W^2\Big(\frac{T_{i-1}}{n}\Big) - W^2(v) \Big) dv + \int_{\frac{T_{k-1}}{n}}^{s} \Big(W^2\Big(\frac{T_{k-1}}{n}\Big) - W^2(v) \Big) dv \Big| \leq \\ & s \max_{1 \leq i \leq k} \sup_{v_1, v_2 \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} |W^2(v_1) - W^2(v_2)|, \end{aligned} \tag{5.11}$$

$$\begin{aligned} |C_n^{12}(s) - \tilde{C}^{12}(s)| &= \Big| \sum_{i=1}^{k-1} \int_{\frac{T_{i-1}}{n}}^{\frac{T_i}{n}} \Big(W\Big(\frac{T_{i-1}}{n}\Big) - W(v) \Big) dv + \int_{\frac{T_{k-1}}{n}}^{s} \Big(W\Big(\frac{T_{k-1}}{n}\Big) - W(v) \Big) dv \Big| \leq \\ & s \max_{1 \leq i \leq k} \sup_{v_1, v_2 \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} |W(v_1) - W(v_2)|. \end{aligned}$$

Thus, for $\frac{T_{k-1}}{n} < N \leq \frac{T_k}{n}$, $k = 1, 2, ...$,

$$\sup_{0 \leq s \leq N} \big| C_n^{11}(s) - \tilde{C}^{11}(s) \big| \leq N \max_{1 \leq i \leq k} \sup_{v_1, v_2 \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} |W^2(v_1) - W^2(v_2)|, \tag{5.12}$$

$$\sup_{0 \leq s \leq N} \big| C_n^{12}(s) - \tilde{C}^{12}(s) \big| \leq N \max_{1 \leq i \leq k} \sup_{v_1, v_2 \in [\frac{T_{i-1}}{n}, \frac{T_i}{n}]} |W(v_1) - W(v_2)|. \tag{5.13}$$

From (4.12), (4.15), (5.12) and (5.13) and the uniform continuity of Brownian sample paths it follows that

$$\sup_{0\le s\le N}\left|C_n^{11}(s)-\tilde{C}^{11}(s)\right| \to_P 0, \tag{5.14}$$

and

$$\sup_{0\le s\le N}\left|C_n^{12}(s)-\tilde{C}^{12}(s)\right| = \sup_{0\le s\le N}\left|C_n^{21}(s)-\tilde{C}^{21}(s)\right| \to_P 0 \tag{5.15}$$

for all $N \in \mathbf{N}$. Relations (5.14) and (5.15), together with $C_n^{22}(s)=\tilde{C}^{22}(s)=s$ evidently imply that

$$\sup_{0\le s\le N}|C_n(s)-C(s,X_n)| \to_P 0,$$

for all $N \in \mathbf{N}$. Consequently, condition $[sup-\gamma]$ (and thus $[\gamma_{loc}-\mathbf{R}_+^2]$) of Theorem 3.1 is satisfied. We therefore have $X_n \to_d X$. This result, together with (4.3) and (5.3), implies relation (5.1) by virtue of Lemma 13.2. ■

We now formulate and prove the corresponding law for the linear process case. The outcome is a semimartingale convergence result.

Theorem 5.2 *Suppose that u_t is the linear process $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j\epsilon_{t-j}$, $C(L)=\sum_{j=0}^{\infty} c_jL^j$, where $\sum_{j=1}^{\infty} j|c_j|<\infty$, $C(1)\neq 0$, and $(\epsilon_t)_{t\in\mathbf{Z}}$ satisfy assumption (D2) with $p>4$. Then*

$$\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)u_t \to_d r\lambda+\omega^2\int_0^r W(v)dW(v), \tag{5.16}$$

where $\lambda=\sum_{j=1}^{\infty} Eu_0u_j$ and $\omega^2={\sigma_\epsilon}^2C^2(1)$.

Remark 5.1 *Suppose that u_t and v_t are two linear processes: $u_t=\Gamma(L)\epsilon_t=\sum_{j=0}^{\infty}\gamma_j\epsilon_{t-j}$, $v_t=\Delta(L)\epsilon_t=\sum_{j=0}^{\infty}\delta_j\epsilon_{t-j}$, $\Gamma(L)=\sum_{j=0}^{\infty}\gamma_jL^j$, $\Delta(L)=\sum_{j=0}^{\infty}\delta_jL^j$, where $\sum_{j=1}^{\infty}j|\gamma_j|<\infty$, $\sum_{j=1}^{\infty}j|\delta_j|<\infty$, $\Gamma(1)\neq 0$, $\Delta(1)\neq 0$, and $(\epsilon_t)_{t\in\mathbf{Z}}$ satisfy assumption (D2) with $p>4$. Then, analogous to the proof of Theorem 5.2, one can show, using Theorem 3.1, that $\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)v_t \to_d r\lambda_{uv}+\omega_u\omega_v\int_0^r W(v)dW(v)$, where $\omega_u^2=\sigma_\epsilon^2\Gamma^2(1)$, $\omega_v^2=\sigma_\epsilon^2\Delta^2(1)$ and $\lambda_{uv}=\sum_{j=1}^{\infty}Eu_0v_j$.*

Proof. As in the proof of Theorem 5.1, one can show that, for all $\mu\in\mathbf{R}$,

$$r\mu+\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\epsilon_i\right)\epsilon_t \to_d r\mu+\sigma_\epsilon^2\int_0^r W(v)dW(v). \tag{5.17}$$

Indeed, for M_n defined in (5.2), we have the following semimartingale representation for the left-hand side of (5.17):

$$r\mu+\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\epsilon_i\right)\epsilon_t =_d r\mu+M_n\Big(\frac{T_{[nr]}}{n}\Big). \tag{5.18}$$

Similar to the proof of Theorem 5.1, for $n\ge 1$, let $X_n=(X_n(s), s\ge 0)$ and $X=(X(s), s\ge 0)$ be the continuous vector semimartingales with $X_n(s)=(s\mu+M_n(s), W(s))$ and $X(s)=(s\mu+\int_0^s W(v)dW(v), W(s))$. The first characteristic of X_n is the process $B_n=(B_n(s), s\ge 0)$ with

$$B_n(s)=(s\mu,0)=(B_n^1(s),B_n^2(s)). \tag{5.19}$$

The second characteristic of X_n is the same as that of (M_n, W) and is given by the process $C_n=(C_n(s), s\ge 0)$ defined in (5.4).

The process X is a solution to stochastic differential equation (11.6) with $g_1(x) = x$, $x \in \mathbf{R}$, and $g_2(x) = \mu$, $x \in \mathbf{R}$. The first and second characteristic of X are the processes $B(X)$ and $C(X)$, where B and C are defined in (11.7) with the above functions $g_i(x)$, $i = 1, 2$ (so that the first predictable characteristic of X is the same as that of $X_n : B(s) = (s\mu, 0)$).

As in the proof of Theorem 5.1, we show that (5.17) holds by verifying the conditions of Theorem 3.1. For $\alpha = (\alpha(s), s \geq 0)$, $\alpha(s) = (\alpha_1(s), \alpha_2(s))$, an element of the Skorohod space $\mathbb{D}(\mathbf{R}^2_+)$, and $B^1(s, \alpha)$ and $B^2(s, \alpha)$ as in (11.7) with $g_1(x) = x$, $g_2(x) = \mu$, $x \in \mathbf{R}$, denote $H(s, \alpha) = Var(B^1)(s, \alpha) + Var(B^2)(s, \alpha) = s|\mu|$ (see Definition 2.2). For the stopping time $S^a(\alpha)$ defined in (3.1) and for all $r < s$ we have

$$H(s \wedge S_a(\alpha), \alpha) - H(r \wedge S^a(\alpha), \alpha) = |\mu|(s \wedge S^a(\alpha) - r \wedge S^a(\alpha)) \leq |\mu|(s - r). \tag{5.20}$$

This result and (5.8) and (5.9) imply that condition (A1) of Theorem 3.1 is satisfied with
$F(s, a) = \max(1, |\mu|, a^2)s$.

Since the functions $g_1(x) = x$ and $g_2(x) = \mu$ are obviously locally Lipschitz continuous and satisfy growth condition (11.8), we conclude, by Corollaries 11.1 and 11.2, that conditions (A2)-(A4) of Theorem 3.1 are satisfied. Condition (A5) of Theorem 3.1 is again trivially satisfied since $X_n(0) = X(0) = 0$. Condition $[sup - \beta]$ (and $[sup - \beta_{loc}]$) holds since $B(s, X_n) = B_n(s) = s\mu$ for all $n \geq 1$. Since, as shown in the proof of Theorem 5.1, $\sup_{0 \leq s \leq N} |C_n(s) - C(s, X_n)| \to_P 0$ for all $N \in \mathbf{N}$ and thus conditions $[sup - \gamma]$ (and $[\gamma_{loc} - R^2_+]$) of Theorem 3.1 holds, we get (5.17).

To complete the proof, let us now show that, for all $N \in \mathbf{N}$,

$$\sup_{0 \leq r \leq N} \Big| \frac{1}{n} \sum_{t=2}^{[nr]} \left(\sum_{i=1}^{t-1} u_i \right) u_t - r\lambda - \frac{1}{n} C^2(1) \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} \epsilon_i \right) \epsilon_t \Big| \to_P 0. \tag{5.21}$$

Using the Phillips-Solo device as in the proof of Theorem 4.2, from (4.5) and (4.6) we find that, for all $N \in \mathbf{N}$ and all $r \in [0, N]$,

$$\begin{aligned}
& \Big| \frac{1}{n} \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} u_i \right) u_t - r\lambda - \frac{1}{n} C^2(1) \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} \epsilon_i \right) \epsilon_t \Big| \\
= & \Big| \frac{1}{n} C(1) \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} u_i \right) \epsilon_t + \frac{1}{n} \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} u_i \right) (\tilde{\epsilon}_{t-1} - \tilde{\epsilon}_t) - r\lambda - \frac{1}{n} C^2(1) \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} \epsilon_i \right) \epsilon_t \Big| \\
= & \Big| \frac{1}{n} C^2(1) \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} \epsilon_i \right) \epsilon_t + \frac{1}{n} C(1) \sum_{t=1}^{[nr]} (\tilde{\epsilon}_0 - \tilde{\epsilon}_{t-1}) \epsilon_t \\
& + \frac{1}{n} \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} u_i \right) (\tilde{\epsilon}_{t-1} - \tilde{\epsilon}_t) - r\lambda - \frac{1}{n} C^2(1) \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} \epsilon_i \right) \epsilon_t \Big| \\
= & \Big| \frac{1}{n} C(1) \sum_{t=1}^{[nr]} (\tilde{\epsilon}_0 - \tilde{\epsilon}_{t-1}) \epsilon_t + \frac{1}{n} \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} u_i \right) (\tilde{\epsilon}_{t-1} - \tilde{\epsilon}_t) - r\lambda |.
\end{aligned}$$

Therefore, the expression on the left-hand side of (5.21) is dominated by

$$\begin{aligned}
& \sup_{0 \leq r \leq N} \Big| \frac{1}{n} C(1) \sum_{t=1}^{[nr]} \tilde{\epsilon}_0 \epsilon_t \Big| + \sup_{0 \leq r \leq N} \Big| \frac{1}{n} C(1) \sum_{t=1}^{[nr]} \tilde{\epsilon}_{t-1} \epsilon_t \Big| \\
& + \sup_{0 \leq r \leq N} \Big| \frac{1}{n} \sum_{t=1}^{[nr]} \left(\sum_{i=1}^{t-1} u_i \right) (\tilde{\epsilon}_{t-1} - \tilde{\epsilon}_t) - r\lambda \Big| \\
= & \ \mathcal{I}_{1n} + \mathcal{I}_{2n} + \mathcal{I}_{3n}.
\end{aligned} \tag{5.22}$$

By Lemma 13.6, $E\tilde{\epsilon}_0^2 < \infty$, under the assumptions of the theorem. Therefore, $\eta_t^{(1)} = \tilde{\epsilon}_0 \epsilon_t$, $t \geq 1$, and $\eta_t^{(2)} = \tilde{\epsilon}_{t-1} \epsilon_t$, $t \geq 1$, are martingale-difference sequences with $E\big(\eta_t^{(1)}\big)^2 = E\big(\eta_t^{(2)}\big)^2 = E\tilde{\epsilon}_0^2 E\epsilon_0^2 < \infty$. Consequently, from Lemma 13.5 we get that $\mathcal{I}_{1n} = \max_{1 \leq k \leq nN} \big| \frac{1}{n} \sum_{t=1}^{k} \eta_t^{(1)} \big| \to_P 0$ and $\mathcal{I}_{2n} = \max_{1 \leq k \leq nN} \big| \frac{1}{n} \sum_{t=1}^{k} \eta_t^{(2)} \big| \to_P 0$ in (5.22).

Using summation by parts, we get that

$$\mathcal{I}_{3n} = \sup_{0\le r\le N}\Big| -\frac{1}{n}\Big(\sum_{t=1}^{[nr]} u_t\Big)\tilde{\epsilon}_{[nr]} + \frac{1}{n}\sum_{t=1}^{[nr]} u_t\tilde{\epsilon}_t - r\lambda\Big|$$

$$(5.23)\qquad \le \sup_{0\le r\le N}\Big|\frac{1}{n}\Big(\sum_{t=1}^{[nr]} u_t\Big)\tilde{\epsilon}_{[nr]}\Big| + \sup_{0\le r\le N}\Big|\frac{1}{n}\sum_{t=1}^{[nr]} u_t\tilde{\epsilon}_t - r\lambda\Big| = \mathcal{I}'_{3n} + \mathcal{I}''_{3n}.$$

Evidently, $\mathcal{I}'_{3n} \to_P 0$ holds if $\max_{1\le k\le nN} n^{-1}\Big|\sum_{t=1}^{k} u_t\Big||\tilde{\epsilon}_k| \to_P 0$. This, on the other hand, follows from (4.8) and the property that, by Theorem 4.2,

$$(5.24)\qquad \max_{1\le k\le nN} n^{-1/2}\Big|\sum_{t=1}^{k} u_t\Big| = O_P(1).$$

Let us show that $\mathcal{I}''_{3n} \to_P 0$. For $\tilde{c}_j$ as in the proof of Theorem 4.2, let $h_r(L)$, $r \ge 0$, be the lag polynomials defined by $h_0(L) = \sum_{k=0}^{\infty} c_k\tilde{c}_k L^k = \sum_{k=0}^{\infty} h_{kk}L^k$, $h_r(L) = \sum_{k=0}^{\infty}(c_k\tilde{c}_{k+r} + \tilde{c}_k c_{k+r})L^k = \sum_{k=0}^{\infty} h_{kr}L^k$, $r \ge 1$. Further, let, for $r \ge 0$, $\tilde{h}_r(L) = \sum_{k=0}^{\infty}\tilde{h}_{kr}L^k$, where $\tilde{h}_{kr} = \sum_{j=k+1}^{\infty} h_{jr}$. Similar to the derivations of second order BN decompositions in Phillips and Solo (1992) and the proof of Theorem 4.4, it is not difficult it is not difficult to see that

$$(5.25)\qquad u_t\tilde{\epsilon}_t = h_0(L)\epsilon_t^2 + \sum_{r=1}^{\infty} h_r(L)\epsilon_t\epsilon_{t-r} = h_0(1)\epsilon_t^2 - (1-L)\tilde{w}_{at} + \epsilon_t\epsilon_{t-1}^h - (1-L)\tilde{w}_{bt},$$

where $\tilde{w}_{at} = \tilde{h}_0(L)\epsilon_t^2$, $\epsilon_{t-1}^h = \sum_{r=1}^{\infty} h_r(1)\epsilon_{t-r}$ and $\tilde{w}_{bt} = \sum_{r=1}^{\infty}\tilde{h}_r(L)\epsilon_t\epsilon_{t-r}$ (the validity of decomposition (5.25) is justified by Lemma 13.9).

By (5.25), we have that, for all $k \ge 1$,

$$\frac{1}{n}\sum_{t=1}^{k} u_t\tilde{\epsilon}_t = \frac{1}{n}h_0(1)\sum_{t=1}^{k}\epsilon_t^2 + \frac{1}{n}\tilde{w}_{a0} - \frac{1}{n}\tilde{w}_{ak} + \frac{1}{n}\sum_{t=1}^{k}\epsilon_t\epsilon_{t-1}^h + \frac{1}{n}\tilde{w}_{b0} - \frac{1}{n}\tilde{u}_{bk}$$

and, thus, for all $N \in \mathbf{N}$,

$$\max_{1\le k\le Nn}\Big|\frac{1}{n}\sum_{t=1}^{k} u_t\tilde{\epsilon}_t - k\lambda\Big| \le \max_{1\le k\le Nn}\Big|\frac{1}{n}\sum_{t=1}^{k}(h_0(1)\epsilon_t^2 - \lambda)\Big| + 2\max_{0\le k\le Nn}\Big|\frac{1}{n}\tilde{w}_{ak}\Big|$$

$$+ \max_{1\le k\le Nn}\Big|\frac{1}{n}\sum_{t=1}^{k}\epsilon_t\epsilon_{t-1}^h\Big| + 2\max_{0\le k\le Nn}\Big|\frac{1}{n}\tilde{w}_{bk}\Big|.$$

It is not difficult to see that

$$(5.26)\qquad \lambda = h_0(1)\sigma_\epsilon^2.$$

Therefore, $\eta_t^{(3)} = h_0(1)\epsilon_t^2 - \lambda$, $t \ge 1$, is a martingale-difference sequence with $E\big(\eta_t^{(3)}\big)^2 = E(h_0(1)\epsilon_t^2 - h_0(1)\sigma_\epsilon^2)^2 \le LE|\epsilon_0|^4$ for some constant $L > 0$ and all t. Similarly, by Lemma 13.11, $\eta_t^{(4)} = \epsilon_t\epsilon_{t-1}^h$, $t \ge 1$, is a martingale-difference sequence with $E\big(\eta_t^{(4)}\big)^2 = E\epsilon_0^2 E\big(\epsilon_{-1}^h\big)^2 < \infty$ for all t.

Thus, using Lemma 13.5, we get that $\max_{1\le k\le Nn}\big|\frac{1}{n}\sum_{t=1}^{k}(h_0(1)\epsilon_t^2 - \lambda)\big| = \max_{1\le k\le Nn}\big|\eta_t^{(1)}\big| \to_P 0$ and $\max_{1\le k\le Nn}\big|\frac{1}{n}\sum_{t=1}^{k}\epsilon_t\epsilon_{t-1}^h\big| = \max_{1\le k\le Nn}\big|\eta_t^{(2)}\big| \to_P 0$.

Since, by Lemma 13.10, $E|\tilde{w}_{a0}|^{p/2} < \infty$ and $E|\tilde{w}_{b0}|^{p/2} < \infty$ under the assumptions of the theorem, from Lemma 13.4 we get that $\max_{0\le k\le nN} n^{-1}|\tilde{w}_{a,k}| \to_P 0$ and $\max_{0\le k\le nN} n^{-1}|\tilde{w}_{b,k}| \to_P 0$. Consequently, $\mathcal{I}''_{3n} \to_P 0$ and, thus, by (5.22), convergence (5.21) indeed holds. By Lemmas 13.1 and 13.3, relations (5.17) and (5.21) imply (5.16). ■

6. Asymptotics for general functionals of partial sums

The martingale convergence approach developed here can also be used to derive asymptotic results for functionals of partial sums of linear processes. These results are particularly useful in practice for models where nonlinear functions of integrated processes arise. In this context, the following generalization of Theorems 5.1 and 5.2 holds.

Theorem 6.1 *Let $f : \mathbf{R} \to \mathbf{R}$ be a twice continuously differentiable function such that f' satisfies the growth condition*[3] *$|f'(x)| \le K(1+|x|^\alpha)$ for some constants $K > 0$ and $\alpha > 0$ and all $x \in \mathbf{R}$. Suppose that u_t is the linear process $u_t = C(L)\epsilon_t = \sum_{j=0}^\infty c_j\epsilon_{t-j}$, $C(L) = \sum_{j=0}^\infty c_jL^j$, where $\sum_{j=1}^\infty j|c_j| < \infty$, $C(1) \ne 0$, and $(\epsilon_t)_{t\in\mathbf{Z}}$ satisfy assumption (D2) with $p \ge \max(6, 4\alpha)$. Then*

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{[nr]} f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\right)u_t \to_d \lambda\int_0^r f'(\omega W(v))dv + \omega\int_0^r f(\omega W(v))dW(v), \tag{6.1}$$

where $\lambda = \sum_{j=1}^\infty Eu_0u_j$ and $\omega^2 = {\sigma_\epsilon}^2C^2(1)$.

Remark 6.1 *The processes on the right-hand side of (6.1) belong to an important class of limit semimartingales for functionals of partial sums of linear processes whose first predictable characteristics (the drift terms) are non-deterministic. The latter is a qualitative difference between the semimartingales in (6.1) and the processes on the right-hand side of (5.16), where the first characteristics are deterministic ($r\lambda, r \ge 0$).*

Remark 6.2 *From the proof of Theorem 6.1 it follows that the assumption that f is twice continuously differentiable can be replaced by the condition that f has a locally Lipschitz continuous first derivative, that is, for every $N \in \mathbf{N}$ there exists a constant K_N such that $|f'(x) - f'(y)| \le K_N|x-y|$ for all $x, y \in \mathbf{R}$ with $|x| \le N$ and $|y| \le N$.*

Remark 6.3 *Similar to Remark 5.1, from the proof of Theorem 6.1 we find that the following extension holds. Let $f : R \to R$ be a twice continuously differentiable function such that f' satisfies the growth condition $|f'(x)| \le K(1+|x|^\alpha)$ for some constants $K > 0$ and $\alpha > 0$ and all $x \in R$. Suppose that u_t and v_t are two linear processes: $u_t = \Gamma(L)\epsilon_t = \sum_{j=0}^\infty \gamma_j\epsilon_{t-j}$, $v_t = \Delta(L)\epsilon_t = \sum_{j=0}^\infty \delta_j\epsilon_{t-j}$, $\Gamma(L) = \sum_{j=0}^\infty \gamma_jL^j$, $\Delta(L) = \sum_{j=0}^\infty \delta_jL^j$, where $\sum_{j=1}^\infty j|\gamma_j| < \infty$, $\sum_{j=1}^\infty j|\delta_j| < \infty$, $\Gamma(1) \ne 0$, $\Delta(1) \ne 0$, and $(\epsilon_t)_{t\in\mathbf{Z}}$ satisfy assumption (D2) with $p \ge \max(6, 4\alpha)$. Then, $\frac{1}{\sqrt{n}}\sum_{t=2}^{[nr]} f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\right)u_t \to_d \lambda_{uv}\int_0^r f'(\omega_uW(v))dv + \omega_v\int_0^r f(\omega_uW(v))dW(v)$, where $\omega_u^2 = \sigma_\epsilon^2\Gamma^2(1)$, $\omega_v^2 = \sigma_\epsilon^2\Delta^2(1)$ and $\lambda_{uv} = \sum_{j=1}^\infty Eu_0v_j$.*

One should also note that, as follows from the proof of the Theorem, if ϵ_t satisfies assumption (D1) with $p > 6$ (so that $\lambda = 0$) then the relation $\frac{1}{\sqrt{n}}\sum_{t=2}^{[nr]} f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}\epsilon_i\right)\epsilon_t \to_d \sigma_\epsilon\int_0^r f(\sigma_\epsilon W(v))dW(v)$ holds if f satisfies the exponential growth condition $|f(x)| \le 1 + \exp(K|x|)$ for some constant $K > 0$ and all $x \in R$.

Remark 6.4 *The assumption $|f'(x)| \le K(1+|x|^\alpha)$, together with the moment condition $E|\epsilon_0|^p < \infty$ for $p >$ $\max(6, 4\alpha)$, guarantees, by Lemma 13.12, that bound (13.12) for moments of partial sums in the Appendix holds. As follows from the proof, Theorem 6.1 in fact holds for $p \ge 6$ and all twice continuously differentiable functions f for which the estimate (13.12) is true and f' (and, thus, f itself) satisfies the exponential growth condition $|f'(x)| \le 1 + \exp(K|x|)$ for some constant $K > 0$ and all $x \in \mathbf{R}$.*

Proof. We first show that

$$\begin{aligned}\mathcal{I}_n &= \frac{\lambda}{n}\sum_{t=2}^{[nr]} f'\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\right) + \frac{1}{\sqrt{n}}\sum_{t=2}^{[nr]} f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\right)\epsilon_t \to_d \\ &\quad \lambda\int_0^r f'(\omega W(v))dv + \omega\int_0^r f(\omega W(v))dW(v).\end{aligned} \tag{6.2}$$

[3] This assumption evidently implies that f satisfies a similar growth condition with the power $1+\alpha$, i.e., $|f(x)| \le K(1+|x|^{1+\alpha})$ for some constant K and all $x \in \mathbf{R}$.

Consider the continuous semimartingale $M_n = (M_n(s), s \geq 0)$, where

$$
\begin{aligned}
M_n(s) \quad &= \quad \frac{\lambda}{n}\sum_{i=2}^{k-1} f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big) + \lambda f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big)\Big(s - \frac{k-1}{n}\Big) \\
&\quad + \sum_{i=1}^{k-1} f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big)\left(W\left(\frac{T_i}{n}\right) - W\left(\frac{T_{i-1}}{n}\right)\right) \\
&\quad + f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big)\left(W(s) - W\left(\frac{T_{k-1}}{n}\right)\right),
\end{aligned} \tag{6.3}
$$

for $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$, $k = 1, 2, \ldots$ By Lemma 12.1, we have the following semimartingale representation for the left-hand side of (6.2) :

$$
\mathcal{I}_n =_d M_n\Big(\frac{T_{[nr]}}{n}\Big). \tag{6.4}
$$

Further, let $X_n = (X_n(s), s \geq 0)$ for $n \geq 1$ and $X = (X(s), s \geq 0)$ be the continuous vector martingales with $X_n(s) = (M_n(s), W(s))$ and $X(s) = (h_0(1)\int_0^s f'(C(1)W(v))dv + \int_0^s f(C(1)W(v))dW(v), W(s))$, where, as in (5.26), $h_0(1) = \lambda/\sigma_\epsilon^2$.

The first characteristic of X_n is the process $(B_n(s), s \geq 0)$, where

$$
B_n(s) = \Big(\frac{\lambda}{n}\sum_{i=2}^{k-1} f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big) + \lambda f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big)\Big(s - \frac{k-1}{n}\Big), 0\Big) = (B_n^1(s), B_n^2(s)) \tag{6.5}
$$

for $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$, $k = 1, 2, \ldots$ The second characteristic of X_n is the process $C_n = (C_n(s), s \geq 0)$ with

$$
C_n(s) = \begin{pmatrix} C_n^{11}(s) & C_n^{12}(s) \\ C_n^{21}(s) & C_n^{22}(s) \end{pmatrix}, \tag{6.6}
$$

where

$$
C_n^{11}(s) = \sum_{i=2}^{k-1} f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big)\left(\frac{T_i}{n} - \frac{T_{i-1}}{n}\right) + f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big)\left(s - \frac{T_{k-1}}{n}\right), \tag{6.7}
$$

$$
C_n^{12}(s) = C_n^{21}(s) = \sum_{i=2}^{k-1} f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big)\left(\frac{T_i}{n} - \frac{T_{i-1}}{n}\right) + f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big)\left(s - \frac{T_{k-1}}{n}\right), \tag{6.8}
$$

for $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$, $k = 1, 2, \ldots$, and

$$
C_n^{22}(s) = s. \tag{6.9}
$$

The process X is a solution to stochastic differential equation (11.6) with $g_1(x) = f(C(1)x)$, $x \in \mathbf{R}$, and $g_2(x) = h_0(1)f'(C(1)x)$, $x \in \mathbf{R}$. The first and second predictable characteristics of X are, respectively, $B(X)$ and $C(X)$, where B and C are defined in (11.7) with the above $g_i(x)$, $i = 1, 2$.

As in the proof of Theorems 5.1 and 5.2, we proceed to show that $X_n \to_d X$ by verifying the conditions of Theorem 3.1 in order.

For $x \in \mathbf{R}$, let $x_+ = max(x, 0)$ and $x_- = max(-x, 0)$ and let $B^i(s, \alpha)$, $i = 1, 2$, and $C^{ij}(s, \alpha)$, $1 \leq i, j \leq 2$, be as in (11.7) with $g_1(x) = f(C(1)x)$ and $g_2(x) = h_0(1)f'(C(1)x)$. Since, obviously, $B^1(s, \alpha) = \int_0^s [h_0(1)f'(C(1)\alpha_2(v))]_+ dv - \int_0^s [h_0(1)f'(C(1)\alpha_2(v))]_- dv$ for $\alpha = ((\alpha_1(s), \alpha_2(s)), s \geq 0) \in \mathbb{D}(\mathbf{R}_+^2)$, one has (see Definition 2.2)

$$
Var(B^1)(s, \alpha) + Var(B^2)(s, \alpha) = \int_0^s [h_0(1)f'(C(1)\alpha_2(v))]_+ dv + \int_0^s [h_0(1)f'(C(1)\alpha_2(v))]_- dv =
$$

$$\int_0^s |h_0(1) f'(C(1)\alpha_2(v))| dv = H(s,\alpha).$$

Let $0 \le r < s$. For the stopping time $S^a(\alpha)$ defined in (3.1) and for all $v \in (r \wedge S^a(\alpha), s \wedge S^a(\alpha))$ we have $|\alpha_2(v)| \le |\alpha(v)| < a$ and thus $|f(C(1)\alpha_2(v))| \le \max_{|x|<a} |f(C(1)x)| = G_1(a)$ and $|f'(C(1)\alpha_2(v))| \le \max_{|x|<a} |f'(C(1)x)| = G_2(a)$. Consequently,

$$H(s \wedge S_a(\alpha),\alpha) - H(r \wedge S^a(\alpha),\alpha) = \int_{r\wedge S^a(\alpha)}^{s\wedge S^a(\alpha)} |h_0(1) f'(C(1)W(v))| dv \le |h_0(1)| G_2(a)(s-r), \tag{6.10}$$

$$C^{11}(s \wedge S^a(\alpha),\alpha) - C^{11}(r \wedge S^a(\alpha),\alpha) = \int_{r\wedge S^a(\alpha)}^{s\wedge S^a(\alpha)} f^2(C(1)\alpha_2(v)) dv \le G_1^2(a)(s-r), \tag{6.11}$$

$$C^{22}(s \wedge S^a(\alpha),\alpha) - C^{22}(r \wedge S^a(\alpha),\alpha) = s \wedge S^a(\alpha) - r \wedge S^a(\alpha) \le (s-r). \tag{6.12}$$

By (6.10)-(6.12), condition (A1) of Theorem 3.1 is satisfied with $F(s,a) = \max(G_1^2(a), |h_0(1)|G_2(a), 1)s$.

Since, under assumptions of the theorem, the functions $g_1(x) = f(C(1)x)$ and $g_2(x) = h_0(1)f'(C(1)x)$ are locally Lipschitz continuous and satisfy growth condition (11.8), from Corollaries 11.1 and 11.2 it follows that conditions (A2)-(A4) of Theorem 3.1 hold. Condition (A5) of Theorem 3.1 is trivially satisfied since $X_n(0) = X(0) = 0$.

Let

$$\tilde{B}_n^1(s) = h_0(1) \sum_{i=2}^{k-1} f'\Big(\frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} u_j\Big)\Big(\frac{T_i}{n} - \frac{T_{i-1}}{n}\Big) + h_0(1) f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big)\Big(s - \frac{T_{k-1}}{n}\Big) \tag{6.13}$$

for $\frac{T_{k-1}}{n} < s \le \frac{T_k}{n}$, $k = 1, 2, ...$. It is not difficult to see that

$$\sup_{0<s\le N} |B_n^1(s) - \tilde{B}_n^1(s)| \to_P 0. \tag{6.14}$$

Indeed, by (5.26), we have that, for $\frac{T_{k-1}}{n} < s \le \frac{T_k}{n}$, $k = 1, 2, ...$,

$$\begin{aligned} |B_n^1(s) - \tilde{B}_n^1(s)| &= \Big| h_0(1) \sum_{i=2}^{k-1} f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big)\Big(\frac{T_i}{n} - \frac{T_{i-1}}{n} - \frac{\sigma_\epsilon^2}{n}\Big) \\ &\quad + h_0(1) f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big)\Big(\frac{k-1}{n}\sigma_\epsilon^2 - \frac{T_{k-1}}{n}\Big)\Big| \\ &\le |h_0(1)| \Big| \sum_{i=2}^{k-1} f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big)\Big(\frac{T_i}{n} - \frac{T_{i-1}}{n} - \frac{\sigma_\epsilon^2}{n}\Big)\Big| \\ &\quad + |h_0(1)| \Big| f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big)\Big| \Big| \frac{T_{k-1}}{n} - \frac{k-1}{n}\sigma_\epsilon^2 \Big|. \end{aligned} \tag{6.15}$$

By (4.12), from (6.15) we conclude that relation (6.14) follows if

$$\max_{1\le k\le KNn} \Big| \sum_{i=2}^{k-1} f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big)\Big(\frac{T_i}{n} - \frac{T_{i-1}}{n} - \frac{\sigma_\epsilon^2}{n}\Big)\Big| \to_P 0 \tag{6.16}$$

and

$$\max_{1\le k\le KNn} \Big| f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big)\Big| \Big| \frac{T_{k-1}}{n} - \frac{k-1}{n}\sigma_\epsilon^2 \Big| \to_P 0. \tag{6.17}$$

By Lemma 12.1 and estimate (13.12), under the assumptions of the theorem, $\eta_{tn} = f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{t-1} u_j\Big)(T_t - T_{t-1} - \sigma_\epsilon^2)$, $t \geq 2$, is a martingale-difference sequence with

$$\max_{1\leq t\leq n} E\eta_{tn}^2 \leq L_1 E\epsilon_0^4 \max_{1\leq t\leq n} E\Big(f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{t-1} u_j\Big)\Big)^2 \leq L_2$$

for some constants $L_1 > 0$ and $L_2 > 0$. Therefore, from Lemma 13.5 we conclude that (6.16) holds. In addition, from Theorem 4.2 it follows that

$$\max_{1\leq k\leq KNn} \Big|f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big)\Big| = O_P(1). \tag{6.18}$$

This, together with (12.3), implies (6.17). Consequently, (6.14) indeed holds.

By definition of $B(s,\alpha)$ and $C(s,\alpha)$ in (11.7) with $g_1(x) = f(C(1)x)$ and $g_2(x) = h_0(1)f'(C(1)x)$, we have that

$$B(s, X_n) = \Big(\int_0^s h_0(1)f'(C(1)W(v))dv, 0\Big) = (\tilde{B}^1(s), \tilde{B}^2(s)), \tag{6.19}$$

where $\tilde{B}^1(s) = \int_0^s h_0(1)f'(C(1)W(v))dv$ and $\tilde{B}^2(s) = 0$, and

$$C(s, X_n) = \begin{pmatrix} \int_0^s f^2(C(1)W(v))dv & \int_0^s f(C(1)W(v))dv \\ \int_0^s f(C(1)W(v))dv & s \end{pmatrix} = \begin{pmatrix} \tilde{C}^{11}(s,\alpha) & \tilde{C}^{12}(s,\alpha) \\ \tilde{C}^{21}(s,\alpha) & \tilde{C}^{22}(s,\alpha) \end{pmatrix}, \tag{6.20}$$

where $\tilde{C}^{11}(s) = \int_0^s f^2(C(1)W(v))dv$, $\tilde{C}^{12}(s) = \tilde{C}^{21}(s) = \int_0^s f(C(1)W(v))dv$ and $\tilde{C}^{22}(s) = s$.

By (6.13) and (6.19), for $\frac{T_{k-1}}{n} < s \leq \frac{T_k}{n}$, $k = 1, 2, ...$,

$$\begin{aligned} |\tilde{B}_n^1(s) - \tilde{B}^1(s)| &= |h_0(1)|\Big|\sum_{i=1}^{k-1}\int_{\frac{T_{i-1}}{n}}^{\frac{T_i}{n}} \Big[f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big) - f'(C(1)W(v))\Big]dv \\ &\quad + \int_{\frac{T_{k-1}}{n}}^{s} \Big[f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1} u_j\Big) - f'(C(1)W(v))\Big]dv\Big| \\ &\leq s|h_0(1)| \max_{1\leq i\leq k} \sup_{v\in[\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \Big|f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big) - f'(C(1)W(v))\Big|. \end{aligned} \tag{6.21}$$

Thus, for $\frac{T_{k-1}}{n} < N \leq \frac{T_k}{n}$, $k = 1, 2, ...$,

$$\sup_{0\leq s\leq N} \Big|\tilde{B}_n^1(s) - \tilde{B}^1(s)\Big| \leq N|h_0(1)| \max_{1\leq i\leq k} \sup_{v\in[\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \Big|f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big) - f'(C(1)W(v))\Big|. \tag{6.22}$$

By (12.1) we have

$$\begin{aligned} \max_{1\leq i\leq k} \sup_{v\in[\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \Big|f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big) - f'(C(1)W(v))\Big| &\leq \max_{1\leq i\leq k} \Big|f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big) - f'\Big(C(1)W\Big(\frac{T_{i-1}}{n}\Big)\Big)\Big| + \\ \max_{1\leq i\leq k} \sup_{v\in[\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \Big|f'\Big(C(1)W\Big(\frac{T_{i-1}}{n}\Big)\Big) - f'(C(1)W(v))\Big| &\leq \max_{1\leq i\leq k} \Big|f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big) - f'\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1} \epsilon_j\Big)\Big| + \\ \max_{1\leq i\leq k} \sup_{v\in[\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \Big|f'\Big(C(1)W\Big(\frac{T_{i-1}}{n}\Big)\Big) - f'(C(1)W(v))\Big| &\leq \max_{1\leq i\leq k} \Big|f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1} u_j\Big) - f'\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1} \epsilon_j\Big)\Big| + \\ &\quad \max_{1\leq i\leq k} \sup_{v_1,v_2\in[\frac{T_{i-1}}{n}, \frac{T_i}{n}]} \Big|f'(C(1)W(v_1)) - f'(C(1)W(v_2))\Big|. \end{aligned} \tag{6.23}$$

Using (4.6) we get

$$\max_{1\le i\le KNn}\Big|f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f'\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\Big|$$

$$= \max_{1\le i\le KNn}\Big|f'\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j+\tilde\epsilon_0-\tilde\epsilon_{i-1}\Big)-f'\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\Big|. \tag{6.24}$$

By (4.8), from (6.24) and uniform continuity of f' we obtain that

$$\max_{1\le i\le KNn}\Big|f'\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f'\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\Big|\to_P 0. \tag{6.25}$$

In addition, relation (4.15), together with uniform continuity of f' and that of the Brownian sample paths, implies

$$\max_{1\le i\le KNn}\sup_{v_1,v_2\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\Big|f'(C(1)W(v_1))-f'(C(1)W(v_2))\Big|\to_P 0, \tag{6.26}$$

By (4.12), from (6.22), (6.23), (6.25) and (6.26) we get

$$\sup_{0\le s\le N}\big|\tilde B_n^1(s)-\tilde B^1(s)\big|\to_P 0 \tag{6.27}$$

for all $N\in\mathbf{N}$. From (6.14) and (6.27) we conclude that

$$\sup_{0\le s\le N}\big|\ B_n^1(s)-\tilde B^1(s)\big|\to_P 0. \tag{6.28}$$

Consequently, condition $[sup-\beta]$ (and thus $[sup-\beta_{loc}]$) of Theorem 3.1 is satisfied.

By (6.7), (6.8) and (6.20), for $\frac{T_{k-1}}{n}<s\le\frac{T_k}{n}$, $k=1,2,...$,

$$\begin{aligned}\big|C_n^{11}(s)-\tilde C^{11}(s)\big| &= \Big|\sum_{i=1}^{k-1}\int_{\frac{T_{i-1}}{n}}^{\frac{T_i}{n}}\Big[f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f^2(C(1)W(v))\Big]dv\\ &\quad+\int_{\frac{T_{k-1}}{n}}^{s}\Big[f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1}u_j\Big)-f^2(C(1)W(v))\Big]dv\Big|\\ &\le s\max_{1\le i\le k}\sup_{v\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\Big|f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f^2(C(1)W(v))\Big|,\end{aligned} \tag{6.29}$$

$$\begin{aligned}&\big|C_n^{12}(s)-\tilde C^{12}(s)\big|=\big|C_n^{21}(s)-\tilde C^{21}(s)\big|=\\ &\Big|\sum_{i=1}^{k-1}\int_{\frac{T_{i-1}}{n}}^{\frac{T_i}{n}}\Big[f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f(C(1)W(v))\Big]dv\\ &+\int_{\frac{T_{k-1}}{n}}^{s}\Big[f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{k-1}u_j\Big)-f(C(1)W(v))\Big]dv\Big|\\ &\le s\max_{1\le i\le k}\sup_{v\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\Big|f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f(C(1)W(v))\Big|.\end{aligned} \tag{6.30}$$

Thus, for $\frac{T_{k-1}}{n}<N\le\frac{T_k}{n}$, $k=1,2,...$,

$$\sup_{0\le s\le N}\big|C_n^{11}(s)-\tilde C^{11}(s)\big|\le N\max_{1\le i\le k}\sup_{v\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\Big|f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f^2(C(1)W(v))\Big|, \tag{6.31}$$

$$\sup_{0\le s\le N}\left|C_n^{12}(s)-\tilde{C}^{12}(s)\right|\le N\max_{1\le i\le k}\sup_{v\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\left|f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f(C(1)W(v))\right|. \tag{6.32}$$

By (12.1) we have

$$\begin{aligned}
\max_{1\le i\le k}\sup_{v\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\left|f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f^2(C(1)W(v))\right| &\le \max_{1\le i\le k}\left|f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f^2\Big(C(1)W\Big(\frac{T_{i-1}}{n}\Big)\Big)\right|+\\
\max_{1\le i\le k}\sup_{v\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\left|f^2\Big(C(1)W\Big(\frac{T_{i-1}}{n}\Big)\Big)-f^2(C(1)W(v))\right| &\le \max_{1\le i\le k}\left|f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f^2\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\right|+\\
\max_{1\le i\le k}\sup_{v\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\left|f^2\Big(C(1)W\Big(\frac{T_{i-1}}{n}\Big)\Big)-f^2(C(1)W(v))\right| &\le \max_{1\le i\le k}\left|f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f^2\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\right|+\\
&\quad \max_{1\le i\le k}\sup_{v_1,v_2\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\left|f^2(C(1)W(v_1))-f^2(C(1)W(v_2))\right|. \qquad (6.33)
\end{aligned}$$

Similarly,

$$\begin{aligned}
\max_{1\le i\le k}\sup_{v\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\left|f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f(C(1)W(v))\right| &\le \max_{1\le i\le k}\left|f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f\Big(C(1)W\Big(\frac{T_{i-1}}{n}\Big)\Big)\right|+\\
\max_{1\le i\le k}\sup_{v\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\left|f\Big(C(1)W\Big(\frac{T_{i-1}}{n}\Big)\Big)-f(C(1)W(v))\right| &\le \max_{1\le i\le k}\left|f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\right|+\\
\max_{1\le i\le k}\sup_{v\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\left|f\Big(C(1)W\Big(\frac{T_{i-1}}{n}\Big)\Big)-f(C(1)W(v))\right| &\le \max_{1\le i\le k}\left|f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\right|+\\
&\quad \max_{1\le i\le k}\sup_{v_1,v_2\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\left|f(C(1)W(v_1))-f(C(1)W(v_2))\right|. \qquad (6.34)
\end{aligned}$$

By (4.6) we have

$$\begin{aligned}
&\max_{1\le i\le KNn}\left|f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f^2\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\right|\\
=&\max_{1\le i\le KNn}\left|f^2\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j+\tilde{\epsilon}_0-\tilde{\epsilon}_{i-1}\Big)-f^2\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\right|, \qquad (6.35)
\end{aligned}$$

$$\begin{aligned}
&\max_{1\le i\le KNn}\left|f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\right|\\
=&\max_{1\le i\le KNn}\left|f\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j+\tilde{\epsilon}_0-\tilde{\epsilon}_{i-1}\Big)-f\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\right|. \qquad (6.36)
\end{aligned}$$

By (4.8), from (6.35) and (6.36) and uniform continuity of f and f^2 we obtain

$$\max_{1\le i\le KNn}\left|f^2\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f^2\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\right|\to_P 0, \tag{6.37}$$

$$\max_{1\le i\le KNn}\left|f\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{i-1}u_j\Big)-f\Big(\frac{C(1)}{\sqrt{n}}\sum_{j=1}^{i-1}\epsilon_j\Big)\right|\to_P 0. \tag{6.38}$$

In addition, relation (4.15), together with uniform continuity of f and f^2 and of the Brownian sample paths implies that

$$\max_{1\le i\le KNn}\sup_{v_1,v_2\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\Big|f^2(C(1)W(v_1))-f^2(C(1)W(v_2))\Big|\to_P 0, \tag{6.39}$$

$$\max_{1\le i\le KNn}\sup_{v_1,v_2\in[\frac{T_{i-1}}{n},\frac{T_i}{n}]}\Big|f(C(1)W(v_1))-f(C(1)W(v_2))\Big|\to_P 0, \tag{6.40}$$

By (4.12), from (6.31)-(6.34) and (6.37)-(6.40) we get

$$\sup_{0\le s\le N}\big|C_n^{11}(s)-\tilde C^{11}(s)\big|\to_P 0, \tag{6.41}$$

$$\sup_{0\le s\le N}\big|C_n^{12}(s)-\tilde C^{12}(s)\big|=\sup_{0\le s\le N}\big|C_n^{21}(s)-\tilde C^{21}(s)\big|\to_P 0, \tag{6.42}$$

for all $N\in\mathbf{N}$. Relations (6.41) and (6.42), together with $C_n^{22}(s)=\tilde C^{22}(s)=s$ evidently imply that

$$\sup_{0\le s\le N}|C_n(s)-C(s,X_n)|\to_P 0,$$

for all $N\in\mathbf{N}$. Consequently, condition $[sup-\gamma]$ (and thus $[\gamma_{loc}-R_+^2]$) of Theorem 3.1 is satisfied. We therefore have $X_n\to_d X$. This, together with (4.3) and (6.4) implies, by Lemma 13.2, relation (6.2).

For $k\ge 2$, denote

$$\mathcal{I}_k=\Big|\frac{1}{\sqrt n}\sum_{t=2}^{k}f\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t-1}u_i\Big)u_t-\frac{\lambda}{n}\sum_{t=2}^{k}f'\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t-1}u_i\Big)-\frac{C(1)}{\sqrt n}\sum_{t=2}^{k}f\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t-1}u_i\Big)\epsilon_t\Big|.$$

To complete the proof, we show that, for all $N\in\mathbf{N}$,

$$\sup_{0\le r\le N}\mathcal{I}_{[nr]}\to_P 0. \tag{6.43}$$

Using (4.5) and summation by parts gives

$$\mathcal{I}_k=\Big|\frac{1}{\sqrt n}\sum_{t=2}^{k}f\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t-1}u_i\Big)(\tilde\epsilon_{t-1}-\tilde\epsilon_t)-\frac{\lambda}{n}\sum_{t=2}^{k}f'\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t-1}u_i\Big)\Big|=$$

$$\Big|-\frac{1}{\sqrt n}f\Big(\frac{1}{\sqrt n}\sum_{i=1}^{k}u_i\Big)\tilde\epsilon_k+\frac{1}{\sqrt n}\sum_{t=2}^{k}\Big(f\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t}u_i\Big)-f\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t-1}u_i\Big)\Big)\tilde\epsilon_t-\frac{\lambda}{n}\sum_{t=2}^{k}f'\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t-1}u_i\Big)\Big|.$$

Consequently, for all $N\in\mathbf{N}$,

$$\begin{aligned}\max_{1\le k\le nN}\mathcal{I}_k &\le \max_{1\le k\le nN}\Big|\frac{1}{\sqrt n}f\Big(\frac{1}{\sqrt n}\sum_{i=1}^{k}u_i\Big)\tilde\epsilon_k\Big|+\max_{1\le k\le nN}\Big|\frac{1}{n}\sum_{t=2}^{k}f'\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t-1}u_i\Big)(u_t\tilde\epsilon_t-\lambda)\Big|+\\ &\quad \max_{1\le k\le nN}\Big|\frac{1}{\sqrt n}\sum_{t=2}^{k}\Big(f\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t}u_i\Big)-f\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t-1}u_i\Big)-f'\Big(\frac{1}{\sqrt n}\sum_{i=1}^{t-1}u_i\Big)\frac{u_t}{\sqrt n}\Big)\tilde\epsilon_t\Big|\\ &= \mathcal{I}_{1n}+\mathcal{I}_{2n}+\mathcal{I}_{3n}.\end{aligned} \tag{6.44}$$

From (4.8) and property (6.18) it follows that $\mathcal{I}_{1n}\to_P 0$.

Similar to the proof of Theorem 5.2, using (5.25) and (5.26), we get that

$$\mathcal{I}_{2n} \leq \max_{1\leq k\leq nN}\Big|\frac{1}{n}\sum_{t=2}^{k} f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)(h_0(1)\epsilon_t^2 - h_0(1)\sigma_\epsilon^2)\Big| + \max_{1\leq k\leq nN}\Big|\frac{1}{n}\sum_{t=2}^{k} f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)\epsilon_t\epsilon_{t-1}^h\Big| +$$

$$\max_{1\leq k\leq nN}\Big|\frac{1}{n}\sum_{t=2}^{k} f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)(\tilde{w}_{at} - \tilde{w}_{a,t-1})\Big| + \max_{1\leq k\leq nN}\Big|\frac{1}{n}\sum_{t=2}^{k} f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)(\tilde{w}_{bt} - \tilde{w}_{b,t-1})\Big|$$

$$= \mathcal{I}_{2n}^{(1)} + \mathcal{I}_{2n}^{(2)} + \mathcal{I}_{2n}^{(3)} + \mathcal{I}_{2n}^{(4)}.$$

As in the proof of Theorem 5.2 and relation (6.14) above, we conclude, by Lemma 13.12, that $\eta_{tn}^{(1)} = f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)(\epsilon_t^2 - \sigma_\epsilon^2)$, $t \geq 2$, is a martingale-difference with

$$\max_{1\leq t\leq n} E\Big(\eta_{tn}^{(1)}\Big) \leq L_1 E\epsilon_0^4 \max_{1\leq t\leq n} E\Big(f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)\Big)^2 \leq L_2$$

for some constants $L_1 > 0$ and $L_2 > 0$.

Similarly, from Lemmas 13.12 and 13.11 it follows, by Hölder's inequality, that the martingale-difference sequence $\eta_{tn}^{(2)} = f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)\eta_t\eta_{t-1}^h$, $t \geq 2$, satisfies

$$\max_{1\leq t\leq n} E\Big(\eta_{tn}^{(2)}\Big) = E\epsilon_0^2 \max_{1\leq t\leq n} E\Big(f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)\Big)^2 \big(\eta_{t-1}^h\big)^2 \leq E\epsilon_0^2\Big[E\big(\eta_{t-1}^h\big)^4\Big]^{1/2} \max_{1\leq t\leq n}\Big[E\Big(f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)\Big)^4\Big]^{1/2} \leq L$$

for some constant $L \geq 0$. Using Theorem 13.5, we, therefore, have

$$\mathcal{I}_{2n}^{(1)} = \max_{1\leq k\leq nN}\Big|\frac{1}{n}\sum_{t=2}^{k}\eta_{tn}^{(1)}\Big| \to_P 0$$

and

$$\mathcal{I}_{2n}^{(2)} \max_{1\leq k\leq nN}\Big|\frac{1}{n}\sum_{t=2}^{k}\eta_{tn}^{(2)}\Big| \to_P 0.$$

In addition, using summation by parts and the smoothness assumptions on f, we find that (below, $S_k = \sum_{i=1}^k u_i$)

$$\mathcal{I}_{2n}^{(3)} \leq \max_{1\leq k\leq nN}\Big|\frac{1}{n} f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{k}u_i\Big)\tilde{w}_{ak}\Big| + \max_{1\leq k\leq nN}\Big|\frac{1}{n}\sum_{t=2}^{k}\Big(f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t}u_i\Big) - f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)\Big)\tilde{w}_{at}\Big| \leq$$

$$\max_{1\leq k\leq nN}\Big|\frac{1}{n} f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{k}u_i\Big)\Big| \max_{1\leq k\leq nN}\frac{1}{n}|\tilde{w}_{ak}| + N \max_{1\leq k\leq nN}\frac{1}{\sqrt{n}}|u_k\tilde{w}_{ak}| \sup_{|t|\leq \max_{0\leq k\leq nN}|S_k|/\sqrt{n}} |f''(t)|, \tag{6.45}$$

$$\mathcal{I}_{2n}^{(4)} \leq \max_{1\leq k\leq nN}\Big|\frac{1}{n} f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{k}u_i\Big)\tilde{w}_{bk}\Big| + \max_{1\leq k\leq nN}\Big|\frac{1}{n}\sum_{t=2}^{k}\Big(f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t}u_i\Big) - f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)\Big)\tilde{w}_{bt}\Big| \leq$$

$$\max_{1\leq k\leq nN}\Big|\frac{1}{n} f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{k}u_i\Big)\Big| \max_{1\leq k\leq nN}\frac{1}{n}|\tilde{w}_{bk}| + N \max_{1\leq k\leq nN}\frac{1}{\sqrt{n}}|u_k\tilde{w}_{bk}| \sup_{|t|\leq \max_{0\leq k\leq nN}|S_k|/\sqrt{n}} |f''(t)|. \tag{6.46}$$

By Lemma 13.10, $\sup_t |\tilde{w}_{at}|^3 \to_P 0$ and $\sup_t |\tilde{w}_{bt}|^3 \to_P 0$ under the assumptions of the theorem. Therefore, using Lemma 13.4 with $p = 6$ we have

$$\max_{1\leq k\leq nN} n^{-1/6}|u_k| \to_P 0, \quad \max_{1\leq k\leq nN} n^{-1/3}|\tilde{w}_{ak}| \to_P 0, \quad \max_{1\leq k\leq nN} n^{-1/3}|\tilde{w}_{bk}| \to_P 0. \tag{6.47}$$

These relations also imply that $\max_{1\le k\le nN} n^{-1/2}|u_k\tilde{w}_{ak}| \to_P 0$ and $\max_{1\le k\le nN} n^{-1/2}|u_k\tilde{w}_{bk}| \to_P 0$. From the above, together with (5.24), (6.18), (6.45) and (6.46), we conclude that $\mathcal{I}_{2n}^{(3)} \to_P 0$ and $\mathcal{I}_{2n}^{(4)} \to_P 0$.

We have, by Taylor expansion, that

$$\max_{0\le k\le nN}\Big|\frac{1}{\sqrt{n}}\sum_{t=2}^{k}\Big(f\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t}u_i\Big)-f\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)-f'\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\Big)\frac{u_t}{\sqrt{n}}\Big)\tilde{\epsilon}_t\Big|$$

$$(6.48)\qquad \le (N/2)\max_{1\le k\le nN}\frac{1}{\sqrt{n}}u_k^2|\tilde{\epsilon}_k|\sup_{|t|\le \max_{0\le k\le nN}|S_k|/\sqrt{n}}|f''(t)|.$$

By Lemmas 13.4 and 13.10, $\max_{1\le k\le nN} n^{-1/6}|\tilde{\epsilon}_k| \to_P 0$. This, together with (5.24) and the first relation in (6.47) leads to $\max_{0\le k\le nN} n^{-1/2}u_k^2|\tilde{\epsilon}_k| \to_P 0$. Consequently, by (6.48) we have $\mathcal{I}_{3n} \to_P 0$.

From (6.44) we deduce that (6.43) indeed holds. By Lemmas 13.1 and 13.3, relations (6.2) and (6.43) imply (6.1). ■

7. Asymptotics in stationary and unit root autoregression

This section shows how the martingale convergence approach provides a unified treatment of the limit theory for autoregression as in (7.1) below that includes both stationary ($\alpha = 0$) and unit root ($\alpha = 1$) cases. Let $(y_t)_{t\in\mathbf{N}}$ be a stochastic process generated in discrete time according to

$$(7.1)\qquad y_t = \alpha y_{t-1} + u_t,$$

where u_t is the linear process $u_t = C(L)\epsilon_t = \sum_{j=0}^{\infty} c_j\epsilon_{t-j}$, $C(L) = \sum_{j=0}^{\infty} c_jL^j$, $\sum_{j=1}^{\infty} jc_j^2 < \infty$, $C(1) \neq 0$, and $(\epsilon_t)_{t\in\mathbf{Z}}$ satisfy assumption (D2) with $p > 4$. The initial condition in (7.1) is set at $t = 0$ and y_0 may be a constant or a random variable. In (7.1) we can use $\alpha = 0$ to represent the stationary case without loss of generality because u_t is defined as an arbitrary linear process.

Let $\hat{\alpha} = \sum_{t=1}^{n} y_{t-1}y_t \Big/ \sum_{t=1}^{n} y_{t-1}^2$ denote the ordinary least squares (OLS) estimator of α and let $t_{\hat{\alpha}}$ be the conventional regression $t-$statistic in model (7.1) with $\alpha = 1$: $t_{\hat{\alpha}} = \Big(\sum_{t=1}^{n} y_{t-1}^2\Big)^{1/2}(\hat{\alpha}-1)/s$, where $s^2 = n^{-1}\sum_{t=1}^{n}(y_t-\hat{\alpha}y_{t-1})^2$. Further, let $\hat{\sigma}_u^2$ be a consistent estimator of $\sigma_u^2 = Eu_0^2$ and let $\hat{\omega}^2$, $\hat{\lambda}$, $\hat{\gamma}$ and $\hat{\eta}$ be, respectively, consistent nonparametric kernel estimates of the nuisance parameters $\lambda = \sum_{j=1}^{\infty} Eu_0u_j$, $\omega^2 = \sigma_\epsilon^2 C^2(1)$, $\gamma = \sigma_\epsilon^2 f_0(1)$ and $\eta = \Big(f_0^2(1) + \sum_{r=1}^{\infty} f_r^2(1)\Big)^{1/2}$, where $f_0(1) = \sum_{k=0}^{\infty} c_kc_{k+1}$ and $f_r(1) = \sum_{k=0}^{\infty} c_kc_{k+r-1}$ $r \ge 1$. Denote by Z_α and Z_t the statistics $Z_\alpha = n(\hat{\alpha}-1) - \hat{\lambda}\Big(n^{-2}\sum_{t=1}^{n} y_{t-1}^2\Big)^{-1}$ and $Z_t = \hat{\sigma}_u\hat{\omega}^{-1}t_{\hat{\alpha}} - \hat{\lambda}\Big\{\hat{\omega}\Big(n^{-2}\sum_{t=1}^{n} y_{t-1}^2\Big)^{1/2}\Big\}^{-1}$.

We prove the following result.

Theorem 7.1 *If, in model (7.1), $\alpha = 1$ and $\sum_{j=1}^{\infty} j|c_j| < \infty$, then, as $n \to \infty$,*

$$(7.2)\qquad n(\hat{\alpha}-1) \to_d \Big(\omega^2\int_0^1 W(v)dW(v) + \lambda\Big)\Big(\omega^2\int_0^1 W^2(v)dv\Big)^{-1},$$

$$(7.3)\qquad t_{\hat{\alpha}} \to_d \sigma_u^{-1}\omega^{-1}\Big(\omega^2\int_0^1 W(v)dW(v) + \lambda\Big)\Big(\int_0^1 W^2(v)dv\Big)^{-1/2},$$

where $\sigma_u^2 = Eu_0^2$, $\lambda = \sum_{j=1}^{\infty} Eu_0u_j$ and $\omega^2 = \sigma_\epsilon^2C^2(1)$. One also has the following nuisance-parameter-free limits for the test statistics Z_α and Z_t in model (7.1) with $\alpha = 1$ and $\sum_{j=1}^{\infty} j|c_j| < \infty$:

$$(7.4)\qquad Z_\alpha \to_d \Big(\int_0^1 W(v)dW(v)\Big)\Big(\int_0^1 W^2(v)dv\Big)^{-1},$$

(7.5) $$Z_t \to_d \Big(\int_0^1 W(v)dW(v)\Big)\Big(\int_0^1 W^2(v)dv\Big)^{-1/2}.$$

If, in model (7.1), $\alpha = 0$ *and* $\sum_{j=1}^{\infty} jc_j^2 < \infty$*, then, as* $n \to \infty$*,*

(7.6) $$\sqrt{n}(\hat{\alpha} - \gamma) \to_d N(0, \eta^2/\sigma_u^2),$$

(7.7) $$\frac{\hat{\sigma}_u \sqrt{n}}{\hat{\eta}}(\hat{\alpha} - \gamma) \to_d N(0,1).$$

Proof. Using the continuous mapping theorem (e.g., JS, VI.3.8) and Theorem 4.2 we get $n^{-2}\sum_{t=1}^{n} y_{t-1}^2 \to_d \omega^2 \int_0^1 W^2(v)dv$, when $\alpha = 1$, as in Phillips (1987a). Also, by Theorem 5.2, $\frac{1}{n}\sum_{t=1}^{n} y_{t-1}u_t \to_d \lambda + \omega^2 \int_0^1 W(v)dW(v)$. These relations then imply by continuous mapping that (7.2) and (7.3) hold. Relations (7.4) and (7.5) are consequences of (7.2) and (7.3). Relations (7.6) and (7.7) follow from Theorem 4.4, the consistency of $\hat{\eta}$, and the fact that $n^{-1}\sum_{t=1}^{n} u_{t-1}^2 \to_p \sigma_u^2$ by the law of large numbers. ■

Remark 7.1 *The martingale convergence approach provides a unifying principle for proving the limit theory in the stationary and unit root cases in the above result. In particular, in the martingale-difference error case (i.e. when Assumption D1 holds and* $u_t = \varepsilon_t$*, allowing for* $\alpha = 1$ *or* $|\alpha| < 1$*) the construction by which the martingale convergence approach is applied is the same in both cases. Thus, in the stationary case we use the construction (4.16) above and in the unit root case we have essentially the same construction in (5.2). In the former case, the numerator satisfies a central limit theorem, while in the latter case we have weak convergence to a stochastic integral. This difference makes a unification of the limit theory impossible in terms of existing approaches which rely on central limit arguments in the stationary case and special weak convergence arguments in the unit root case. However, the martingale convergence approach readily accommodates both results and, at the same time, also allows for the difference in the rates of convergence. In effect, in both the stationary and unit root cases, we have convergence of a discrete time martingale to a continuous martingale, thereby unifying the limit theory for autoregression. Section 9 makes this formulation explicit.*

8. Useful multivariate extensions

The present section shows how to skip the Skorohod embedding at the beginning of the proofs, which is used above to convert discrete time martingales and semimartingales to continuous versions (e.g. in (4.2), (4.10), (5.3) and (6.4)) and simplify some of the arguments. In fact, we may work directly by treating the discrete time processes as discontinuous processes and seek to verify conditions for martingale and semimartingale convergence that involve the predictable measures of jumps for the discontinuous processes. This may be accomplished by using suitable additional conditions beyond those we have already employed in Theorems 3.1 and 3.2. Dealing with these additional conditions is not problematic, and the increase in the technical difficulty is justified in view of the wide range of applications covered by these more general results. The extensions include results on convergence to multivariate stochastic integrals and a precise formulation of the unification theorem for stationary and nonstationary autoregression. To simplify presentation of the results, we treat the bivariate case here and extensions to general multivariate cases follow in the same fashion.

We start with the following martingale convergence result, which provides a limit theory for multivariate stochastic integrals and enables later extension to the case of general linear processes.

Theorem 8.1 *Let* $\{(\epsilon_t, \eta_t)\}_{t=0}^{\infty}$ *be a sequence of i.i.d. mean-zero random vectors such that* $E\epsilon_0^2 = \sigma_\epsilon^2$, $E\eta_0^2 = \sigma_\eta^2$, $E\epsilon_0\eta_0 = \sigma_{\epsilon\eta}$, $E|\epsilon_0|^p < \infty$ *and* $E|\eta_0|^p < \infty$ *for some* $p > 4$*. Let* $(W, V) = \big((W(s), V(s)), s \geq 0\big)$ *be bivariate Brownian motion with covariance matrix*

$$\begin{pmatrix} \sigma_\epsilon^2 & \sigma_{\epsilon\eta} \\ \sigma_{\epsilon\eta} & \sigma_\eta^2 \end{pmatrix}.$$

Then

$$\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}\epsilon_i\right)\eta_t \to_d \int_0^r W(v)dV(v). \tag{8.1}$$

Proof. For $n \geq 1$, let $X_n = (X_n(s), s \geq 0)$ and $X = (X(s), s \geq 0)$ be the vector martingales

$$X_n(s) = \left(\frac{1}{n}\sum_{t=2}^{[ns]}\Big(\sum_{i=1}^{t-1}\epsilon_i\Big)\eta_t, \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\epsilon_t, \frac{1}{\sqrt{n}}\sum_{t=1}^{[ns]}\eta_t\right)$$

and

$$X(s) = \Big(\int_0^s W(v)dV(v), W(s), V(s)\Big) = (X^1(s), X^2(s), X^3(s)).$$

Let $B'_n = (B'_n(s), s \geq 0)$ denote the first characteristic without truncation of X_n, let $\tilde{C}\,'_n = (\tilde{C}\,'_n(s), s \geq 0)$ stand for its modified second characteristic without truncation and let $\nu_n = (\nu_n(ds, dx))$ denote its predictable measure of jumps (see JS, Ch. II, §2 and IX.3.25). The process B'_n is identically zero so $B'_n(s) = (0, 0, 0) \in \mathbf{R}^3$, $s \geq 0$. For the modified second characteristic without truncation of X_n we have $\tilde{C}\,'_n(s) = (\tilde{C}_n^{ij}(s))_{1\leq i,j\leq 3}$, where

$$\tilde{C}_n^{11}(s) = \frac{\sigma_\eta^2}{n^2}\sum_{t=2}^{[ns]}\Big(\sum_{i=1}^{t-1}\epsilon_i\Big)^2,$$

$$\tilde{C}_n^{12}(s) = \tilde{C}_n^{21}(s) = \frac{\sigma_{\epsilon\eta}}{n^{3/2}}\sum_{t=2}^{[ns]}\Big(\sum_{i=1}^{t-1}\epsilon_i\Big),$$

$$\tilde{C}_n^{13}(s) = \tilde{C}_n^{31}(s) = \frac{\sigma_\eta^2}{n^{3/2}}\sum_{t=2}^{[ns]}\Big(\sum_{i=1}^{t-1}\epsilon_i\Big),$$

$$\tilde{C}_n^{22}(s) = \frac{\sigma_\epsilon^2[ns]}{n},$$

$$\tilde{C}_n^{23}(s) = \tilde{C}_n^{32}(s) = \frac{\sigma_{\epsilon\eta}[ns]}{n},$$

$$\tilde{C}_n^{33}(s) = \frac{\sigma_\eta^2[ns]}{n}.$$

For an element $\alpha = (\alpha(s), s \geq 0)$, $\alpha(s) = (\alpha_1(s), \alpha_2(s), \alpha_3(s))$ of the Skorohod space $\mathbb{D}(\mathbf{R}^3)$ and for a Borel subset Γ of $\mathbf{R}^3$, let $B(s, \alpha) = (0, 0, 0)$,

$$C(s,\alpha) = \begin{pmatrix} \sigma_\eta \int_0^s \alpha_2^2(v)dv & \sigma_{\epsilon\eta}\int_0^s \alpha_2(v)dv & \sigma_\eta^2 \int_0^s \alpha_2(v)dv \\ \sigma_{\epsilon\eta}\int_0^s \alpha_2(v)dv & \sigma_\epsilon^2 s & \sigma_{\epsilon\eta}s \\ \sigma_\eta^2\int_0^s \alpha_2(v)dv & \sigma_{\epsilon\eta}s & \sigma_\eta^2 s \end{pmatrix}, \tag{8.2}$$

and $\nu([0, s], \Gamma)(\alpha) = 0$. Further, let $B(\alpha) = (B(s, \alpha), s \geq 0)$, $C(\alpha) = (C(s, \alpha), s \geq 0)$ and $\nu(\alpha) = (\nu(ds, dx)(\alpha))$. The process X is a solution to the stochastic differential equation

$$\begin{aligned} dX^1(s) &= X^2(s)dV(s); \\ dX^2(s) &= dW(s); \\ dX^3(s) &= dV(s), \end{aligned} \tag{8.3}$$

or, equivalently, to stochastic differential equation (11.1) with $d = 3$ and $m = 2$ and functions $b : \mathbf{R}^3 \to \mathbf{R}^3$ and $\sigma : \mathbf{R}^3 \to \mathbf{R}^{3\times 2}$ given by $b(x_1, x_2, x_3) = (0, 0, 0)$ and

$$\sigma(x_1, x_2, x_3) = \begin{pmatrix} \sigma_\eta x_2 & 0 \\ \sigma_{\epsilon\eta}/\sigma_\eta & \sqrt{\sigma_\epsilon^2\sigma_\eta^2 - \sigma_{\epsilon\eta}^2}\Big/\sigma_\eta \\ \sigma_\eta & 0 \end{pmatrix}. \tag{8.4}$$

According to (11.2), the predictable characteristics of X are $B(X)$, $C(X)$ and $\nu(X)$, with B, C and ν defined as above (so that the first and the third predictable characteristics of X are identically zero, i.e., $B = (0, 0, 0) \in \mathbf{R}^3$ and $\nu = 0$). Since X is continuous, its predictable triplet without truncation is the same.

For $a \geq 0$ and an element $\alpha = (\alpha(s), s \geq 0)$ of the Skorohod space $\mathbb{D}(\mathbf{R}^3_+)$, define, similar to (3.1) and as in IX.3.38 of JS,

$$\begin{aligned} S^a(\alpha) &= \inf(s : |\alpha(s)| \geq a \text{ or } |\alpha(s-)| \geq a), \\ S^a_n &= \inf(s : |X_n(s)| \geq a \text{ or } |X_n(s-)| \geq a), \end{aligned} \tag{8.5}$$

where $\alpha(s-)$ and $X_n(s-)$ denote, respectively, the left-hand limits of α and X_n at s. Let $\mathbb{C}_1(\mathbf{R}^3)$ denote the set of continuous bounded functions $g : \mathbf{R}^3 \to \mathbf{R}$ which are equal to zero in a neighborhood of zero. By Theorem IX.3.48 of JS (see also Remark IX.3.40, Theorem III.2.40 and Lemma IX.4.4 in JS and also the proof of Theorem 2.1 in Coffman, Puhalskii and Reiman, 1998), in order to prove that $X_n \to_d X$, it suffices to check that the following conditions hold in addition to conditions (A1)-(A5) of Theorem 3.1 :

(A6a) $[\boldsymbol{\delta}_{\mathbf{loc}} - \mathbf{R}_+]$ $\int_0^{s\wedge S_a(\alpha)} \int_{\mathbf{R}^3} g(x)\nu_n(dw, dx) \to_P 0$ for all $s > 0$, $a > 0$ and $g \in \mathbb{C}_1(\mathbf{R}^3)$.

$[\mathbf{sup} - \boldsymbol{\beta}'_{\mathbf{loc}}]$ $\sup_{0<s\leq N} |B'_n(s \wedge S^a_n) - B(s \wedge S^a, X_n)| \to_P 0$ for all $N \in \mathbf{N}$ and all $a > 0$.

$[\boldsymbol{\gamma}'_{\mathbf{loc}} - \mathbf{R}_+]$ $\tilde{C}\,'_n(s \wedge S^a_n) - C(s \wedge S^a, X_n) \to_P 0$ for all $s > 0$ and $a > 0$.

(A7) $\lim_{b\to\infty} \overline{\lim}_{n\to\infty} P\Big(\int_0^{s\wedge S^a_n} \int_{\mathbf{R}^3} |x|^2 I(|x| > b)\nu_n(dw, dx) > \epsilon\Big) = 0$ for all $s > 0$, $a > 0$ and $\epsilon > 0$.

The following is a sufficient condition for $[\boldsymbol{\gamma}'_{\mathbf{loc}} - \mathbf{R}_+]$ in (A6a):

$[\mathbf{sup} - \boldsymbol{\gamma}']$ $\sup_{0<s\leq N} |\tilde{C}\,'_n(s) - C(s, X_n)| \to_P 0$ for all $N \in \mathbf{N}$.

In addition, from the definition of the class $\mathbb{C}_1(\mathbf{R}^3)$ and Lemma 5.5.1 in Liptser and Shiryaev (1989) it follows in a similar way to the proof of Theorem 2.1 in Coffman et. al. that the following is a sufficient condition for $[\boldsymbol{\delta}_{\mathbf{loc}} - \mathbf{R}_+]$:

$[\mathbf{sup} - \boldsymbol{\Delta}]$ $\sup_{0<s\leq N} |\Delta X_n(s)| \to_P 0$ for all $N \in \mathbf{N}$, where $\Delta X_n(s) = X_n(s) - X_n(s-)$.

Note that since X is continuous, in the corresponding results in JS, $\nu = 0$, $B' = B$ and $\tilde{C}\,' = C$.

Conditions (A1)-(A5) of Theorem 3.1 in the present context can be verified in complete similarity to the proof of Theorem 5.1. In particular, conditions (A2) and (A3) follow from the straightforward extension of Corollary 11.1 to the case of a three-dimensional homogenous diffusion driven by two Brownian motions.

Condition $[sup - \beta']$ (and thus $[sup - \beta'_{loc}]$) is trivially satisfied since $B'_n(s) = 0$, $s \geq 0$, and $B_n(s, X_n) = 0$, $s \geq 0$.

From formula (8.2) we have that $C_n(s, X_n) = (\tilde{\tilde{C}}{}^{ij}_n(s))_{1\leq i,j\leq 3}$, where

$$\tilde{\tilde{C}}{}^{11}_n(s) = \frac{\sigma^2_\eta}{n^2} \sum_{t=2}^{[ns]} \Big(\sum_{i=1}^{t-1} \epsilon_i\Big)^2 + \frac{\sigma^2_\eta}{n^2} \Big(\sum_{i=1}^{[ns]} \epsilon_i\Big)^2 (ns - [ns]) = \tilde{C}^{11}_n(s) + \frac{\sigma^2_\eta}{n^2} \Big(\sum_{i=1}^{[ns]} \epsilon_i\Big)^2 (ns - [ns]),$$

$$\tilde{\tilde{C}}_n^{12}(s) = \tilde{\tilde{C}}_n^{21}(s) = \frac{\sigma_{\epsilon\eta}}{n^{3/2}} \sum_{t=2}^{[ns]} \Big(\sum_{i=1}^{t-1} \epsilon_i \Big) + \frac{\sigma_{\epsilon\eta}}{n^{3/2}} \Big(\sum_{i=1}^{[ns]} \epsilon_i \Big)(ns - [ns]) = \tilde{C}_n^{12} + \frac{\sigma_{\epsilon\eta}}{n^{3/2}} \Big(\sum_{i=1}^{[ns]} \epsilon_i \Big)(ns - [ns]),$$

$$\tilde{\tilde{C}}_n^{13}(s) = \tilde{\tilde{C}}_n^{31}(s) = \frac{\sigma_\eta^2}{n^{3/2}} \sum_{t=2}^{[ns]} \Big(\sum_{i=1}^{t-1} \epsilon_i \Big) + \frac{\sigma_\eta^2}{n^{3/2}} \Big(\sum_{i=1}^{[ns]} \epsilon_i \Big)(ns - [ns]) = \tilde{C}_n^{13} + \frac{\sigma_\eta^2}{n^{3/2}} \Big(\sum_{i=1}^{[ns]} \epsilon_i \Big)(ns - [ns]),$$

$$\tilde{\tilde{C}}_n^{22}(s) = \sigma_\epsilon^2 s = \tilde{C}_n^{22} + \sigma_\epsilon^2 \frac{ns - [ns]}{n},$$

$$\tilde{\tilde{C}}_n^{23}(s) = \tilde{\tilde{C}}_n^{32}(s) = \sigma_{\epsilon\eta} s = \tilde{C}_n^{23} + \sigma_{\epsilon\eta} \frac{ns - [ns]}{n},$$

$$\tilde{\tilde{C}}_n^{33}(s) = \sigma_\eta^2 s = \tilde{C}_n^{33} + \sigma_\eta^2 \frac{ns - [ns]}{n}.$$

Since, by Lemma 13.5, $n^{-1} \max_{1\le k\le Nn} \Big| \sum_{i=1}^k \epsilon_i \Big| \to_P 0$ for all $N \in \mathbf{N}$, we thus have

$$\sup_{0<s\le N} \big|\tilde{C}_n^{11}(s) - \tilde{\tilde{C}}_n^{11}(s)\big| \le \max_{0<k\le nN} \Big| \frac{\sigma_\eta^2}{n^2} \Big(\sum_{i=1}^k \epsilon_i \Big)^2 \Big| \to_P 0,$$

$$\sup_{0<s\le N} \big|\tilde{C}_n^{12}(s) - \tilde{\tilde{C}}_n^{12}(s)\big| = \sup_{0<s\le N} \big|\tilde{C}_n^{21}(s) - \tilde{\tilde{C}}_n^{21}(s)\big| \le \max_{0<k\le nN} \Big| \frac{\sigma_{\epsilon\eta}}{n^{3/2}} \Big(\sum_{i=1}^k \epsilon_i \Big) \Big| \to_P 0,$$

$$\sup_{0<s\le N} \big|\tilde{C}_n^{13}(s) - \tilde{\tilde{C}}_n^{13}(s)\big| = \sup_{0<s\le N} \big|\tilde{C}_n^{31}(s) - \tilde{\tilde{C}}_n^{31}(s)\big| \le \max_{0<k\le nN} \Big| \frac{\sigma_\eta^2}{n^{3/2}} \Big(\sum_{i=1}^k \epsilon_i \Big) \Big| \to_P 0$$

for all $N \in \mathbf{N}$. In addition, evidently, $\sup_{0<s\le N} \big|\tilde{C}_n^{22}(s) - \tilde{\tilde{C}}_n^{22}(s)\big| \le \sigma_\epsilon^2/n \to_P 0$, $\sup_{0<s\le N} \big|\tilde{C}_n^{23}(s) - \tilde{\tilde{C}}_n^{23}(s)\big| = \sup_{0<s\le N} \big|\tilde{C}_n^{32}(s) - \tilde{\tilde{C}}_n^{32}(s)\big| \le \sigma_{\epsilon\eta}/n \to_P 0$ and $\sup_{0<s\le N} \Big|\tilde{C}_n^{33}(s) - \tilde{\tilde{C}}_n^{33}(s)\Big| \le \sigma_\eta^2/n \to_P 0$ for all $N \in \mathbf{N}$. The above obviously implies that $\sup_{0<s\le N} |\tilde{C}'_n(s) - C(s, X_n)| \to_P 0$ for all $N \in \mathbf{N}$ and thus condition $[sup - \gamma']$ (and condition $[\gamma'_{loc} - \mathbf{R}_+]$) is satisfied.

For all $N \in \mathbf{N}$, we have

$$\sup_{0\le s\le N} |\Delta X_n(s)| \le \max_{0\le k\le nN} \frac{1}{\sqrt{n}} \Big| \sum_{i=1}^k \epsilon_i \Big| \max_{0\le k\le nN} \frac{1}{\sqrt{n}} |\epsilon_k| + \max_{0\le k\le nN} \frac{1}{\sqrt{n}} |\epsilon_k| + \max_{0\le k\le nN} \frac{1}{\sqrt{n}} |\eta_k|.$$

By Theorem 4.1, the sequence $\max_{0\le k\le nN} \frac{1}{\sqrt{n}} \Big| \sum_{i=1}^k \epsilon_i \Big|$ is bounded in probability and $\max_{0\le k\le nN} \frac{1}{\sqrt{n}} \Big| \sum_{i=1}^k \epsilon_i \Big| = O_P(1)$. In addition, by Lemma 13.4, $\max_{0\le k\le nN} \frac{1}{\sqrt{n}} |\epsilon_k| \to_P 0$ and $\max_{0\le k\le nN} \frac{1}{\sqrt{n}} |\eta_k| \to_P 0$. Using the above, we therefore find that $\sup_{0\le s\le N} |\Delta X_n(s)| \to_P 0$ for all $N \in \mathbf{N}$. Thus, condition $[sup - \Delta]$ holds and $[\delta_{loc} - \mathbf{R}_+]$ holds in consequence.

Finally, we demonstrate that (A7) holds. It is not difficult to see that

$$\begin{aligned} E \int_0^{s\wedge S_n^a} \int_{\mathbf{R}^3} |x|^2 I(|x| &> b)\nu_n(dw, dx) \le E \int_0^s \int_{\mathbf{R}^3} |x|^2 I(|x| > b)\nu_n(dw, dx) \le \\ \frac{1}{b^2} E \int_0^s \int_{\mathbf{R}^3} |x|^4 \nu_n(dw, dx) &\le \frac{3}{b^2} E \int_0^s \int_{x=(x_1,x_2,x_3)\in\mathbf{R}^3} (x_1^4 + x_2^4 + x_3^4)\nu_n(dw, dx). \end{aligned} \tag{8.6}$$

Continuing, we have

$$E\int_0^s \int_{x=(x_1,x_2,x_3)\in\mathbf{R}^3} (x_1^4+x_2^4+x_3^4)\nu_n(dw,dx) = \frac{1}{n^4}\sum_{t=2}^{[ns]} E\Big(\sum_{i=1}^{t-1}\epsilon_i\Big)^4 E\eta_t^4 + \tag{8.7}$$

$$\frac{1}{n^2}\sum_{t=2}^{[ns]} E\epsilon_t^4 + \frac{1}{n^2}\sum_{t=2}^{[ns]} E\eta_t^4 = \frac{E\eta_0^4}{n^4}\sum_{t=2}^{[ns]} E\Big(\sum_{i=1}^{t-1}\epsilon_i\Big)^4 + 2\frac{E\epsilon_0^4[ns]}{n^2},$$

and, using inequality (13.13) in Appendix 11, we find that

$$\frac{E\eta_0^4}{n^4}\sum_{t=2}^{[ns]} E\Big(\sum_{i=1}^{t-1}\epsilon_i\Big)^4 \le \frac{K(E\epsilon_0^4)^2}{n^2}\sum_{t=2}^{[ns]} t^2 \le K(E\epsilon_0^4)^2/n \to 0$$

for all $s>0$. Evidently, $[ns]/n^2 \to 0$ for all $s>0$, and from (8.6) and (8.7) we deduce that

$$E\int_0^{s\wedge S_n^a}\int_{\mathbf{R}^3} |x|^2 I(|x|>b)\nu_n(dw,dx) \to 0$$

for all $a,b,s>0$. By Chebyshev's inequality, this evidently implies that condition (A7) holds.

Consequently, conditions (A1)-(A5) of Theorem 3.1, together with conditions (A6a) and (A7) above are satisfied for X_n and X. The convergence (8.1) therefore holds as required. ■

In complete similarity to the proof of relation (8.1) and to Theorem 6.1, we may deduce, with the help of straightforward extensions of Corollary 11.1, that the following analogues of (8.1) and Theorem 6.1 hold in the present context.

Theorem 8.2 *Let $f:\mathbf{R}\to\mathbf{R}$ be a twice continuously differentiable function such that f' satisfies the growth condition $|f'(x)| \le K(1+|x|^\alpha)$ for some constants $K>0$ and $\alpha>0$ and all $x\in\mathbf{R}$. Suppose that $\{(\epsilon_t,\eta_t)\}_{t=0}^\infty$ is a sequence of i.i.d. mean-zero random vectors such that $E\epsilon_0^2=\sigma_\epsilon^2$, $E\eta_0^2=\sigma_\eta^2$, $E\epsilon_0\eta_0=\sigma_{\epsilon\eta}$, $E|\epsilon_0|^p<\infty$ and $E|\eta_0|^p<\infty$ for some with $p\ge\max(6,4\alpha)$. Then*

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{[nr]} f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}\epsilon_i\right)\eta_t \to_d \int_0^r f(W(v))dV(v). \tag{8.8}$$

Further, using the Phillips-Solo device as in the proof of Theorems 5.2 and 6.1, we obtain the following generalizations of relations (8.1) and (8.8) to the case of linear processes.

Theorem 8.3 *Suppose that $w_t=(u_t,v_t)'$ is the linear process $w_t=G(L)\epsilon_t=\sum_{j=0}^\infty G_j\epsilon_{t-j}$, with $G(L)=\sum_{j=0}^\infty G_jL^j$, $\sum_{j=1}^\infty j||G_j||<\infty$, $G(1)$ of full rank, and $\{\epsilon_t\}_{t=0}^\infty$ a sequence of i.i.d. mean-zero random vectors such that $E\epsilon_0\epsilon_0'=\Sigma_\epsilon>0$ and $\max_i E|\epsilon_{i0}|^p<\infty$ for some $p>4$. Then*

$$\frac{1}{n}\sum_{t=2}^{[nr]}\left(\sum_{i=1}^{t-1}u_i\right)v_t \to_d r\lambda_{uv} + \int_0^r W(v)dV(v), \tag{8.9}$$

where $(W,V)=\big((W(s),V(s)),s\ge0\big)$ is bivariate Brownian motion with covariance matrix $\Omega=G(1)\Sigma G(1)$ and $\lambda_{uv}=\sum_{j=1}^\infty Eu_0v_j$.

Further, if $f:R\to R$ is a twice continuously differentiable function such that f' satisfies the growth condition $|f'(x)|\le K(1+|x|^\alpha)$ for some constants $K>0$ and $\alpha>0$ and all $x\in R$, and if $p\ge\max(6,4\alpha)$, then

$$\frac{1}{\sqrt{n}}\sum_{t=2}^{[nr]} f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{t-1}u_i\right)v_t \to_d \lambda_{uv}\int_0^r f'(W(v))dv + \int_0^r f(W(v))dV(v). \tag{8.10}$$

Remark 8.1 *Using the approach developed in the present section together with the Phillips-Solo device as in the proof of Theorems 5.2 and 6.1, one can also obtain limit results for sample covariances of nonlinear functions of unstandardized integrated processes and martingale differences. In particular, our approach provides an alternative way of proving the following analogue of Theorems 8.1-8.3 for the case of integrable functions, which was first given by Park and Phillips (1999). Suppose that* (ϵ_t) *satisfy assumption (D1) with* $p > 4$ *and* (u_t) *is a linear process* $u_t = C(L)\eta_t = \sum_{j=0}^{\infty} c_j \eta_{t-j}$ *with* $C(1) \neq 0$ *and* $\sum_{j=0}^{\infty} j|c_j| < \infty$ *generated by a sequence of i.i.d. mean-zero random variables* (η_t) *independent of* (ϵ_t)*. Assume that* $E|\eta_0|^q < \infty$ *for some* $q > 4$ *and the distribution of* η_0 *is absolutely continuous with respect to Lebesgue measure and has characteristic function* $\phi(t)$ *satisfying* $\lim_{t\to\infty} t^r \phi(t) = 0$ *for some* $r > 0$*. Let a function* $f : R \to R$ *be such that* f^2 *is integrable and satisfies the Lipschitz condition* $|f^2(x) - f^2(y)| \leq C|x-y|^k$ *over its support for some constants* $C > 0$ *and* $k > 6/(q-2)$*. Then*

$$\frac{1}{n^{1/4}} \sum_{t=1}^{[nr]} f\Big(\sum_{i=1}^{t} u_i\Big)\epsilon_t \to_d \Big(L(r,0)\int_{-\infty}^{\infty} f^2(s)ds\Big)^{1/2} W(r), \tag{8.11}$$

where $L(r,0) = \lim_{a\to 0+} \frac{1}{2a}\int_0^r I(|V(s)| \leq a)ds$ *is the local time at the origin over the interval* $[0,r]$ *of a Brownian motion* V *which is independent of the Brownian motion* W*. For a fixed* r*, the limit in (8.11) is mixed normal with a mixing variate given by* L*.*

Furthermore, using the martingale convergence approach as in this paper, one can easily obtain analogues of relation (8.11) for the case of functions $f : R \times \Pi \to R$ *indexed by some parameter* π *from a compact set* Π*, as in Park and Phillips (2001).*

9. Unification of the limit theory of autoregression

The present section demonstrates how the martingale convergence approach developed in this paper provides a unified formulation of the limit theory for first order autoregression, including stationary, unit root, local to unity and (together with the conventional martingale convergence theorem) explosive settings.

Specializing (7.1), we consider here the autoregression

$$y_t = \alpha y_{t-1} + \epsilon_t, \quad t = 1, ..., n \tag{9.1}$$

with martingale-difference errors ϵ_t that satisfy assumption (D1) with $p > 4$. As in (7.1), the initial condition in (9.1) is set at $t = 0$ and y_0 may be any $O_p(1)$ random variable, including a constant. We treat the stationary $|\alpha| < 1$, unit root $\alpha = 1$, local to unity and explosive cases together in what follows and show how the limit theory for all these cases may be formulated in a unified manner within the martingale convergence framework.

We start with the stationary and unit root cases. For $r \in (0,1]$, define the recursive least squares estimator $\hat{\alpha}_r = \sum_{t=1}^{[nr]} y_{t-1}y_t \Big/ \sum_{t=1}^{[nr]} y_{t-1}^2$, and write

$$\left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma_\epsilon^2}\right)^{1/2} (\hat{\alpha}_r - \alpha) = \frac{\sum_{t=1}^{[nr]} y_{t-1}\epsilon_t}{\left(\sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2\right)^{1/2}} = \frac{X_n(r)}{\left(\tilde{C}'_n(r)\right)^{1/2}}, \tag{9.2}$$

where $X_n(r)$ is the martingale given by

$$X_n(r) = \begin{cases} \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} y_{t-1}\epsilon_t & |\alpha| < 1 \\ \frac{1}{n} \sum_{t=1}^{[nr]} y_{t-1}\epsilon_t & \alpha = 1 \end{cases}, \tag{9.3}$$

and $\tilde{C}'_n = (\tilde{C}'_n(s), s \geq 0)$ is the modified second characteristic without truncation of X_n (see JS, Ch. II, §2 and IX.3.25):

$$\tilde{C}'_n(r) = \begin{cases} \frac{1}{n} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2 & |\alpha| < 1 \\ \frac{1}{n^2} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2 & \alpha = 1 \end{cases}. \tag{9.4}$$

By virtue of Remark 4.1 and Theorem 8.1 we have

$$X_n(r) \to_d X(r) = \begin{cases} \sigma_\alpha \sigma_\epsilon W(r) & |\alpha| < 1 \\ \sigma_\epsilon^2 \int_0^r W(v) dW(v) & \alpha = 1 \end{cases}, \tag{9.5}$$

and

$$\tilde{C}_n'(r) \to_d C(r) = \begin{cases} \sigma_\alpha^2 \sigma_\epsilon^2 r & |\alpha| < 1 \\ \sigma_\epsilon^4 \int_0^r W(v)^2 dv & \alpha = 1 \end{cases}, \tag{9.6}$$

where $C = (C(s), s \geq 0)$ is the second predictable characteristic of the continuous martingale X and $\sigma_\alpha^2 = 1/(1-\alpha^2)$. Thus,

$$\begin{aligned} \left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma_\epsilon^2} \right)^{1/2} (\hat{\alpha}_r - \alpha) &= \frac{X_n(r)}{\left(\tilde{C}'_n(r)\right)^{1/2}} \to_d \frac{X(r)}{(C(r))^{1/2}} \\ &= \begin{cases} \frac{1}{r^{1/2}} W(r) & |\alpha| < 1 \\ \frac{\int_0^r W(v) dW(v)}{\left(\int_0^r W(v)^2 dv\right)^{1/2}} & \alpha = 1 \end{cases} \\ &=_d \begin{cases} N(0,1) & |\alpha| < 1 \\ \frac{\int_0^1 W(v) dW(v)}{\left(\int_0^1 W(v)^2 dv\right)^{1/2}} & \alpha = 1 \end{cases}, \end{aligned} \tag{9.7}$$

which unifies the limit theory for the stationary and unit root autoregression.

Defining the error variance estimator $s_r^2 = [nr]^{-1} \sum_{t=1}^{[nr]} (y_t - \hat{\alpha}_r y_{t-1})^2$ and noting that $s_r^2 \to_p \sigma_\epsilon^2$ for $r > 0$, we have the corresponding limit theory for the recursive $t-$ statistic

$$\begin{aligned} t_{\hat{\alpha}}(r) &= \left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{s_r^2} \right)^{1/2} (\hat{\alpha}_r - \alpha) = \frac{\sum_{t=1}^{[nr]} y_{t-1}\epsilon_t}{\left(\sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2\right)^{1/2}} \frac{\sigma_\epsilon}{s_r} = \frac{X_n(r)}{\left(\tilde{C}'_n(r)\right)^{1/2}} \frac{\sigma_\epsilon}{s_r} \\ &\to_d \begin{cases} N(0,1) & |\alpha| < 1 \\ \frac{\int_0^1 W(v) dW(v)}{\left(\int_0^1 W(v)^2 dv\right)^{1/2}} & \alpha = 1 \end{cases}. \end{aligned}$$

The theory also extends to cases where α lies in the neighborhood of unity. In complete similarity to the proof of Theorem 8.1 and to derivations above in this section, one can show that, for $\alpha = 1 + \frac{c}{n}$, (9.2) - (9.4) hold with the same normalization as in the unit root case, but in place of (9.5) and (9.6) one now has

$$X_n(r) \to_d X(r) = \sigma_\epsilon^2 \int_0^r J_c(v) dW(v), \quad \alpha = 1 + \tfrac{c}{n}, \tag{9.8}$$

$$\tilde{C}'_n(r) \to_d C(r) = \sigma_\epsilon^4 \int_0^r J_c(v)^2 dv, \quad \alpha = 1 + \tfrac{c}{n}, \tag{9.9}$$

where $J_c(v) = \int_0^v e^{c(v-s)} dW(s)$ is a linear diffusion (Phillips, 1987b). We then have

$$\begin{aligned} \left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma_\epsilon^2} \right)^{1/2} (\hat{\alpha}_r - \alpha) &= \frac{X_n(r)}{\left(\tilde{C}'_n(r)\right)^{1/2}} \to_d \frac{X(r)}{(C(r))^{1/2}} \\ &=_d \frac{\int_0^1 J_c(v) dW(v)}{\left(\int_0^1 J_c(v)^2 dv\right)^{1/2}}. \end{aligned}$$

Further, when there are moderate deviations from unity of the form $\alpha = 1 + \frac{c}{n^b}$ for some $b \in (0,1)$ and $c < 0$ (as in Phillips and Magdalinos, 2004, and Giraitis and Phillips, 2004), (9.2) continues to hold but with

$$X_n(r) = \frac{1}{n^{\frac{(1+b)}{2}}} \sum_{t=1}^{[nr]} y_{t-1}\epsilon_t, \quad \alpha = 1 + \tfrac{c}{n^b}, \quad c < 0, \ b \in (0,1),$$

and $\tilde{C}_n'(r) = \frac{1}{n^{1+b}} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2$. Then, $X_n(r) \to_d X(r) =_d N\left(0, \frac{\sigma_\epsilon^2}{-2c} r\right)$ and $\tilde{C}_n'(r) \to_p C(r) = \frac{\sigma_\epsilon^2}{-2c} r$. Then, (9.7) again holds with the limit process being $X(r) / (C(r))^{1/2} =_d N(0,1)$.

Next consider the explosive autoregressive case where $\alpha > 1$. In this case, (9.2) applies with $X_n(r) = \frac{1}{\alpha^{[nr]}} \sum_{t=1}^{[nr]} y_{t-1}\epsilon_t$ and $\tilde{C}_n'(r) = \alpha^{-2[nr]} \sum_{t=1}^{[nr]} y_{t-1}^2 \sigma_\epsilon^2$. By the martingale convergence theorem, $\alpha^{-t} y_t \to_{a.s.} Y_\alpha$, where $Y_\alpha = \sum_{s=1}^{\infty} \alpha^{-s}\epsilon_s + y_0$, and, correspondingly, $\tilde{C}:_n'(r) \to_{a.s.} C(r) = Y_\alpha^2 \frac{\sigma_\epsilon^2}{\alpha^2-1}$. By further application of the martingale convergence theorem we find that

$$X_n(r) = \frac{1}{\alpha^{[nr]}} \sum_{t=1}^{[nr]} y_{t-1}\epsilon_t = \sum_{t=1}^{[nr]} \frac{y_{t-1}}{\alpha^{t-1}} \frac{\epsilon_t}{\alpha^{[nr]-(t-1)}} \to_{a.s.} Y_\alpha Z_\alpha, \tag{9.10}$$

with $Z_\alpha = \sum_{s=1}^{\infty} \alpha^{-s}\epsilon_s'$ where (ϵ_s') is an i.i.d. sequence that is distributionally equivalent to (ϵ_s). In (9.10), the limit of $X_n(r)$ is the product $Y_\alpha Z_\alpha$ of the two independent random variables Y_α and Z_α. In place of (9.4) we therefore have

$$X_n(r) \to_{a.s.} X(r) = Y_\alpha Z_\alpha.$$

In place of (9.5) we now have $\tilde{C}_n'(r) \to_{a.s.} C(r)$, where $C(r)$ denotes $C(r) = Y_\alpha^2 \sum_{s=1}^{\infty} \alpha^{-2s}\sigma_\epsilon^2 = Y_\alpha^2 \frac{\sigma_\epsilon^2}{\alpha^2-1}$. We therefore find that

$$\begin{aligned}\left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma_\epsilon^2}\right)^{1/2} (\hat{\alpha}_r - \alpha) &= \frac{X_n(r)}{\left(\tilde{C}_n'(r)\right)^{1/2}} \to_{a.s.} \frac{X(r)}{(C(r))^{1/2}} \\ &= \frac{Y_\alpha Z_\alpha}{|Y_\alpha| \left(\frac{\sigma_\epsilon^2}{\alpha^2-1}\right)^{1/2}} = \operatorname{sign}(Y_\alpha) \left(\frac{\alpha^2-1}{\sigma_\epsilon^2}\right)^{1/2} Z_\alpha.\end{aligned}$$

If $y_0 = 0$ and ϵ_s is i.i.d. $N(0, \sigma_\epsilon^2)$, then Y_α and Z_α are independent $N\left(0, \frac{\sigma_\epsilon^2}{\alpha^2-1}\right)$ variates and we have

$$\left(\frac{\sum_{t=1}^{[nr]} y_{t-1}^2}{\sigma_\epsilon^2}\right)^{1/2} (\hat{\alpha}_r - \alpha) \to_{a.s.} \frac{X(r)}{\left(\tilde{C}_n'(r)\right)^{1/2}} =_d N(0,1),$$

as shown in early work by White (1958) and Anderson (1959).

In concluding this section we note that, using Remark 8.1, results for regression asymptotics with transformed integrated regressors, such as those given in Park and Phillips (1999, 2001), may also be derived using martingale convergence arguments. The present approach therefore provides a unified treatment of asymptotics for stationary autoregression, autoregression with roots at or near unity and explosive cases as well that of regression with nonlinearly transformed integrated processes.

10. Concluding remarks

The last four sections illustrate the power of the martingale convergence approach in dealing with functional limit theory, weak convergence to stochastic integrals and time series asymptotics for both stationary and nonstationary processes. These examples reveal that the method encompasses much existing asymptotic theory in econometrics and is applicable to a wide class of interesting new problems where the limits involve stochastic integrals and mixed normal distributions. The versatility of the approach is most apparent in the unified treatment that it provides for the limit theory of autoregression, covering stationary, unit root, local to unity and explosive cases. No other approach to the limit theory has yet succeeded in accomplishing this unification.

While the technical apparatus of martingale convergence as it has been developed in Jacod and Shiryaev (2003) is initially somewhat daunting, it should be apparent from these econometric implementations that the machinery has a very broad reach in tackling asymptotic distribution problems in econometrics. Following the example of the applications given here, the methods may be applied directly to deliver asymptotic theory in many interesting econometric models, including models with some roots near unity and some cointegration as well as models with certain nonlinear forms of cointegration.

11. Appendix A1. Uniqueness and measurability hypotheses and continuity conditions for homogenous diffusion processes.

An important class of limit semimartingales X for which the conditions of uniqueness and measurability (A2) and (A3) of Theorem 3.1 are satisfied is given by homogenous diffusion processes with infinitesimal characteristics satisfying quite general conditions. These conditions also assure that the uniqueness hypothesis (B2) of Theorem 3.2) holds. We review some key results from that literature here together with some new results on multivariate diffusion processes that are used in the body of the paper.

For $d, m \in \mathbf{N}$, let $\sigma^{ij} : \mathbf{R}^d \to \mathbf{R}$, $i = 1, ..., d$, $j = 1, ..., m$, and $b^i : \mathbf{R}^d \to \mathbf{R}$, $i = 1, ..., d$, be continuous functions and let $\tilde{W} = (\tilde{W}(s), s \geq 0)$, $\tilde{W}(s) = (W^1(s), ..., W^m(s))$, be a standard $m-$dimensional Brownian motion. Consider the stochastic differential equation system $dX^i(s) = \sum_{j=1}^m \sigma^{ij}(X(s))dW^j(s) + b^i(X(s))ds$, $i = 1, ..., d$, or, in matrix form,

$$(dX(s))^T = \sigma(X(s))(d\tilde{W}(s))^T + b^T(X(s))ds, \tag{11.1}$$

where $\sigma : \mathbf{R}^d \to \mathbf{R}^{d\times m}$ and $b : \mathbf{R}^d \to \mathbf{R}^d$ are defined by $\sigma(x) = (\sigma^{ij}(x))_{1\leq i\leq d, 1\leq j\leq m} \in \mathbf{R}^{d\times m}$ and $b(x) = (b^1(x), ..., b^d(x)) \in \mathbf{R}^d$, $x \in \mathbf{R}^d$, and y^T denotes the transpose of the vector y.

Definition 11.1 *(see Definition IV.1.2 in Ikeda and Watanabe, 1989, and Definition III.2.24 in JS). A* ***solution*** *to (11.1) is a continuous $d-$dimensional process $X = (X(s), s \geq 0)$, $X(s) = (X^1(s), ..., X^d(s)) \in \mathbf{R}^d$, such that, for all $s \geq 0$ and all $i = 1, ..., d$, $X^i(s) - X^i(0) = \sum_{j=1}^m \int_0^s \sigma^{ij}(X(v))dW^j(v) + \int_0^s b^i(X(v))dv$. Such a solution is called a* ***homogenous diffusion process.***

Definition 11.2 *(Ikeda and Watanabe, 1989, Definition VI.1.4). It is said that* ***uniqueness of solutions*** *(in the sense of probability laws) holds for (11.1) if, whenever X_1 and X_2 are two solutions for (11.1) such that $X_1(0) = z$ a.s. and $X_2(0) = z$ a.s. for some $z \in \mathbf{R}^d$, then the laws on the space $\mathbb{D}(\mathbf{R}^d_+)$ of the processes X_1 and X_2 coincide.*

For an element $\alpha = (\alpha(s), s \geq 0)$ of the Skorohod space $\mathbb{D}(\mathbf{R}^d)$ and $i, j = 1, ..., d$, define

$$\begin{aligned} B^i(s, \alpha) &= \int_0^s b^i(\alpha(v))dv, \\ C^{ij}(s, \alpha) &= \sum_{k=1}^m \int_0^s \sigma^{ik}(\alpha(v))\sigma^{jk}(\alpha(v))dv = \int_0^s a^{ij}(\alpha(v))dv, \end{aligned} \tag{11.2}$$

where, for $x \in \mathbf{R}^d$ and $1 \leq i, j \leq d$,

$$a^{ij}(x) = \sum_{k=1}^m \sigma^{ik}(x)\sigma^{jk}(x). \tag{11.3}$$

Further, let $B(\alpha) = (B(s, \alpha), s \geq 0)$ and $C(\alpha) = (C(s, \alpha), s \geq 0)$, where $B(s, \alpha) = (B^1(s, \alpha), ..., B^d(s, \alpha))$, and $C(s, \alpha) = (C^{ij}(s, \alpha))_{1\leq i,j\leq d}$. A solution $X = (X(s), s \geq 0)$ to equation (11.1) is a semimartingale with the predictable characteristics $B(X)$ and $C(X)$.

The following lemma gives simple sufficient conditions for a homogenous diffusion (a solution to (11.1)) to satisfy continuity conditions (A4) and (B3).

Lemma 11.1 *If $\sigma(x)$ and $b(x)$ are continuous in $x \in \mathbf{R}^d$, then continuity conditions (A4) and (B3) of Theorems 3.1 and 3.2 are satisfied for the mappings $\alpha \to B(s, \alpha)$ and $\alpha \to C(s, \alpha)$ defined in (11.2).*

Proof. The lemma immediately follows from the definition of $B(s,\alpha)$ and $C(s,\alpha)$ and continuity of the matrix-valued function $a(x) = \sigma(x)\sigma^T(x) = (a^{ij}(x))_{1\le i,j\le d}$, where $a^{ij}(x)$, $1 \le i,j \le d$, are defined in (11.3). ■

For $B(s,\alpha)$ and $C(s,\alpha)$ defined above, one has, in notations (3.3) and (3.4), $\overline{B}_{(r)}(s,\alpha) = (\overline{B}^1_{(r)}(s,\alpha), ..., \overline{B}^d_{(r)}(s,\alpha))$ and $\overline{C}_{(r)}(s,\alpha) = (\overline{C}^1_{(r)}(s,\alpha), ..., \overline{C}^d_{(r)}(s,\alpha))$, where

$$\overline{B}^i_{(r)}(s,\alpha) = B^i(s+r,\overline{\alpha}_{(r)}) - B^i(r,\overline{\alpha}_{(r)}) = \int_r^{s+r} b^i(\alpha(v-r))dv = \int_0^s b^i(\alpha(v))dv = B^i(s,\alpha),$$

$$\begin{aligned}\overline{C}^{ij}_{(r)}(s,\alpha) &= C^{ij}(s+r,\overline{\alpha}_{(r)}) - C^{ij}(r,\overline{\alpha}_{(r)}) \\ &= \sum_{k=1}^m \int_r^{s+r} \sigma^{ik}(\alpha(v-r))\sigma^{jk}(\alpha(v-r))dv \\ &= \int_0^s \sigma^{ik}(\alpha(u))\sigma^{jk}(\alpha(v))dv = C^{ij}(s,\alpha),\end{aligned} \tag{11.4}$$

$i,j = 1, ..., d$, that is, $\overline{B}_{(r)} = B$ and $\overline{C}_{(r)} = C$ for all $r \ge 0$ in the uniqueness hypothesis (A2) in Theorem 3.1. Thus, in the case where, in Theorem 3.1, the predictable characteristics of the limit semimartingale X are $B(X)$ and $C(X)$ with B and C defined in (11.2) (the limit semimartingale X is a solution to differential equation (11.1)), conditions (A2) and (A3) simplify to the following:

*(A2') **Uniqueness hypothesis:** Let $\mathcal{H}$ denote the $\sigma-$field generated by $X(0)$ and let $\mathcal{L}_0$ denote the distribution of $X(0)$. For each $z \in \mathbf{R}^d$, the martingale problem associated with $(\mathcal{H}, X)$ and $(\mathcal{L}_0, B, C, \nu)$, where $X(0) = z$ a.s. and $\nu = 0$, has a unique solution P_z (see Definition 3.2).*

*(A3') **Measurability hypothesis:** The mapping $z \in \mathbf{R}^d \to P_z(A)$ is Borel for all $A \in \Im$.*

The following Theorems 11.1 and 11.2 give sufficient conditions for a homogenous diffusion (a solution to (11.1)) to satisfy conditions (A2) and (A3) (equivalently, (A2') and (A3')). They follow from Theorems IV.2.3, IV.2.4 and IV.3.1 in Ikeda and Watanabe (1989) and Theorem 5.3.1 in Durrett (1996) (see also the discussion following Theorem IV.6.1 on p. 215 in Ikeda and Watanabe, 1989, and Theorem III.2.32 in JS).

Theorem 11.1 *Conditions (A2) and (A3) of Theorem 3.1 are satisfied for a semimartingale $X = (X(s), s \ge 0)$ with the predictable characteristics $B(X)$ and $C(X)$ and B and C defined in (11.2) if and only if uniqueness of solutions (in the sense of probability laws) holds for (11.1).*

Theorem 11.2 *For any $z \in \mathbf{R}^d$, equation (11.1) has a unique (in the sense of probability laws) solution $X_{(z)} = (X_{(z)}(s), s \ge 0)$ with $X_{(z)}(0) = z$ if*

(C1) $\sigma(x)$ and $b(x)$ are locally Lipschitz continuous, that is, for every $N \in \mathbf{N}$ there exists a constant K_N such that $|\sigma(x) - \sigma(y)|^2 + |b(x) - b(y)|^2 \le K_N|x-y|^2$ for all $x, y \in \mathbf{R}^d$ such that $|x| \le N$ and $|y| \le N$.

(C2) There is a constant $K < \infty$ and a function $\phi(x) \ge 0$, $x \in \mathbf{R^d}$, with $\lim_{|x|\to\infty} \phi(x) = \infty$, so that if $X = (X(s), s \ge 0)$ is a solution of (11.1), then $(e^{-Ks}\phi(X(s)), s \ge 0)$ is a local supermartingale.

Let $a(x) = \sigma(x)\sigma^T(x)$ (in the component form, $a(x) = (a^{ij}(x))_{1\le i,j\le d}$, where $a^{ij}(x)$ are defined in 11.3). Condition (C2) above holds with $K = \tilde{K}$ if

(C3) $\sum_{i=1}^d 2x_i b_i(x) + a_{ii}(x) \le \tilde{K}(1+|x|^2)$ for some positive constant $\tilde{K}$ and all $x \in \mathbf{R}^d$.

Remark 11.1 *Analysis of the proof of Theorem 3.1 in Durrett (1996) reveals that condition $\lim_{|x|\to\infty} \phi(x) = \infty$ does indeed need to be imposed in the theorem, as indicated in (C2).*

Remark 11.2 *Conditions (C1) and (C2) (and, thus, (C1) and (C3)) of Theorem 11.2 guarantee existence of a global solution to (11.1) (that is, a solution defined for all $s \in \mathbf{R}_+$) and its uniqueness. Formally, for any $x \in \mathbf{R}$, a solution $X_{(x)}$ to (11.1) with the initial condition $X_{(x)}(0) = x$ and the stopping times $\tilde{S}_n$ defined by $\tilde{S}_n = \inf\{s \geq 0 : |X_{(x)}(s)| \geq n\}$, one has that the explosion time $\tilde{S}$ for $X_{(x)}$ given by $\tilde{S} = \lim_{n\to\infty} S_n$ is infinite a.s.: $\tilde{S} = \infty$ a.s.*

Remark 11.3 *In fact, conditions (C1) and (C2) (and, thus, (C1) and (C3)) of Theorem 11.2 are sufficient not only for existence and uniqueness of solutions for (11.1) in the sense of probability laws (Definition 11.2), but also for pathwise uniqueness of solutions (see Ikeda and Watanabe, 1989, Ch. IV). Theorems 11.1 and 11.2 have a counterpart, due to Stroock and Varadhan, according to which existence and uniqueness of solutions in the sense of probability laws holds for (11.1) if the following conditions are satisfied:*

(C1') $b(x)$ is bounded;

(C2') $a(x) = \sigma(x)\sigma^T(x)$ is bounded and continuous and everywhere invertible.

(see Theorem IV.3.3 and the discussion following Theorem IV.6.1 on p. 215 in Ikeda and Watanabe, 1989, Theorem III.2.34 and Corollary III.2.41 in JS, and Chapters 6 and 7 in Stroock and Varadhan, 1979).

For the proof of the main results in the paper, we will need a corollary of Theorems 11.1 and 11.2 in the case $d = 2$ and $m = 1$ (that is, in the case of a two-dimensional homogenous diffusion driven by a single Brownian motion) and functions $\sigma : \mathbf{R}^2 \to \mathbf{R}^{2\times 1}$ and $b : \mathbf{R}^2 \to \mathbf{R}^2$ given by

$$\begin{aligned} \sigma(x_1, x_2) &= (g_1(x_2), 1)^T, \\ b(x_1, x_2) &= (g_2(x_2), 0), \end{aligned} \tag{11.5}$$

where $g_i : \mathbf{R} \to \mathbf{R}$, $i = 1, 2$, are some continuous functions. In other words, we consider the stochastic differential equation

$$\begin{aligned} dX_1(s) &= g_1(X_2(s))dW(s) + g_2(X_2(s))ds; \\ dX_2(s) &= dW(s). \end{aligned} \tag{11.6}$$

A solution $X = (X(s), s \geq 0)$, $X(s) = (X_1(s), X_2(s))$ to (11.6) is a two-dimensional semimartingale with the predictable characteristics $B(X)$ and $C(X)$, where, for an element $\alpha = (\alpha(s), s \geq 0)$, $\alpha(s) = (\alpha_1(s), \alpha_2(s))$ of the Skorohod space $\mathbb{D}(\mathbf{R}^2_+)$,

$$\begin{aligned} B(s,\alpha) &= \Big(\int_0^s g_2(\alpha_2(v))dv, 0 \Big) = (B^1(s,\alpha), B^2(s,\alpha)), \\ C(s,\alpha) &= \begin{pmatrix} \int_0^s g_1^2(\alpha_2(v))dv & \int_0^s g_1(\alpha_2(v))dv \\ \int_0^s g_1(\alpha_2(v))dv & s \end{pmatrix} = \begin{pmatrix} C^{11}(s,\alpha) & C^{12}(s,\alpha) \\ C^{21}(s,\alpha) & C^{22}(s,\alpha) \end{pmatrix}. \end{aligned} \tag{11.7}$$

Corollary 11.1 *Suppose that the conditions hold:*

($\tilde{C}$1) The functions g_1 and g_2 are locally Lipschitz continuous, that is, for every $N \in \mathbf{N}$ there exists a constant K_N such that $|g_i(x) - g_i(y)| \leq K_N|x - y|$, $i = 1, 2$, for all $x, y \in \mathbf{R}$ such that $|x| \leq N$ and $|y| \leq N$;

($\tilde{C}$2) g_1 and g_2 satisfy the growth condition

$$|g_i(x)| \leq e^{K|x|}, \quad i = 1, 2, \tag{11.8}$$

for some positive constant K and all $x \in \mathbf{R}$.

Then, for any $z \in \mathbf{R}^2$, stochastic differential equation (11.6) has a unique solution $X_{(z)} = (X_{(z)}(s), s \geq 0)$ with $X_{(z)}(0) = z$ and, thus, by Theorem 11.1, conditions (A2) and (A3) of Theorem 3.1 are satisfied for a semimartingale $X = (X(s), s \geq 0)$, $X(s) = (X_1(s), X_2(s))$ with the predictable characteristics $B(X)$ and $C(X)$ and B and C defined in (11.7).

Proof. Clearly, under the assumptions of the corollary, condition (C1) of Theorem 11.2 is satisfied for the mappings σ and b defined in (11.5). Let us show that condition (C2) of Theorem 11.2 is satisfied with $A = 2 + 2K^2$ and $\phi(x_1, x_2) = x_1^2 + e^{2Kx_2} + e^{-2Kx_2}$. Clearly, $\lim_{|(x_1,x_2)|\to\infty} \phi(x_1, x_2) = \infty$. Similar to the proof of Theorem 5.3.1 in Durrett (1996), by Itô's formula we have that

$$\begin{aligned} d\Big[e^{-As}\phi(X_1(s), X_2(s))\Big] &= e^{-As}\Big[-A\Big(X_1^2(s) + e^{2KX_2(s)} + e^{-2KX_2(s)}\Big) \\ &\quad +2X_1(s)g_2(X_2(s)) + g_1^2(X_2(s)) + 2K^2\Big(e^{2KX_2(s)} + e^{-2KX_2(s)}\Big)\Big]ds \\ &\quad +e^{-As}\Big[2X_1(s)g_1(X_2(s)) + 2K\Big(e^{2KX_2(s)} - e^{-2KX_2(s)}\Big)\Big]dW(s). \end{aligned}$$

Since

$$\begin{aligned} -A\Big(X_1^2(s) + e^{2KX_2(s)} + e^{-2KX_2(s)}\Big) + 2X_1(s)g_2(X_2(s)) + g_1^2(X_2(s)) + 2K^2\Big(e^{2KX_2(s)} + e^{-2KX_2(s)}\Big) &= \\ -AX_1^2(s) + 2X_1(s)g_2(X_2(s)) + g_1^2(X_2(s)) - 2\Big(e^{2KX_2(s)} + e^{-2KX_2(s)}\Big) &\leq \\ (1-A)X_1^2(s) + g_2^2(X_2(s)) + g_1^2(X_2(s)) - 2\Big(e^{2KX_2(s)} + e^{-2KX_2(s)}\Big) &\leq 0 \end{aligned}$$

by condition ($\tilde{C}2$), we have that the process $(e^{-s}\phi(X(s)), s \geq 0)$ is a local supermartingale. Consequently, (C2) indeed holds and, by Theorems 11.1 and 11.2, the proof is complete. ■

Remark 11.4 *It is important to note that condition (C2') of Remark 11.3 is not satisfied for stochastic differential equation (11.6) since, as it is easy to see, the matrix $a(x) = \sigma(x)\sigma^T(x)$ is degenerate for σ defined in (11.5). The same applies, in general, to condition (C3) of Theorem 11.2. Therefore, the counterpart to Theorems 11.1 and 11.2 given by Remark 11.3 and, in general, linear growth condition (C3) cannot be employed to justify uniqueness and measurability hypothesis of Theorem 3.1 for the limit martingale X with the predictable characteristics $B(X)$ and $C(X)$ and B and C defined in (11.7). This is crucial in the proof of convergence to stochastic integrals in Sections 5 and 6 in the paper, where the limit semimartingales are solutions to (11.6), and we employ the result given by Corollary 11.1 to justify that conditions (A2) and (A3) of Theorem 3.1 hold for them.*

The following is a straightforward corollary of Lemma 11.1 in the case of stochastic equation (11.6).

Corollary 11.2 *Continuity conditions (A4) and (B3) of Theorems 3.1 and 3.2 hold for the mappings $\alpha \to B(s, \alpha)$ and $\alpha \to C(s, \alpha)$ defined in (11.7) if the functions $g_1(x)$ and $g_2(x)$ are continuous (in particular, (A4) and (B3) hold under assumption of local Lipschitz continuity ($\tilde{C}1$) of Corollary 11.1).*

12. Appendix A2. Embedding of a martingale into a Brownian motion

The following lemma gives the Skorohod embedding of martingales and a strong approximation to their quadratic variation. It was obtained in Park and Phillips (1999) in the case of the space $\mathbb{D}([0, 1])$ (see also Theorem A.1 in Hall and Heyde, 1980, Phillips and Ploberger, 1999, and Park and Phillips, 2001). The argument in the case of the space $\mathbb{D}(\mathbf{R}_+)$ is the same as in Park and Phillips (1999).

Lemma 12.1 *(Park and Phillips,1999, Lemma 6.2). Let assumption (D1) hold. Then there exists a probability space supporting a standard Brownian motion W and an increasing sequence of nonnegative stopping times $(T_k)_{k\geq 0}$ with $T_0 = 0$ such that*

(12.1)
$$\frac{1}{\sqrt{n}}\sum_{k=1}^{t}\epsilon_k =_d W\Big(\frac{T_t}{n}\Big),$$

$t \in \mathbf{N}$, *and*

$$\max_{1 \le t \le Nn} \frac{|T_t - \sigma_\epsilon^2 t|}{n^q} \to_{a.s.} 0, \tag{12.2}$$

$$\sup_{0 \le r \le N} \left| \frac{T_{[nr]}}{n} - \sigma_\epsilon^2 r \right| =_{a.s.} o(n^{q-1}) \tag{12.3}$$

for all $N \in \mathbf{N}$ and any $q > max(1/2, 2/p)$. In addition to the above, T_t is $\mathcal{E}_t-$measurable and, for all $\beta \in [1, p/2]$,

$$E((T_t - T_{t-1})^\beta | \mathcal{E}_{t-1}) \le K_\beta E(|\epsilon_t|^{2\beta} | \Im_{t-1}) \ \ a.s.$$

for some constant K_β depending only on β,

$$E(T_t - T_{t-1} | \mathcal{E}_{t-1}) = \sigma_\epsilon^2 \ \ a.s.,$$

where $\mathcal{E}_t$ is the $\sigma-$field generated by $(\epsilon_k)_{k=1}^t$ and $W(s)$ for $0 \le s \le T_t$.

13. Appendix A3. Auxiliary lemmas

Lemma 13.1 *(Billingsley, 1968, Theorem 4.1). Let $(\Omega, \Im, P)$ be a probability space and let $(E, \mathcal{E})$ be a metric space with a metric ρ. Let X_n, $Y_n, n \ge 1$, and X be $E-$valued random elements on $(\Omega, \Im, P)$ such that $X_n \to_d X$ and $\rho(X_n, Y_n) \to_P 0$. Then $Y_n \to X$.*

For $\alpha, \beta \in \mathbb{D}(\mathbf{R}_+)$, let $\alpha \circ \beta \in \mathbb{D}(\mathbf{R}_+)$ denote the composition of α and β, that is, the function $(\alpha \circ \beta)(s) = \alpha(\beta(s))$, $s \ge 0$.

Lemma 13.2 *If $X_n \to_d X$ and $Y_n \to_P Y$, where $X = (X(s), s \ge 0)$ and $Y = (Y(s), s \ge 0)$ are continuous processes, then $X_n \circ Y_n \to_d X \circ Y$.*

For the proof of Lemma 13.2, we need the following well-known result. Let $\rho(x, y)$ denote the Skorohod metric on $\mathbb{D}(\mathbf{R}_+)$ and let $\mathbb{C}(\mathbf{R}_+)$ denote the space of continuous functions on $\mathbf{R}_+$.

Lemma 13.3 *(Proposition VI.1.17 in JS; see also Theorem 15.12 in HWY). Let $x_n \in \mathbb{D}(\mathbf{R}_+)$, $n \ge 1$, and $x \in \mathbb{D}(\mathbf{R}_+)$. Then*

$$\sup_{0 \le s \le N} |x_n(s) - x(s)| \to 0 \tag{13.1}$$

for all $N \in \mathbf{N}$ implies that

$$\rho(x_n, x) \to 0. \tag{13.2}$$

If, in addition, $x \in \mathbb{C}(\mathbf{R}_+)$, then relations (13.1) and (13.2) are equivalent.

Proof of Lemma 13.2. Relations $X_n \to_d X$ and $Y_n \to_P Y$ imply (see Theorem 4.4 in Billingsley, 1968) that

$$(X_n, Y_n) \to_d (X, Y). \tag{13.3}$$

It is not difficult to see that the mapping $\psi : \mathbb{D}(\mathbf{R}_+^2) \to \mathbb{D}(\mathbf{R}_+)$ defined by $\psi(\alpha, \beta) = \alpha \circ \beta$ for $(\alpha, \beta) \in \mathbb{D}(\mathbf{R}_+^2)$ is continuous at (α, β) such that $\alpha, \beta \in \mathbb{C}(\mathbf{R}_+)$. Indeed suppose that, for the Skorohod metric ρ, $\rho(\alpha_n, \alpha) \to 0$ and $\rho(\beta_n, \beta) \to 0$, where $\alpha_n, \beta_n \in \mathbb{D}(\mathbf{R}_+)$, $n \ge 1$, and $\alpha, \beta \in \mathbb{C}(\mathbf{R}_+)$. We have that, for any $N \in \mathbf{N}$,

$$\sup_{0 \le s \le N} |\alpha_n \circ \beta_n(s) - \alpha \circ \beta(s)| \le \sup_{0 \le s \le N} |\alpha_n \circ \beta_n(s) - \alpha \circ \beta_n(s)| + \sup_{0 \le s \le N} |\alpha \circ \beta_n(s) - \alpha \circ \beta(s)| \tag{13.4}$$

Using Lemma 13.3 with $x_n = \beta_n$ and $x = \beta$ and continuity of β we get that, for all $n \geq 1$, $\sup_{0\leq s\leq N} |\beta_n(s)| \leq \sup_{0\leq s\leq N} |\beta_n(s) - \beta(s)| + \sup_{0\leq s\leq N} |\beta(s)| \leq K(N) < \infty$. Consequently, from the same lemma with $x_n = \alpha_n$ and $x = \alpha$ it follows that, for all $N \in \mathbf{N}$,

$$\sup_{0\leq s\leq N} |\alpha_n \circ \beta_n(s) - \alpha \circ \beta_n(s)| \leq \sup_{0\leq s\leq K(N)} |\alpha_n(s) - \alpha(s)| \to 0. \tag{13.5}$$

Using again Lemma 13.3 with $x_n = \beta_n$ and $x = \beta$ and uniform continuity of α, we also get that, for all $N \in \mathbf{N}$,

$$\sup_{0\leq s\leq N} |\alpha \circ \beta_n(s) - \alpha \circ \beta(s)| \to 0. \tag{13.6}$$

Relations (13.4)-(13.6) imply that (13.1) holds with $x_n = \alpha_n \circ \beta_n$ and $x = \alpha \circ \beta$ and thus, by Lemma 13.3, $\rho(\alpha_n \circ \beta_n, \alpha \circ \beta) \to 0$, as required.

Continuity of ψ and property (13.3) imply, by continuous mapping theorem (see JS, VI.3.8, and Billingsley, 1968, Corollary 1 to Theorem 5.1 and the discussion on pp. 144-145) that $X_n \circ Y_n = \psi(X_n, Y_n) \to_d \psi(X, Y) = X \circ Y$. ■

Lemma 13.4 *Let $p > 0$. Suppose that a sequence of identically distributed random variables $(\xi_t)_{t\in\mathbf{N}_0}$ is such that $E|\xi_0|^p < \infty$. Then*

$$n^{-1/p} \max_{0\leq k\leq nN} |\xi_k| \to_P 0 \tag{13.7}$$

for all $N \in \mathbf{N}$.

Proof. Evidently, (13.7) is equivalent to $n^{-1} \max_{0\leq k\leq nN} |\xi_k|^p \to_P 0$. Similar to the discussion preceding Theorem 3.4 in Phillips and Solo (1992) and the discussion in Hall and Heyde (1980, p. 53) we get that this relation, in turn, is equivalent to

$$J_n = \frac{1}{n}\sum_{k=1}^{Nn} |\xi_k|^p I(|\xi_k|^p > n\delta) \to_P 0$$

for all $\delta > 0$. The latter property holds because $EJ_n \leq NE|\xi_0|^p I(|\xi_0|^p > n\delta) \to 0$ by the dominated convergence theorem (see Theorem A.7 in Hall and Heyde, 1980) since $E|\xi_0|^p < \infty$. ■

As it is well known, the conclusion of Lemma 13.7 can be strengthened in the case of martingales. In particular, the following lemma holds.

Lemma 13.5 *Suppose that $(\eta_{tn}, \Im_t)_{t\in\mathbf{N}}$, $n \geq 1$, is an array of martingale-difference sequences with* $\max_{1\leq t\leq n} E\eta_{tn}^2 \leq L$ *for some constant $L > 0$ and all $n \in \mathbf{N}$. Then*

$$n^{-1} \max_{1\leq k\leq Nn} \Big|\sum_{t=1}^{k} \eta_{tn}\Big| \to_P 0$$

for all $N \in \mathbf{N}$.

Proof. By Kolmogorov's inequality for martingales (Hall and Heyde, 1980, Corollary 2.1) we get that, for all $\delta > 0$,

$$P\Big(n^{-1} \max_{1\leq k\leq Nn} \Big|\sum_{t=1}^{k} \eta_{tn}\Big| > \delta\Big) \leq E\Big(\sum_{t=1}^{Nn} \eta_{tn}\Big)^2/(\delta^2 n^2) \leq N \max_{1\leq t\leq Nn} E\eta_{tn}^2/n \leq NL/n \to 0,$$

as required. ■

Lemma 13.6 *For the random variables $\tilde{\epsilon}_t$ defined in the proof of Theorem 4.2, one has $E|\tilde{\epsilon}_0|^p < \infty$ if $(\epsilon_t)_{t\in\mathbf{Z}}$ satisfy assumption (D2) with $p > 2$.*

Proof. Since $E|\epsilon_0|^p < \infty$, by the triangle inequality for the L_p-norm $||\cdot||_p = (E|\cdot|^p)^{1/p}$ and Lemma 2.1 in Phillips and Solo (1992) we have $||\tilde{\epsilon}_0||_p = ||\sum_{j=0}^{\infty} \tilde{c}_j \epsilon_{-j}||_p \le ||\epsilon_0||_p \sum_{j=0}^{\infty} |\tilde{c}_j| < \infty$. ■

Lemma 13.7 *For g_{jk} defined in the proof of Theorem 4.4, one has $\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} |g_{rj}| < \infty$ for all r if $\sum_{j=1}^{\infty} jc_j^2 < \infty$.*

Proof. Using change of summation indices and Hölder inequality, we have that

$$\begin{aligned}\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} |g_{rj}| &= \sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} |c_j||c_{j+r}| = \sum_{j=1}^{\infty} j|c_j||c_{j+r}| = \\ \sum_{j=1}^{\infty} j^{1/2}|c_j|j^{1/2}|c_{j+r}| &\le \Big(\sum_{j=1}^{\infty} j|c_j|^2\Big)^{1/2}\Big(\sum_{j=1}^{\infty} j|c_{j+r}|^2\Big)^{1/2} < \infty,\end{aligned}$$

as required. ■

Lemma 13.8 *For the random variables $\tilde{u}_{at}$ and $\tilde{u}_{bt}$ defined in the proof of Theorem 4.4, one has $Eu_{a0}^2 < \infty$ and $Eu_{b0}^2 < \infty$ if $(\epsilon_t)_{t\in\mathbf{Z}}$ satisfy assumption (D2) with $p > 2$.*

Proof. The property $Eu_{b0}^2 < \infty$ holds by Lemma 5.9 in Phillips and Solo (1992). By the triangle inequality for the L_2-norm $||\cdot||_2 = (E(\cdot)^2)^{1/2}$ and Lemma 13.7, $||\tilde{u}_{a0}||_2 = \Big|\Big|\sum_{k=0}^{\infty} \tilde{g}_{mk}\epsilon_{-k}^2\Big|\Big|_2 \le ||\epsilon_0^2||_2 \sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} |g_{mj}| < \infty$ Consequently, $E\tilde{u}_{a0}^2 = O\Big(\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} |g_{mj}|\Big)^2 < \infty$. ■

Lemma 13.9 *For $\tilde{h}_{kr}$ defined in the proof of Theorem 5.2, one has $\sum_{r=0}^{\infty}\sum_{k=0}^{\infty} |\tilde{h}_{kr}| < \infty$ if $\sum_{j=1}^{\infty} j|c_j| < \infty$.*

Proof. By definition of $\tilde{h}_{kr}$, it suffices to prove that

$$\sum_{r=0}^{\infty}\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} |c_j||\tilde{c}_{j+r}| < \infty \tag{13.8}$$

and

$$\sum_{r=0}^{\infty}\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} |\tilde{c}_j||c_{j+r}| < \infty. \tag{13.9}$$

Using change of summation indices, we have that

$$\sum_{r=0}^{\infty}\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} |c_j||\tilde{c}_{j+r}| \le \sum_{r=0}^{\infty}\sum_{j=1}^{\infty} j|c_j||\tilde{c}_{j+r}| = \sum_{j=1}^{\infty} j|c_j| \sum_{k=j}^{\infty} |\tilde{c}_k| \le \Big(\sum_{j=1}^{\infty} j|c_j|\Big)\Big(\sum_{k=1}^{\infty} |\tilde{c}_k|\Big) < \infty, \tag{13.10}$$

$$\begin{aligned}\sum_{r=0}^{\infty}\sum_{k=0}^{\infty}\sum_{j=k+1}^{\infty} |\tilde{c}_j||c_{j+r}| &\le \sum_{r=0}^{\infty}\sum_{j=1}^{\infty} j|\tilde{c}_j||c_{j+r}| \le \sum_{r=0}^{\infty}\sum_{j=1}^{\infty} j|c_{j+r}| \sum_{k=j+1}^{\infty} |c_k| \le \\ \sum_{r=0}^{\infty}\sum_{j=1}^{\infty} |c_{j+r}| \sum_{k=j+1}^{\infty} k|c_k| &\le \Big(\sum_{j=1}^{\infty}\sum_{s=j}^{\infty} |c_s|\Big)\Big(\sum_{k=1}^{\infty} k|c_k|\Big) < \infty \end{aligned} \tag{13.11}$$

because, as in Lemma 2.1 in Phillips and Solo (1992) and its proof, $\sum_{j=1}^{\infty} j|c_j| < \infty$ implies that $\sum_{j=1}^{\infty} |\tilde{c}_j| < \infty$ and, even stronger, $\sum_{j=1}^{\infty}\sum_{s=j}^{\infty} |c_s| < \infty$. ■

Lemma 13.10 *For the random variables $\tilde{w}_{ak}$ and $\tilde{w}_{bk}$ defined in the proof of Theorem 5.2, one has $E|\tilde{w}_{a0}|^{p/2} < \infty$ and $E|\tilde{w}_{b0}|^{p/2} < \infty$ if $(\epsilon_t)_{t\in\mathbf{Z}}$ satisfy assumption (D2) with $p > 2$ and $\sum_{j=1}^{\infty} j|c_j| < \infty$.*

Proof. Denote $q = p/2$. Since $E|\epsilon_0|^p < \infty$, by the triangle inequality for the L_q-norm $||\cdot||_q = (E|\cdot|^q)^{1/q}$ and Lemma 13.9, we get

$$||\tilde{w}_{a0}||_q = \Big|\Big|\sum_{k=0}^{\infty} \tilde{h}_{k0}\epsilon_{-k}^2\Big|\Big|_q \leq ||\epsilon_0||_p \sum_{k=0}^{\infty} |\tilde{h}_{k0}| < \infty,$$

$$||\tilde{w}_{b0}||_q \leq \sum_{r=1}^{\infty} ||\tilde{h}_r(L)\epsilon_0\epsilon_{-r}||_q \leq (||\epsilon_0||_q)^2 \sum_{r=1}^{\infty}\sum_{k=0}^{\infty} |\tilde{h}_{kr}| < \infty.$$

Consequently, $E|\tilde{w}_{a0}|^q < \infty$ and $E|\tilde{w}_{b0}|^q < \infty$, as required. ■

Lemma 13.11 *For the random variables η_{t-1}^h defined in the proof of Theorem 5.2, one has $E(\eta_{-1})^4 < \infty$ if $(\epsilon_t)_{t\in\mathbf{Z}}$ satisfy assumption (D2) with $p \geq 4$ and $\sum_{j=1}^{\infty} j|c_j| < \infty$.*

Proof. As in Lemma 2.1 in Phillips and Solo (1992) and its proof, $\sum_{j=1}^{\infty} j|c_j| < \infty$ implies that $\sum_{j=1}^{\infty} |\tilde{c}_j| < \infty$ and, even stronger, $\sum_{j=1}^{\infty}\sum_{s=j}^{\infty} |c_s| < \infty$. Therefore, under the assumptions of the theorem, $\sum_{r=1}^{\infty} |h_r(1)| \leq \sum_{r=1}^{\infty}\sum_{k=0}^{\infty} |c_k||\tilde{c}_{k+r}| + \sum_{r=1}^{\infty}\sum_{k=0}^{\infty} |\tilde{c}_k||c_{k+r}| \leq 2\big(\sum_{j=0}^{\infty}|c_j|\big)\big(\sum_{j=0}^{\infty}|\tilde{c}_j|\big) < \infty$. Using the triangle inequality for the L_4-norm $||\cdot||_4 = (E|\cdot|^4)^{1/4}$, we get, therefore,

$$||\eta_{-1}||_4 = \Big|\Big|\sum_{r=1}^{\infty} h_r(1)\epsilon_{-r}\Big|\Big|_4 \leq ||\epsilon_0||_4 \sum_{r=1}^{\infty} |h_r(1)| < \infty.$$

Consequently, $E(\epsilon_{-1}^h)^4 = O\big(\sum_{r=1}^{\infty} h_r(1)\big) < \infty$. ■

Lemma 13.12 *Under the assumptions of Theorem 6.1, one has*

$$\max_{1\leq k\leq n} E\Big(f'\Big(\frac{1}{\sqrt{n}}\sum_{t=1}^{k} u_t\Big)\Big)^4 \leq L$$

for some constant $L > 0$ and all $n \in \mathbf{N}$.

Proof. The growth condition $|f'(x)| \leq K(1+|x|^\alpha)$ evidently implies that $(f'(x))^4 \leq K(1+x^{4\alpha})$. Consequently, using (4.6), we get that, for all k,

$$\Big(f'\Big(\frac{1}{\sqrt{n}}\sum_{t=1}^{k} u_t\Big)\Big)^4 \leq K\Big(1+\Big|\frac{1}{\sqrt{n}}\sum_{t=1}^{k} u_t\Big|^{4\alpha}\Big) = K\Big(1+\Big|\frac{C(1)}{\sqrt{n}}\sum_{t=1}^{k}\epsilon_t + \frac{\tilde{\epsilon}_0}{\sqrt{n}} - \frac{\tilde{\epsilon}_k}{\sqrt{n}}\Big|^{4\alpha}\Big) \leq$$

$$K\Big(1+\Big|\frac{C(1)}{\sqrt{n}}\sum_{t=1}^{k}\epsilon_t\Big|^{4\alpha} + \Big|\frac{\tilde{\epsilon}_0}{\sqrt{n}}\Big|^{4\alpha} + \Big|\frac{\tilde{\epsilon}_k}{\sqrt{n}}\Big|^{4\alpha}\Big).$$

Thus, for some constant $K > 0$,

$$\max_{1\leq k\leq n} E\Big(f'\Big(\frac{1}{\sqrt{n}}\sum_{t=1}^{k} u_t\Big)\Big)^4 \leq K\Big(1+\max_{1\leq k\leq n} E\Big|\frac{C(1)}{\sqrt{n}}\sum_{t=1}^{k}\epsilon_t\Big|^{4\alpha} + E\Big|\frac{\tilde{\epsilon}_0}{\sqrt{n}}\Big|^{4\alpha}\Big). \tag{13.12}$$

Since, by the assumptions of the theorem, $E|\epsilon_0|^p < \infty$ for some $p \geq max(6, 4\alpha)$, we get, by Lemma 13.4, that $E|\tilde{\epsilon}_0|^{4\alpha} < \infty$. Since for i.i.d. random variables η_t, $t \geq 1$, and $p > 2$,

$$E\Big|\sum_{t=1}^{k}\eta_t\Big|^p \leq Kn^{p/2}E|\eta_1|^p \tag{13.13}$$

(see, e.g., Dharmadhikari, Fabian and Jogdeo, 1968, and also de la Peña, Ibragimov and Sharakhmetov, 2003), we also conclude, using Jensen's inequality, that

$$\max_{1\le k\le n} E\Big|\frac{C(1)}{\sqrt{n}}\sum_{t=1}^{k}\epsilon_t\Big|^{4\alpha} \le \Big(E\Big|\frac{C(1)}{\sqrt{n}}\sum_{t=1}^{k}\epsilon_t\Big|^{p}\Big)^{p/(4\alpha)} \le K(E|\epsilon_0|^p)^{4\alpha/p}$$

for some constant $K > 0$. These estimates evidently imply, together with (13.12), that bound (13.12) indeed holds. ■

REFERENCES

Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* **30**, 676-687.

Avram, F. (1988). Weak convergence of the variations, iterated integrals and Doleans-Dade exponentials of sequences of semimartingales. *Annals of Probability* **16**, 246-250.

Billingsley, P. (1968). *Convergence of probability measures.* Wiley, New York.

Borodin, A. N. and Ibragimov, I. A. (1995). Limit theorems for functionals of random walks. *Proceedings of the Steklov Institute of Mathematics*, no. 2 (195).

Chan, N. H. and C. Z. Wei (1987). "Asymptotic inference for nearly nonstationary AR(1) processes, *Annals of Statistics* 15, 1050–1063.

Chan, N. H. and C. Z. Wei (1988). "Limiting distributions of least squares estimates of unstable autoregressive processes," *Annals of Statistics* 16, 367–401.

Coffman, E. G., Puhalskii, A. A. and Reiman, M. I. (1998). Polling systems in heavy traffic: A Bessel process limit. *Mathematics of Operations Research* **23**, 257-304.

Dharmadhikari, S. W., Fabian, V. and Jogdeo, K. (1968). Bounds on the moments of martingales. *Annals of Mathematical Statistics* **39**, 1719-1723.

de la Peña, V. H., Ibragimov R. and Sharakhmetov, S. (2003). On extremal distributions and sharp L_p-bounds for sums of multilinear forms. *Annals of Probability* **31**, 630-675.

de Jong, R. M. (2002). Nonlinear estimators with integrated regressors but without exogeneity. Working paper, Michigan State University.

Durrett, R. (1996). *Stochastic calculus. A practical introduction.* CRC Press, Boca Raton, FL.

Dynkin, E. B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals. *Annals of Statistics* **11**, 739-745.

Giraitis, L. and P. C. B. Phillips (2004). Uniform Limit Theory for Stationary Autoregression. University of York, mimeographed.

Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic Press, New York.

He, S.-W., Wang, J.-G. and Yan, J.-A. (1992). *Semimartingale theory and stochastic calculus.* CRC Press, Boca Raton.

Hu, L. and P. C. B. Phillips (2001). Dynamics of the Federal Funds Target Rate: a Nonstationary Discrete Choice Approach. *Journal of Applied Econometrics* (forthcoming)

Ikeda, N. and Watanabe, S. (1989). *Stochastic differential equations and diffusion processes.* North-Holland, Amsterdam.

Jacod, J. and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes.* 2nd edition. Springer-Verlag, Berlin.

Jeganathan, P. (2003a). Convergence of functionals of sums of r.v.s. to local times of fractional stable motions. Working paper, Indian Statistical Institute.

Jeganathan, P. (2003b). Second order limits of functionals of sums of linear processes that converge to fractional stable motions. Working paper, Indian Statistical Institute.

Liptser, R. Sh. and Shiryaev, A. N. (1989). *Theory of martingales.* Kluwer, Dordecht.

Mandelbaum, A. and Taqqu, M. S. (1984) Invariance principle for symmetric statistics. *Annals of Statistics* **12**, 483-496.

Park J. Y. and Phillips P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* **15**, 269-298.

Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69**, 117-161.

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

Phillips, P. C. B. (1987b). Towards a unified asymptotic theory for autoregression. *Biometrika*, **74**, 535-547.

Phillips, P. C. B. and Magdalinos, T. (2004). Limit theory for moderate deviations from unity. Cowles Foundation Discussion Paper, Yale University.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression *Biometrika* **75**, 335-346.

Phillips, P. C. B. and Ploberger, W. (1996). An asymptotic theory of Bayesian inference for time series. *Econometrica* **64**, 381-413.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics* **20**, 971-1001.

Phillips, P. C. B. (1999). Unit root log periodogram regression. Cowles Foundation Discussion Paper 1244, Yale University.

Phillips, P. C. B. and Magdalinos, T. (2004). Limit theory for moderate deviations from a unit root. Working paper, Yale University.

Pöetscher, B. M. (2004). Nonlinear functions and convergence to Brownian motion: beyond the continuous mapping theorem. *Econometric Theory* **20**, 1-22.

Saikkonen, P. and Choi, I. (2004). Cointegrating smooth transition regressions. *Econometric Theory* **20**, 301-340.

Stroock, D. W. and Varadhan, S. R. S. (1979). *Multidimensional diffusion processes.* Springer-Verlag, New York.

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *Annals of Mathematical Statistics* **29**, 1188-1197.