Convergence to Stochastic Integrals with Non-linear Integrands

Carlos Caceres
Nuffield College, University of Oxford

Bent Nielsen*
Nuffield College, University of Oxford

12 February 2007

Abstract

In this paper we present a general result concerning the convergence to stochastic integrals with non-linear integrands. The key finding represents a generalization of Chan and Wei's (1988) Theorem 2.4. and that of Ibragimov and Phillips' (2004) Theorem 8.2. This result is necessary for analysing the asymptotic properties of mis-specification tests, when applied to a unit root process, for which Wooldridge (1999) mentioned that the exiting results in the literature were not sufficient.

Key words: *non-stationarity, unit roots, convergence, autoregressive processes, martingales, stochastic integrals, non-linearity.*

*This co-author received financial support from ESRC grant RES-000-27-0179.

1 Introduction

The asymptotic analysis of unit root statistics relies on the use of the Functional Central Limit Theorem, the Continuous Mapping Theorem, and on convergence to stochastic integrals. However, Wooldridge (1999) mentioned that these results are insufficient for the analysis of, for instance, White's (1980) test for heteroskedasticity applied to a unit root process. For that analysis the asymptotic properties are needed for terms like

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n-1} f\left(\frac{X_{1,n}\left(\frac{t}{n}\right)}{\sqrt{n}}\right) u_{2,t+1} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-1} f\left(\frac{\sum_{k=1}^{t} u_{1,k}}{\sqrt{n}}\right) u_{2,t+1} \tag{1}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a non-linear function, and $(u_{1,n}, u_{2,n})_{n\in\mathbb{Z}}$ is a martingale difference sequence with respect to a filtration $\mathcal{F}_n$. Moreover, for most applications it is necessary to establish this convergence jointly with the convergence of the process $n^{-1/2}X_{1,n}$. Such a joint convergence result will be developed in this paper. In the most general result presented, f is a function of a vector random walk including the integrator as one of its elements, as well as of time.

The result presented in this paper generalizes two types of results available in the literature. First, Chan and Wei (1988, Theorem 2.4) considered the linear case where $f(x) = x$ and established the joint convergence result

$$\left(\frac{1}{\sqrt{n}}X_{1,n}, \frac{1}{\sqrt{n}}X_{2,n}, \frac{1}{n}\sum_{k=1}^{n-1} X_{1,n}\left(\frac{k}{n}\right) u_{2,k+1}\right) \xrightarrow{d} \left(B_1, B_2, \int_0^1 B_1\, dB_2\right) \tag{2}$$

on $D^2[0,1]\times\mathbb{R}$, where B_1 and B_2 are two Brownian motions with respect to an increasing sequence of σ-fields $\mathcal{G}_t$. Their proof will be followed to a large extent. Secondly, the case where f is non-linear has received less attention. Strasser (1986) considered the case where f is Lipschitz which excludes the polynomial functions often found in econometrics. Ibragimov and Phillips (2004) considered more general non-linear functions, and established a convergence result for sums like that presented in (1) using general convergence results for semi-martingales. This was done under the restrictive assumption that the innovations are independent and identically distributed, which prohibits important applications.

The result presented in this paper will be proved along the lines of Chan and Wei (1988) with one important change. Since they analysed the case of linear functions f, so that the involved sums are quadratic forms, they could make extensive use of the uncorrelatedness of martingale differences. In the proof presented in this paper, this argument will have to be replaced. The outline of the paper is therefore: section 2 will present the main result, whereas section 3 presents some lemmas used to replace Chan and Wei's argument based on uncorrelatedness. The proof of the main result follows in section 4.

Throughout this paper the notation $[[1, h]]$ is used for a sequence $1, 2, \ldots, h$ of natural numbers.

2 Main Results

Two versions of the main result are presented. Theorem 1 covers the quantity given in equation (1), so that the function f does not involve the integrator. Thereby it is possible to get an overview of the necessary moment conditions. Theorem 2 generalizes the result to functions of vectors of random walks that can include the integrator.

In order to formulate the main results the following sets of assumptions are needed.

Assumption 1. *Let $f : \mathbb{R}^h \times [0,1] \to \mathbb{R}$ be a differentiable vector function satisfying the growth condition: for any vector $\boldsymbol{x} = (x_1, \ldots, x_{h+1})' \in \mathbb{R}^{h+1}$, we have $\forall r \in [[1, h+1]]$,*

$$\left|\frac{\partial f(\boldsymbol{x})}{\partial x_r}\right| \leq K_r \left(1 + \sum_{s=1}^{h+1} |x_s|^{\alpha_{r,s}}\right) \tag{3}$$

for a set of $h+1$ positive constants $(K_1, \ldots, K_{h+1})$, and $(h+1)^2$ positive, but finite, integers

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{1,1} & \cdots & \alpha_{1,h} & \alpha_{1,h+1} \\ \vdots & \ddots & \vdots & \vdots \\ \alpha_{h,1} & \cdots & \alpha_{h,h} & \alpha_{h,h+1} \\ \alpha_{h+1,1} & \cdots & \alpha_{h+1,h} & \alpha_{h+1,h+1} \end{bmatrix} \tag{4}$$

Assumption 1 is satisfied for instance by power functions: $f(x) = x^p$ and by Lipschitz functions. To see the latter note that a function $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz if $|f(y) - f(x)| \leq K^*|x-y|$ for a constant $K^* > 0$ and any $x, y \in \mathbb{R}$. For such a function it can be established that $|f'(x)| \leq K_0(1 + |x|^\alpha)$ for some constant K_0 and some integer α.

Assumption 2. *Let $\boldsymbol{X}_n = (X_{1,n}, X_{2,n}, \ldots X_{h,n})$ be the partial sum process defined by: $\forall s \in [[1, h]]$, $X_{s,n}(t) = \sum_{k=1}^{[nt]} u_{s,k}$. Suppose it satisfies*

$$n^{-1/2}\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{B} \tag{5}$$

on $D[0,1]^h$, where $\boldsymbol{B} = (B_1, B_2, \ldots, B_h)$ is a h-dimensional Brownian motion.

Note that in the case where all the $u_{s,n}$'s are also i.i.d. with zero mean and constant covariances, Assumption 2 is equivalent to the Multivariate Donsker Theorem (c.f. White (2000), Theorem 7.27).

Some higher order martingale difference sequence properties and some moment conditions are needed that depend on the constants $\alpha_{r,s}$ in (4). In what follows, we consider the case where $h = 2$, and we replace the term x_{h+1} by t to emphasize that the latter refers to the use of a deterministic term. First the case where $f(x_1, x_2, t)$ does not vary with x_2 is considered.

Assumption 3. *Let $(u_{1,n}, u_{2,n})'$ be a martingale difference sequence with respect to a filtration $\mathcal{F}_n$. If $K_1 > 0$,*

$$E(u_{1,n}^i|\mathcal{F}_{n-1}) \overset{a.s.}{=} c_{1,i} \quad \text{for} \quad i = 2, \ldots, (\alpha_{1,1}+2) \cdot I_{(\alpha_{1,1}>0)}, \tag{6}$$

$$E\left(u_{1,n}^{2\alpha_{1,1}+2}\right) \overset{a.s.}{\leq} c_{1,1}^* \tag{7}$$

If $K_{h+1} > 0$,

$$E\left(u_{1,n}^i|\mathcal{F}_{n-1}\right) \overset{a.s.}{=} \tilde{c}_{1,i} \quad \text{for} \quad i = 2, \ldots, \min\{\alpha_{h+1,1}+1, 2\alpha_{h+1,1}-1\}, \tag{8}$$

$$E\left(u_{1,n}^{2\alpha_{h+1,1}}\right) \overset{a.s.}{\leq} c_{1,h+1}^* \tag{9}$$

In addition,

$$E(u_{2,n}^2|\mathcal{F}_{n-1}) \overset{a.s.}{\leq} c_{2,2}^* \tag{10}$$

for a set of positive integers $(\alpha_{1,1}, \alpha_{h+1,1})$*,* K_1 *and* K_{h+1} *as in Assumption 1, some positive constants* $c_{1,1}^*$*,* $c_{1,h+1}^*$*,* $c_{2,2}^*$ *and some set of constants* $(c_{1,i}, \tilde{c}_{1,i})$*.*

Assumption 3 is required to replace the uncorrelatedness property of martingale differences used in Chan and Wei (1988). Conditions (6) and (8) are higher order martingale different properties. These ensure uncorrelatedness of powers of the innovations. Conditions (7) and (9) are corresponding unconditional moment conditions, whereas condition (10) is the only conditional moment bound. It relates to the integrator and is not related to the form of f. Note that in contrast to Chan and Wei (1988), no bounds are required for the conditional variance of the innovation $u_{1,n}$ related to the integrand. In addition, if no deterministic terms are included in f, then conditions (8) and (9) do not apply.

The following three examples illustrates some of the uses and differences when applying the above three assumptions to different functional forms for the function f.

Example 1. $f(x_1, x_2, t) = 1$
Then $K_1 = K_2 = K_{h+1} = 0$ *and conditions (6) to (9) do not apply. Thus, in this case, we are interested in showing*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} u_{2,k+1} \overset{d}{\longrightarrow} B_2(1) \tag{11}$$

which, given Assumptions 1 and 2, is basically equivalent to Donsker's theorem as given in Chan and Wei (1988, Theorem 2.2).

Example 2. $f(x_1, x_2, t) = x_1$
Then $K_1 > 0$*,* $K_2 = K_{h+1} = 0$ *,* $\alpha_{1,1} = 0$*,* $\alpha_{h+1,1} = 0$ *and conditions (6), (8) and (9) do not apply, but (7) is needed. Thus, in this case, we are interested in showing*

$$\frac{1}{n} \sum_{k=1}^{n-1} X_{1,n}\left(\frac{k}{n}\right) \cdot u_{2,k+1} \overset{d}{\longrightarrow} \int_0^1 B_1(u)\, dB_2(u) \tag{12}$$

Given Assumptions 1 and 2, the above example is similar to Theorem 2.4 of Chan and Wei (1988), noting however that (7) only requires the unconditional variance of $u_{1,n}$ *to be bounded as opposed to a bound on the conditional variance.*

Example 3. $f(x_1, x_2, t) = x_1^2$
Then $K_1 > 0$*,* $K_2 = K_{h+1} = 0$ *,* $\alpha_{1,1} = 1$*,* $\alpha_{h+1,1} = 0$*, so conditions (6) and (7) imply that the first three conditional moments of* $(u_{1,n})$ *are constant and the fourth moment bounded. Thus, in this case, we are interested in showing*

$$\frac{1}{n^{3/2}} \sum_{k=1}^{n-1} \left[X_{1,n}\left(\frac{k}{n}\right)\right]^2 u_{2,k+1} \overset{d}{\longrightarrow} \int_0^1 \left[B_1(u)\right]^2 dB_2(u) \tag{13}$$

The term in (13) cannot be dealt with by using Chan and Wei's result due to the presence of the non-linearity in $X_{1,n}$*. In fact, this point is the limitation in the existing literature that was implicitly* *pointed out by Wooldridge (1999). Thus, the need of the result presented in this paper, in order to be able to deal with terms such as that appearing in (13) above. If in addition $(u_{1,n}, u_{2,n})$ is an i.i.d. sequence, then the result in (13) also follows from the result presented by Ibragimov and Phillips (2004, Theorem 8.2). However, their result cannot be used to show (13) for the case where we have an i.i.d. sequence (ξ_n) with standard normal distribution generating $u_{1,n} = \xi_n$ and $u_{2,n} = \xi_n \xi_{n-1}$, which arises in some of the problems alluded to by Wooldridge (1999).*

The following theorem is a generalization of the results presented by Chan and Wei (1988) and Ibragimov and Phillips (2004). However, in here we allow for the possibility of including more general 'functional' terms such as the one presented in equation (1), while still maintaining the joint convergence result of equation (2).

Theorem 1. *Let $f : \mathbb{R} \times [0,1] \to \mathbb{R}$ be a function satisfying Assumption 1. Suppose that Assumptions 2 and 3 are satisfied. Then, on $D^2[0,1] \times \mathbb{R}$,*

$$\left(\frac{1}{\sqrt{n}} X_{1,n}, \frac{1}{\sqrt{n}} X_{2,n}, \frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} f\left[\frac{1}{\sqrt{n}} X_{1,n}\left(\frac{k}{n}\right), \frac{k}{n} \right] u_{2,k+1} \right) \xrightarrow{d} \left(B_1, B_2, \int_0^1 f\left[B_1(u), u\right] dB_2(u) \right) \tag{14}$$

Theorem 1 can be generalized so that the function f can include not only $X_{1,n}$ but also $X_{2,n}$. In fact this result also holds for a finite number of martingales $X_{s,n}$. This is presented in the following theorem. But first we need to modify Assumption 3 accordingly.

Assumption 4. *Let $u_n = (u_{1,n}, \ldots, u_{h,n})'$ be a martingale difference sequence with respect to a filtration $\mathcal{F}_n$. Let $\boldsymbol{\alpha}$ be a set of integers as defined in Assumption 1, and let $(c^*_{r,r}, c^*_{j,r}, c^*_{j,h+1}, c^*_{h,2})$ be positive constants and $(c_{r,i}, c_{j,i}, \tilde{c}_{j,i})$ some constants.*
If $K_r > 0$, $r \in [[1,h]]$,

$$E\left(u^i_{r,n} | \mathcal{F}_{n-1}\right) \overset{a.s.}{=} c_{r,i} \quad \text{for} \quad i = 2, \ldots, \max\{3 \cdot I_{(\sum_{j=1}^h \alpha_{r,j} > 0)}, (\alpha_{r,r}+2) \cdot I_{(\alpha_{r,r}>0)}\}, \tag{15}$$

$$E\left(u^{2\beta_r}_{r,n}\right) \overset{a.s.}{\leq} c^*_{r,r} \quad \text{where} \quad \beta_r = \max\{2 \cdot I_{(\sum_{j=1}^h \alpha_{r,j} > 0)}, \alpha_{r,r}+1\}. \tag{16}$$

If $K_r > 0$, $r \in [[1,h]]$, $j \in [[1,h]]$ so that $j \neq r$,

$$E\left(u^i_{j,n} | \mathcal{F}_{n-1}\right) \overset{a.s.}{=} c_{j,i} \quad \text{for} \quad i = 2, \ldots, \min\{2\alpha_{r,j}+1, 4\alpha_{r,j}-1\}, \tag{17}$$

$$E\left(u^{4\alpha_{r,j}}_{j,n}\right) \overset{a.s.}{\leq} c^*_{j,r} \tag{18}$$

If $K_{h+1} > 0$, $j \in [[1,h]]$,

$$E\left(u^i_{j,n} | \mathcal{F}_{n-1}\right) \overset{a.s.}{=} \tilde{c}_{j,i} \quad \text{for} \quad i = 2, \ldots, \min\{\alpha_{h+1,j}+1, 2\alpha_{h+1,j}-1\}, \tag{19}$$

$$E\left(u^{2\alpha_{h+1,j}}_{j,n}\right) \overset{a.s.}{\leq} c^*_{j,h+1} \tag{20}$$

In addition,

$$E(u^2_{h,n} | \mathcal{F}_{n-1}) \overset{a.s.}{\leq} c^*_{h,2} \tag{21}$$

Example 4. *$f(\boldsymbol{X}) = f(x_1, t)$ and $h = 2$*
In this case, $K_2 = 0$, $\alpha_{1,2} = 0$ and $\alpha_{2,s} = 0$ for $s \in \{1, 2\}$. Assumption 4 then reduces to Assumption 3. Nevertheless, it is important to note here that we need conditions (15) to (20) to hold for all the $u_{s,n}$, $s \in [[1, h]]$ - which are included in the function f. In Assumption 3, these conditions were only required to hold for $u_{1,n}$ (i.e. for the term included inside the function f), whereas $u_{2,n}$ was only required to satisfy condition (21).

Once again, we present here an example to illustrate the use of Assumption 4 in conjunction with Assumptions 1 and 2.

Example 5. *$f(\boldsymbol{X}) = x_1^2 \cdot x_2$ and $h = 2$ and no deterministic terms*
Then $K_1 > 0$, $K_2 > 0$. Since $\frac{\partial f(\boldsymbol{x})}{\partial X_1} = 2X_1X_2$. In order to get the growth condition satisfied the inequality $2|X_1X_2| \leq |X_1|^2 + |X_2|^2$ is used, so $\alpha_{1,1} = 2$ and $\alpha_{1,2} = 2$. Further $\alpha_{2,1} = 2$ and $\alpha_{2,2} = 0$. In this case, we are interested in knowing the limiting distribution of the term

$$\frac{1}{n^2}\sum_{k=1}^{n-1}\left[X_{1,n}\left(\frac{k}{n}\right)\right]^2 X_{2,n}\left(\frac{k}{n}\right) \cdot u_{2,k+1} \tag{22}$$

Note that, in the above example, condition (18) requires the existence of the eighth moment for $X_{1,n}$ and $X_{2,n}$ (or equivalently for $u_{1,n}$ and $u_{2,n}$). Alternatively, this condition could be replaced by an assumption involving the existence of moments for certain cross-product of $X_{1,n}$ and $X_{2,n}$. In particular $E(X_{1,n}^4 X_{2,n}^2)$. This is needed to get the bound in (40), in the proof of the main result. In other words, there is a trade off between the moments required for $X_{1,n}$ and $X_{2,n}$ individually, in the above example up to the eight moment, and the moments required for the cross-product $X_{1,n} \cdot X_{2,n}$, up to the fourth moment in this example.

The following theorem represents the main result in this paper. It is a generalization of Theorem 1 into a higher dimensional case and more general functional form.

Theorem 2. *Let $f : \mathbb{R}^h \times [0, 1] \to \mathbb{R}$ be a function satisfying Assumption 1. Suppose that Assumptions 2 and 4 are satisfied. Then, on $D^h[0, 1] \times \mathbb{R}$,*

$$\left(\frac{1}{\sqrt{n}}\boldsymbol{X}_n, \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f\left[\frac{1}{\sqrt{n}}\boldsymbol{X}_n\left(\frac{k}{n}\right), \frac{k}{n}\right] u_{h,k+1}\right) \xrightarrow{d} \left(\boldsymbol{B}, \int_0^1 f\left[\boldsymbol{B}(u), u\right] \, dB_h(u)\right) \tag{23}$$

Note that Theorem 1 is a special case of Theorem 2, in which only X_1 was included in f. Additionally, Theorem 2 represents a generalization of the work by Chan and Wei (1988), Ibragimov and Phillips (2004) and Strasser (1986) as discussed in the introduction.

Remark 1. *The proof of Theorem 2 is based on the convergence in probability of the Skorokhod embedding of $\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f[\frac{1}{\sqrt{n}}\boldsymbol{X}_n\left(\frac{k}{n}\right), \frac{k}{n}]u_{h,k+1}$. Thus (23) can be further generalized in the sense that, if there are h functions so that $f_s : \mathbb{R}^h \times [0, 1] \to \mathbb{R}$ satisfy Assumption 1, for all $s \in [[1, h]]$, and* *suppose that Assumptions 2 and 4 are satisfied for each function f_s. Then,*

$$\left(\frac{1}{\sqrt{n}}\boldsymbol{X}_n, \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f_1\left[\frac{1}{\sqrt{n}}\boldsymbol{X}_n\left(\frac{k}{n}\right), \frac{k}{n}\right]u_{1,k+1}, \ldots, \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f_h\left[\frac{1}{\sqrt{n}}\boldsymbol{X}_n\left(\frac{k}{n}\right), \frac{k}{n}\right]u_{h,k+1}\right)$$

$$\xrightarrow{d} \left(\boldsymbol{B}, \int_0^1 f_1\left[\boldsymbol{B}(u), u\right] dB_1(u), \ldots, \int_0^1 f_h\left[\boldsymbol{B}(u), u\right] dB_h(u)\right)$$

This last result is a consequence of the fact that each of the terms $\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f_s[\frac{1}{\sqrt{n}}\boldsymbol{X}_n(\frac{k}{n}), \frac{k}{n}]u_{s,k+1}$ converges in probability to $\int_0^1 f_s[\boldsymbol{B}(u), u]\, dB_s(u)$ for all $s \in [[1, h]]$. Thus the joint convergence in distribution.

The following remark introduces another result which is required in the analysis of some misspecification tests mentioned earlier - when these are applied to a marginally stable (i.e. with all its roots equal to unity) autoregressive process, with or without deterministic terms (e.g. constant or linear trend). This is due to the presence of sums of polynomial terms that appear for instance in White's test for heteroskedasticity (White, 1980).

Remark 2. *Suppose Assumptions 1, 2 and 4 are satisfied and let $g : \mathbb{R}^h \times [0, 1] \to \mathbb{R}$ be continuous. Then (23) holds jointly with*

$$g\left[\frac{1}{\sqrt{n}}\boldsymbol{X}_n\left(\frac{k}{n}\right), \frac{k}{n}\right] \xrightarrow{d} g\left[\boldsymbol{B}(u), u\right] \tag{24}$$

since applying g to $\boldsymbol{X}_n$ and the identity to $\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f[\frac{1}{\sqrt{n}}\boldsymbol{X}_n(\frac{k}{n}), \frac{k}{n}]u_{h,k+1}$, is continuous. As an example let $g[\boldsymbol{X}_n(t), t] = \frac{1}{n}\sum_{k=1}^{n-1}(f[\frac{1}{\sqrt{n}}\boldsymbol{X}_n(\frac{k}{n}), \frac{k}{n}])^2$ so that the denominator and the numerator of the least squares-type statistic

$$Z_n = \frac{\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f\left[\frac{1}{\sqrt{n}}\boldsymbol{X}_n\left(\frac{k}{n}\right), \frac{k}{n}\right]u_{h,k+1}}{\frac{1}{n}\sum_{k=1}^{n-1}\left(f\left[\frac{1}{\sqrt{n}}\boldsymbol{X}_n\left(\frac{k}{n}\right), \frac{k}{n}\right]\right)^2}$$

converge jointly, and hence Z_n converges.

3 Preliminary Lemmas

At this point, it is worth mentioning that the following two lemmas and a corollary are required in order to establish the proof for Theorem 2.

First of all, a key point in the argument used by Chan and Wei (1988) is that they work with linear terms, so they can square these terms and use the fact that the square martingale is uncorrelated. When dealing with squared term, Chan and Wei used the following result

$$E\left[\left(\sum_{t=1}^{N}\varepsilon_t\right)^2\right] = \sum_{t_1=1}^{N}\cdot\sum_{t_2=1\neq t_1}^{N} E(\varepsilon_{t_1}\varepsilon_{t_2}) + \sum_{r_1=1}^{N} E({\varepsilon_{r_1}}^2) = \sum_{r_1=1}^{N} E({\varepsilon_{r_1}}^2) \leq c \cdot N \tag{25}$$

given that, when $t_1 \neq t_2$, $E(\varepsilon_{t_1}\varepsilon_{t_2}) = E(\varepsilon_{\min\{t_1,t_2\}} \cdot E\{\varepsilon_{\max\{t_1,t_2\}}|\mathcal{F}_{\max\{t_1,t_2\}-1}\}) = 0$, and $E(\varepsilon_{r_1}^2) =$

$E\left(E\{\varepsilon_{r_1}^2|\mathcal{F}_{r_1-1}\}\right) \leq c$ for some constant $c > 0$. However, due to the functional terms and growth condition used in Theorem 2, we require the existence of higher order moments for the sum $\sum_{t=1}^{N} \varepsilon_t$. In fact, we are interested in the power $2p$ (p is a positive integer) of the above sum. Thus, a higher order martingale difference assumption and higher order moment bounds are needed, as stated in Assumption 4. The following lemma summarizes the abovementioned points in a more explicit manner.

Lemma 1. *Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence of martingale differences with respect to increasing σ-fields $\mathcal{F}_n$. If,*

$$E(\varepsilon_n^i|\mathcal{F}_{n-1}) \overset{a.s.}{=} c_i \qquad \text{for some constants } c_i, \quad i = 2,3,...,\min\{p+1, 2p-1\}, \tag{26}$$

$$E\left(\varepsilon_n^{2p}\right) \overset{a.s.}{\leq} c^* \qquad \text{for some constant } c^* > 0, \quad \text{for a fixed integer } p. \tag{27}$$

Then, there exists a constant $K_p > 0$ such that

$$E\left[\left(\sum_{t=1}^{N} \varepsilon_t\right)^{2p}\right] \leq K_p \cdot N^p \tag{28}$$

Note that when $p = 1$, equation (27) is reduced to condition (2.14) in Chan and Wei (1988), and equation (26) is then empty.

Proof of Lemma 1. : We have that

$$E\left[\left(\sum_{t=1}^{N} \varepsilon_t\right)^{2p}\right] = \sum_{t_1=1}^{N}\sum_{t_2=1}^{N}\cdots\sum_{t_{2p-1}=1}^{N}\sum_{t_{2p}=1}^{N} E(\varepsilon_{t_1}\varepsilon_{t_2}\ldots\varepsilon_{t_{2p-1}}\varepsilon_{t_{2p}}) \tag{29}$$

Suppose that there are k different indexes $s_1, ..., s_k$, each replicated $r_1, ..., r_k$ times in the above expectation. Thus for all $k \in [[1, 2p]]$, we have that $\sum_{i=1}^{k} r_i = 2p$. Hence,

$$E\left[\left(\sum_{t=1}^{N} \varepsilon_t\right)^{2p}\right] = \sum_{s_1=1}^{N}\sum_{s_2=1}^{N}\cdots\sum_{s_{k-1}=1}^{N}\sum_{s_k=1}^{N} E\left[\prod_{i=1}^{k} \varepsilon_{s_i}^{r_i}\right] \tag{30}$$

The bound (28) needs to be established in each of four different cases.
- **Case 1**: $k < p$

Then, from the triangle inequality and Hölder's inequality (c.f. Magnus and Neudecker, 1999), we obtain that

$$\left|E\left[\prod_{i=1}^{k} \varepsilon_{s_i}^{r_i}\right]\right| \leq E\left|\prod_{i=1}^{k} \varepsilon_{s_i}^{r_i}\right| \leq \prod_{i=1}^{k}\left(E\left|\varepsilon_{s_i}^{r_i}\right|^{\frac{2p}{r_i}}\right)^{\frac{r_i}{2p}} \qquad \text{where} \qquad \sum_{i=1}^{k}\frac{r_i}{2p} = 1.$$

By assumption (27), we have that

$$\prod_{i=1}^{k}\left(E\left|\varepsilon_{s_i}\right|^{2p}\right)^{\frac{r_i}{2p}} \;\leq\; \prod_{i=1}^{k}(c^*)^{\frac{r_i}{2p}} \;\leq\; c^*$$

Using (30) and the triangle inequality,

$$\left| E\left[\left(\sum_{t=1}^{N}\varepsilon_t\right)^{2p}\right]_{k<p}\right| \leq \sum_{s_1=1}^{N}\sum_{s_2=1}^{N}\cdots\sum_{s_{k-1}=1}^{N}\sum_{s_k=1}^{N}\left|E\left[\prod_{i=1}^{k}\varepsilon_{s_i}^{r_i}\right]\right|$$

Therefore, in combination with the bound established above, the expectation in (28) is of order $O\left(N^k\right) = o\left(N^p\right)$

- **Case 2**: $k > p$

In this case there is at least one group with one element, so a $j \in [[1, h]]$ exists so that $r_j = 1$, and the largest group size is p. Thus, by taking iterated expectations

$$\begin{aligned} E\left[\prod_{i=1}^{k}\varepsilon_{s_i}^{r_i}\right] &= E\left[\left(\prod_{i=1}^{j-1}\varepsilon_{s_i}^{r_i}\right)\varepsilon_{s_j}\left(\prod_{i=j+1}^{k}\varepsilon_{s_i}^{r_i}\right)\right] \\ &= E\left[\left(\prod_{i=1}^{j-1}\varepsilon_{s_i}^{r_i}\right)\varepsilon_{s_j}\, E\left(\prod_{i=j+1}^{k}\varepsilon_{s_i}^{r_i}|\mathcal{F}_{s_j}\right)\right] \end{aligned}$$

Due to the maximal group size and condition (26), the conditional expectations are constant. Thus

$$\begin{aligned} E\left[\prod_{i=1}^{k}\varepsilon_{s_i}^{r_i}\right] &= \left(\prod_{i=j+1}^{k}c_{r_i}\right)E\left[\left(\prod_{i=1}^{j-1}\varepsilon_{s_i}^{r_i}\right)\varepsilon_{s_j}\right] \\ &= \left(\prod_{i=j+1}^{k}c_{r_i}\right)E\left[\left(\prod_{i=1}^{j-1}\varepsilon_{s_i}^{r_i}\right)E\left(\varepsilon_{s_j}|\mathcal{F}_{s_j-1}\right)\right] \end{aligned}$$

which is zero by the martingale difference assumption. Thus the expectation in (28) is zero in this case.

- **Case 3**: $k = p$ and there is a singleton group

The argument is the same as in Case 2, noting that the longest group size is now $p+1$. Thus the expectation in (28) is again zero.

- **Case 4:** $k = p$ and there are no singletons

In this case, all the s_i's arranged in pairs (i.e. $\forall i \in [[1, k]]$, $r_i = 2$). Using the law of iterated expectation successively to obtain that

$$E(\varepsilon_{t_1}\varepsilon_{t_2}\ldots\varepsilon_{t_{2p-1}}\varepsilon_{t_{2p}})_p = E\left[\prod_{i=1}^{p}\varepsilon_{s_i}^2\right] = \prod_{i=1}^{p}E\{\varepsilon_{s_i}^2|\mathcal{F}_{s_i-1}\} = c_2^p$$

using the moment conditions presented in equation (26), i.e. $E({\varepsilon_n}^2|\mathcal{F}_{n-1}) = c_2$. The number of possible combinations in which we can order the different pairs represented by s_1, s_2, ..., s_p is given by

$$a_p \cdot \binom{N}{1} \times \binom{N-1}{1} \times \ldots \times \binom{N-p+1}{1} = a_p \cdot \frac{N!}{(N-p)!} = O\left(N^p\right)$$

where a_p is a constant for a given p. Hence the bound in (28) is satisfied. □

As mentioned earlier, Theorem 2 allows for non-linearities to be included via the function f which satisfies the growth condition. Thus, concerning the proof of Theorem 2, we are interested in the difference $|f(Y) - f(X)|$ (for some X and Y in the domain of f) rather than the simple difference $|Y - X|$ used in the proof of Chan and Wei (1988). Thus, the need for the following lemma.

Lemma 2. *Let $f : \mathbb{R}^{h+1} \to \mathbb{R}$ satisfy Assumption 1. Then, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{h+1}$, such that $\boldsymbol{x} = (x_1, \dots, x_{h+1})'$ and $\boldsymbol{y} = (y_1, \dots, y_{h+1})'$, we have*

$$|f(\boldsymbol{y}) - f(\boldsymbol{x})| \leq \sum_{r=1}^{h+1} K_r \left[1 + \sum_{j=1}^{h+1} 2^{(\alpha_{r,j}-1)} \left(|x_j|^{\alpha_{r,j}} + |y_j - x_j|^{\alpha_{r,j}} \right) \right] |y_r - x_r| \tag{31}$$

$$|f(\boldsymbol{y}) - f(\boldsymbol{x})|^2 \leq 2(h+1) \sum_{r=1}^{h+1} K_r^2 \left[1 + (h+1) \sum_{j=1}^{h+1} 2^{(2\alpha_{r,j}-1)} \left(|x_j|^{2\alpha_{r,j}} + |y_j - x_j|^{2\alpha_{r,j}} \right) \right] |y_r - x_r|^2 \tag{32}$$

Proof of Lemma 2. : First of all, from Jensen's inequality we have that for any positive integers a and N,

$$\left(\sum_{i=1}^{N} X_i \right)^a \leq N^{a-1} \sum_{i=1}^{N} {X_i}^a \tag{33}$$

Since f is differentiable, the mean value theorem and the triangle inequality give, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{h+1}$,

$$|f(\mathbf{y}) - f(\mathbf{x})| = \left| \sum_{r=1}^{h+1} \frac{\partial f(\mathbf{x} + \boldsymbol{\delta})}{\partial x_r} (y_r - x_r) \right| \leq \sum_{r=1}^{h+1} \left| \frac{\partial f(\mathbf{x} + \boldsymbol{\delta})}{\partial x_r} \right| \cdot |y_r - x_r| \tag{34}$$

where $\boldsymbol{\delta} = (\delta_1, \dots, \delta_{h+1})' \in \mathbb{R}^{h+1}$ such that $\forall i \in [[1, h+1]]$, $|\delta_i| \leq |y_i - x_i|$. Using first the growth condition in Assumption 1 and then the inequality (33), we obtain that

$$\left| \frac{\partial f(\mathbf{x} + h)}{\partial x_r} \right| \leq K_r \left[1 + \sum_{j=1}^{h+1} |x_j + \delta_j|^{\alpha_{r,j}} \right] \leq K_r \left[1 + \sum_{j=1}^{h+1} 2^{(\alpha_{r,j}-1)} \left(|x_j|^{\alpha_{r,j}} + |\delta_j|^{\alpha_{r,j}} \right) \right]$$

Inserting this in (34) gives (31). We recover the result in (32) by using the inequality (33) repeatedly to inequality (31). First, apply (33) to the sum in r

$$|f(\mathbf{y}) - f(\mathbf{x})|^2 \leq (h+1) \sum_{r=1}^{h+1} {K_r}^2 \left[1 + \sum_{j=1}^{h+1} 2^{(\alpha_{r,j}-1)} \left(|x_j|^{\alpha_{r,j}} + |\delta_j|^{\alpha_{r,j}} \right) \right]^2 \cdot |y_r - x_r|^2$$

Then apply (33) to the sum in squared bracket

$$|f(\mathbf{y}) - f(\mathbf{x})|^2 \leq 2(h+1) \sum_{r=1}^{h+1} {K_r}^2 \left[1 + \left(\sum_{j=1}^{h+1} 2^{(\alpha_{r,j}-1)} \left(|x_j|^{\alpha_{r,j}} + |\delta_j|^{\alpha_{r,j}} \right) \right)^2 \right] \cdot |y_r - x_r|^2$$

Then apply (33) to the sum in j

$$|f(\mathbf{y}) - f(\mathbf{x})|^2 \leq 2(h+1) \sum_{r=1}^{h+1} {K_r}^2 \left[1 + (h+1) \sum_{j=1}^{h+1} 2^{2(\alpha_{r,j}-1)} \left(|x_j|^{\alpha_{r,j}} + |\delta_j|^{\alpha_{r,j}} \right)^2 \right] \cdot |y_r - x_r|^2$$

Finally, apply (33) to the square terms involving x_j and δ_j. This gives (32). $\square$

4 Proof of Theorem 2

Having established the proof of Lemma 1 and Lemma 2, we can now proceed to the proof of Theorem 2. Also, note that Theorem 1 is a special case of Theorem 2 so it is proved thereby.

Proof of Theorem 2. : From Billingsley (1968) we know that since B_1, B_2, ... , B_h have continuous paths, the convergence in the Skorokhod topology is equivalent to the uniform convergence. Additionally, D can be equipped with a complete metric so that the induced topology is equivalent to the Skorokhod topology (Billingsley, 1968). Then, by the Skorokhod representation theorem (Skorokhod, 1956), there are a probability space Ω and random elements $\tilde{U}_{1,n}$, ... ,$\tilde{U}_{h,n}$ in $D[0,1]$ such that

$$\|(\tilde{U}_{1,n}, \ldots, \tilde{U}_{h,n}) - (B_1, \ldots, B_h)\|_\infty \longrightarrow 0 \ a.s. \tag{35}$$

and

$$\tilde{\mathbf{U}}_n = (\tilde{U}_{1,n}, \ldots, \tilde{U}_{h,n}) \stackrel{d}{=} n^{-1/2}(X_{1,n}, \ldots, X_{h,n}) = n^{-1/2}\mathbf{X}_n$$

let

$$G_{h,n} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} f\left[\frac{X_{1,n}\left(\frac{k}{n}\right)}{\sqrt{n}}, \ldots, \frac{X_{h,n}\left(\frac{k}{n}\right)}{\sqrt{n}}, \frac{k}{n}\right] u_{h,k+1}$$

and

$$\tilde{G}_{h,n} = \sum_{k=1}^{n-1} f\left[\tilde{\mathbf{U}}_n\left(\frac{k}{n}\right), \frac{k}{n}\right] \cdot \left[\tilde{U}_{h,n}\left(\frac{k+1}{n}\right) - \tilde{U}_{h,n}\left(\frac{k}{n}\right)\right]$$

Thus

$$(\tilde{\mathbf{U}}_n, \tilde{G}_{h,n}) = (\tilde{U}_{1,n}, \ldots, \tilde{U}_{h,n}, \tilde{G}_{h,n}) \stackrel{d}{=} (n^{-1/2}X_{1,n}, \ldots, n^{-1/2}X_{h,n}, G_{h,n})$$

Now, from equation (35) and Egorov's theorem (c.f. Billingsley, 1995), given $\varepsilon > 0$, there is an event $\Omega_\varepsilon \subset \Omega$ such that $P(\Omega_\varepsilon) \geq 1 - \varepsilon$ and

$$\sup\{\|(\tilde{U}_{1,n}(\omega), \ldots, \tilde{U}_{h,n}(\omega)) - (B_1(\omega), \ldots, B_h(\omega))\|_\infty : \omega \in \Omega_\varepsilon\} = \delta_n \longrightarrow 0 \tag{36}$$

where δ_n is a sequence of constants. Then, we can take integers $N(n) \longrightarrow \infty$ such that

$$N(n) \cdot \delta_n \longrightarrow 0 \quad and \quad \frac{N(n)}{n} \longrightarrow 0 \tag{37}$$

Following Chan and Wei (1988), for each n, we can choose a partition $\{t_0, ..., t_N(n)\}$ of $[0,1]$ such that

$$\begin{gathered} 0 = t_0 < t_1 = \frac{n_1}{n} < t_2 = \frac{n_2}{n} < ... < t_{N(n)} = \frac{n_{N(n)}}{n} = 1, \\ \max\{|t_{i+1} - t_i| : 0 \leq i \leq N(n) - 1\} = o(1). \end{gathered} \tag{38}$$

In order to prove the result shown in equation (23), it is sufficient to show that,

$$\tilde{G}_{h,n} \xrightarrow{P} \int_0^1 f\left[B_1(u), \ldots, B_h(u), u\right] dB_h(u)$$

In fact, we follow closely the argument of Chan and Wei (1988) in that we can write

$$\tilde{G}_{h,n} - \int_0^1 f\left[B_1(u), \ldots, B_h(u), u\right] dB_h(u) = J_{h,n} + H_{h,n} + L_{h,n} + M_{h,n}$$

where,

$$J_{h,n} = \tilde{G}_{h,n} - \sum_{k=1}^{N(n)} f\left[\tilde{\mathbf{U}}_n(t_{k-1}), t_{k-1}\right] \cdot \left[\tilde{U}_{h,n}(t_k) - \tilde{U}_{h,n}(t_{k-1})\right]$$

$$H_{h,n} = \sum_{k=1}^{N(n)} \left(f\left[\tilde{\mathbf{U}}_n(t_{k-1}), t_{k-1}\right] - f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right]\right) \cdot I_{\Omega_\varepsilon} \cdot \left[\tilde{U}_{h,n}(t_k) - \tilde{U}_{h,n}(t_{k-1})\right]$$

$$L_{h,n} = \sum_{k=1}^{N(n)} f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right] \cdot I_{\Omega_\varepsilon} \cdot \left(\left[\tilde{U}_{h,n}(t_k) - \tilde{U}_{h,n}(t_{k-1})\right] - \left[B_h(t_k) - B_h(t_{k-1})\right]\right)$$

$$M_{h,n} = \sum_{k=1}^{N(n)} f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right] \cdot \left[B_h(t_k) - B_h(t_{k-1})\right] - \int_0^1 f\left[\mathbf{B}(t), t\right] dB_h(t)$$

where $\forall t \in [0,1]$, $\mathbf{B}(t) = (B_1(t), B_2(t), \ldots, B_h(t))$.

Thus, the proof of Theorem 2 completed by showing that each of the four terms $J_{h,n}$, $H_{h,n}$, $L_{h,n}$ and $M_{h,n}$ is of order $o_p(1)$. This is done in the following four lemmas. □

Lemma 3. *If Assumptions 1 and 4 are satisfied. Then,* $J_{h,n} = o_p(1)$.

The proof of this Lemma is based on Chan an Wei (1988). This is where the higher order martingale difference sequence assumptions and Lemma 1 are needed.

Proof of Lemma 3. : We have

$$J_{h,n} = \sum_{k=1}^{N(n)} \sum_{i=n_{k-1}}^{n_k-1} A_{k,i} R_{i+1}$$

where

$$\begin{aligned} A_{k,i} &= f\left[\tilde{\mathbf{U}}_n\left(\frac{i}{n}\right), \frac{i}{n}\right] - f\left[\tilde{\mathbf{U}}_n(t_{k-1}), t_{k-1}\right] \\ R_{i+1} &= \tilde{U}_{h,n}\left(\frac{i+1}{n}\right) - \tilde{U}_{h,n}\left(\frac{i}{n}\right) \end{aligned}$$

so $A_{k,i}$ is $\mathcal{F}_i$-measurable, and R_i is a $\mathcal{F}_i$-martingale difference sequence. Therefore

$$E\left(J_{h,n}^2\right) = \sum_{k=1}^{N(n)} \sum_{i=n_{k-1}}^{n_k-1} E\left[A_{k,i}^2 R_{i+1}^2\right] = \sum_{k=1}^{N(n)} \sum_{i=n_{k-1}}^{n_k-1} E\left[A_{k,i}^2 \cdot E\left(R_{i+1}^2 | \mathcal{F}_i\right)\right]$$

Due to (21) in Assumption 4, we have $E\left(R_{i+1}^2 | \mathcal{F}_i\right) \overset{a.s.}{\leq} n^{-1} c_{h,2}^*$. If in addition it is shown, as suggested by Chan and Wei (1988), that

$$E\left(A_{k,i}^{2}\right) \leq K_{\alpha}\cdot\left(\frac{n_k}{n}-\frac{n_{k-1}}{n}\right) \tag{39}$$

where K_α is a positive constant, then

$$E\left(J_{s,n}^{2}\right) \leq K_{\alpha}\frac{c_{h,2}^{*}}{n}\sum_{k=1}^{N(n)}\sum_{i=n_{k-1}}^{n_k-1}\left(\frac{n_k}{n}-\frac{n_{k-1}}{n}\right) \leq K_{\alpha}\cdot c_{h,2}^{*}\cdot\max_{1\leq k\leq N(n)}\{t_k-t_{k-1}\}\longrightarrow 0$$

from equation (38). The desired result will then follow from Chebychev's inequality. Hence, it is now sufficient to show that equation (39) is satisfied. Now, to prove (39) the result (32) in Lemma 2 is applied to $A_{k,i}^2$ which gives

$$A_{k,i}^{2} \;\leq\; 2(h+1)\sum_{r=1}^{h+1}K_r^2\left[1+(h+1)\sum_{j=1}^{h+1}2^{(2\alpha_{r,j}-1)}\left(C_j^{2\alpha_{r,j}}+D_j^{2\alpha_{r,j}}\right)\right]D_r^2$$

where, for $j\in[[1,h]]$

$$C_j=\tilde{U}_{j,n}\left(t_{k-1}\right) \qquad \text{and} \qquad D_j=\tilde{U}_{j,n}\left(\frac{i}{n}\right)-\tilde{U}_{j,n}\left(t_{k-1}\right)$$

whereas $j=h+1$ gives terms related to the deterministic terms

$$C_{h+1}=\frac{n_{k-1}}{n} \qquad \text{and} \qquad D_{h+1}=\frac{i}{n}-\frac{n_{k-1}}{n}$$

The summands for $r\in[[1,h]]$ and for $r=h+1$ are treated separately.

First, let $r\in[[1,h]]$. If $K_r=0$ there is no contribution. Thus, for $K_r>0$ bounds are needed for $E(D_r^2)$, $E(C_j^{2\alpha_{r,j}}D_r^2)$ and $E(D_j^{2\alpha_{r,j}}D_r^2)$. Starting with the latter, for $j=r$

$$E\left(D_r^{2\alpha_{r,r}+2}\right)\leq c_{r,r}^{*}\left(\frac{n_k}{n}-\frac{n_{k-1}}{n}\right)$$

by (15) and (16) in Assumption 4 and Lemma 1, and noting that $(i-n_{k-1})\leq(n_k-n_{k-1})$.

For $j\in[[1,h]]$ so that $j\neq r$ and $\alpha_{r,j}\neq 0$, the Cauchy-Schwarz inequality gives

$$\left[E\left(D_j^{2\alpha_{r,j}}D_r^2\right)\right]^2\leq E\left(D_j^{4\alpha_{r,j}}\right)\cdot E\left(D_r^4\right)\leq c_{r,j}^{*}c_{r,r}^{*}\left(\frac{n_k}{n}-\frac{n_{k-1}}{n}\right)^2 \tag{40}$$

by conditions (15) to (18) in Assumption 4 and Lemma 1.

For $j=h+1$, or $j\in[[1,h]]$ so that $j\neq r$ and $\alpha_{r,j}=0$, then it suffices that

$$E\left(D_r^2\right)\leq c_{r,r}^{*}\left(\frac{n_k}{n}-\frac{n_{k-1}}{n}\right) \tag{41}$$

by (16) in Assumption 4 and Lemma 1.

Continue with $E(C_j^{2\alpha_{r,j}}D_r^2)$. If $j\in[[1,h]]$ and $\alpha_{r,j}\neq 0$

$$E\left(C_j^{2\alpha_{r,j}}D_r^2\right)=E\left[C_j^{2\alpha_{r,j}}\cdot E\left(D_r^2|\mathcal{F}_{n_{k-1}}\right)\right]$$

Now,

$$E\left(D_r^2|\mathcal{F}_{n_{k-1}}\right) = c_{r,2}\left(\frac{i}{n} - \frac{n_{k-1}}{n}\right)$$

by (15) in Assumption 4 and Lemma 1, whereas $E(C_j^{2\alpha_{r,j}})$ is bounded by conditions (15) to (18) in Assumption 4 and Lemma 1. For the case $j = h+1$, and likewise for the term $E(D_r^2)$, the desired result follows from (41).

Secondly, let $r = h+1$. Again, no bounds are needed if $K_{h+1} = 0$. So, let $K_{h+1} > 0$. In this situation it suffices that $E(C_j^{2\alpha_{h+1,j}})$ and $E(D_j^{2\alpha_{h+1,j}})$ are bounded. For $j = h+1$ this is trivial, whereas for $j \in [[1, h]]$ this follows from (19) and (20) in Assumption 4 and Lemma 1. □

Lemma 4. *If Assumptions 1, 2 and condition (21) in Assumption 4 are satisfied. Then,* $H_{h,n} = o_p(1)$.

Proof of Lemma 4. : We have

$$H_{h,n} = \sum_{k=1}^{N(n)} A_k I_{\Omega_\varepsilon} F_k$$

where

$$\begin{aligned} A_k &= f\left[\tilde{\mathbf{U}}_n(t_{k-1}), t_{k-1}\right] - f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right] \\ F_k &= \tilde{U}_{h,n}(t_k) - \tilde{U}_{h,n}(t_{k-1}) \end{aligned}$$

From the triangle inequality we have

$$|H_{h,n}| = \left|\sum_{k=1}^{N(n)} A_k I_{\Omega_\varepsilon} F_k\right| \leq \sum_{k=1}^{N(n)} |A_k I_{\Omega_\varepsilon} F_k|$$

Now, applying (31) in Lemma 2 to A_k, we obtain that

$$|A_k| \leq \sum_{r=1}^{h+1} K_r \left[1 + \sum_{j=1}^{h+1} 2^{(\alpha_{r,j}-1)}\left(|C_j|^{\alpha_{r,j}} + |D_j|^{\alpha_{r,j}}\right)\right] \cdot |D_r|$$

where, for $j \in [[1, h]]$

$$C_j = B_j(t_{k-1}) \qquad \text{and} \qquad D_j = \tilde{U}_{j,n}(t_{k-1}) - B_j(t_{k-1})$$

whereas $j = h+1$ gives the terms related to the deterministic terms

$$C_{h+1} = t_{k-1} \qquad \text{and} \qquad D_{h+1} = 0$$

Therefore, there is no contribution from the case $r = h+1$. Hence, $A_{k,h+1} = 0$ and

$$|H_{h,n}| \leq \sum_{k=1}^{N(n)} \sum_{r=1}^{h} K_r \left[1 + \sum_{j=1}^{h+1} 2^{(\alpha_{r,j}-1)}\left(|C_j|^{\alpha_{r,j}} + |D_j|^{\alpha_{r,j}}\right)\right] \cdot |D_r|\, I_{\Omega_\varepsilon}\, |F_k|$$

Now, from equation (36) we have the bound $I_{\Omega_\varepsilon}|D_r| \leq \delta_n$. Thus, we obtain that

$$|H_{h,n}| \;\; \leq \;\; \delta_n \sum_{k=1}^{N(n)} \sum_{r=1}^{h} K_r \cdot G_{k,r} \; |F_k|$$

where,

$$G_{k,r} = 1 + \sum_{j=1}^{h} 2^{(\alpha_{r,j}-1)} \left(|C_j|^{\alpha_{r,j}} + {\delta_n}^{\alpha_{r,j}} \right) + 2^{(\alpha_{r,h+1}-1)} t_{k-1}^{\alpha_{r,h+1}}$$

Taking expectations gives,

$$E|H_{h,n}| \;\; \leq \;\; \delta_n \sum_{k=1}^{N(n)} \sum_{r=1}^{h} K_r \cdot E\left\{ G_{k,r}\, |F_k| \right\}$$

Now, from the Cauchy-Schwarz inequality $E\{G_{k,r}\,|F_k|\} \leq (E[G_{k,r}^2] \cdot E[F_k^2])^{1/2}$. Due to (21) in Assumption 4, $E[F_k^2] \leq c^*_{h,2}(t_k - t_{k-1}) \leq c^*_{h,2}$. Furthermore, $E[G_{k,r}^2]$ involves powers of the Brownian motion and thus has finite moments. Therefore, for some constant $K > 0$, it follows that

$$E|H_{h,n}| \;\; \leq \;\; \left(\sum_{r=1}^{h} K_r \right) \cdot K \cdot N(n) \cdot \delta_n \longrightarrow 0$$

from equation (37). Thus, by the Markov inequality we obtain the desired result. □

Lemma 5. *If Assumptions 1 and 2 are satisfied. Then,* $L_{h,n} = o_p(1)$.

Note that when dealing with the term $L_{h,n}$, the corresponding proof was not explicitly shown by Chan and Wei (1988). This is due to triviality of this step in the proof when f is a linear function. However, for a more general non-linear function f like the one considered here, the following proof is in order.

Proof of Lemma 5. : First, summation by parts gives

$$L_{h,n} = I_{\Omega_\varepsilon} \sum_{k=1}^{N(n)} f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right] \cdot \left(\left[\tilde{U}_{h,n}(t_k) - B_h(t_k)\right] - \left[\tilde{U}_{h,n}(t_{k-1}) - B_h(t_{k-1})\right] \right) = \tilde{L}_{1,n} - \tilde{L}_{2,n}$$

where

$$\begin{aligned} \tilde{L}_{1,n} &= I_{\Omega_\varepsilon}\, f[\mathbf{B}(1), 1] \cdot \left[\tilde{U}_{h,n}(1) - B_h(1)\right] \\ \tilde{L}_{2,n} &= I_{\Omega_\varepsilon} \sum_{k=1}^{N(n)} \left[\tilde{U}_{h,n}(t_k) - B_h(t_k)\right] \cdot \left(f\left[\mathbf{B}(t_k), t_k\right] - f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right] \right) \end{aligned}$$

Hence, the proof of Lemma 5 consist of showing that both $\tilde{L}_{1,n}$ and $\tilde{L}_{2,n}$ are $o_p(1)$. First, from (36), we have

$$E|\tilde{L}_{1,n}| = E\left| f[\mathbf{B}(1), 1]\, I_{\Omega_\varepsilon} \left[\tilde{U}_{h,n}(1) - B_h(1)\right] \right| \;\; \leq \;\; E\left| f[\mathbf{B}(1), 1] \right| \cdot \delta_n \longrightarrow 0$$

for a twice continuously differentiable vector function $f : \mathbb{R}^h \times [0,1] \to \mathbb{R}$. Thus, $\tilde{L}_{1,n} = o_p(1)$.

Secondly, using the triangle inequality and (36), respectively, we obtain that

$$|\tilde{L}_{2,n}| \leq \sum_{k=1}^{N(n)} \left|\tilde{U}_{h,n}(t_k) - B_h(t_k)\right| I_{\Omega_\varepsilon} |A_k| \leq \delta_n \sum_{k=1}^{N(n)} |A_k|$$

where

$$A_k = f\left[\mathbf{B}(t_k), t_k\right] - f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right]$$

Thus, taking expectations

$$E|\tilde{L}_{2,n}| \leq \delta_n \sum_{k=1}^{N(n)} E\,|A_k| \tag{42}$$

Now, applying (31) in Lemma 2 to A_k, we obtain that

$$|A_k| \leq \sum_{r=1}^{h+1} K_r \left[1 + \sum_{j=1}^{h+1} 2^{(\alpha_{r,j}-1)} \left(|C_j|^{\alpha_{r,j}} + |D_j|^{\alpha_{r,j}}\right)\right] \cdot |D_r|$$

where, for $j \in [[1,h]]$

$$C_j = B_j(t_{k-1}) \qquad \text{and} \qquad D_j = B_j(t_k) - B_j(t_{k-1})$$

whereas $j = h+1$ gives terms related to the deterministic terms

$$C_{h+1} = t_{k-1} \qquad \text{and} \qquad D_{h+1} = t_k - t_{k-1}$$

Therefore,

$$\begin{aligned} E|A_k| &\leq \sum_{r=1}^{h+1} K_r \cdot E\left\{\left[1 + \sum_{j=1}^{h+1} 2^{(\alpha_{r,j}-1)} \left(|C_j|^{\alpha_{r,j}} + |D_j|^{\alpha_{r,j}}\right)\right] \cdot |D_r|\right\} \\ &\leq \sum_{r=1}^{h+1} K_r \left(E\left[1 + \sum_{j=1}^{h+1} 2^{(\alpha_{r,j}-1)} \left(|C_j|^{\alpha_{r,j}} + |D_j|^{\alpha_{r,j}}\right)\right]^2 \cdot E\left[D_r^2\right]\right)^{1/2} \end{aligned}$$

from the Cauchy-Schwarz inequality.

If $K_r = 0$ there is no contribution. Thus, for $K_r > 0$ the properties of the Brownian motion imply that both expectations are bounded. In particular, for $r \in [[1,h]]$,

$$E\left(D_r^2\right) = E\left([B_r(t_k) - B_r(t_{k-1})]^2\right) = (t_k - t_{k-1}) \leq 1$$

While $D_{h+1}^2 = (t_k - t_{k-1})^2 \leq 1$. Therefore,

$$E|A_k| \leq \tilde{K}\left(\sum_{r=1}^{h+1} K_r\right) \tag{43}$$

where $\tilde{K}$ is a positive constant.

Hence, combining (42) and (43), we obtain that

$$E|\tilde{L}_{2,n}| \quad \leq \quad \tilde{K}\left(\sum_{r=1}^{h} K_r\right)\cdot N(n)\cdot\delta_n \longrightarrow 0$$

from (37). Then, by the Markov inequality, we have that $\tilde{L}_{2,n} = o_p(1)$. Thus $L_{h,n} = o_p(1)$. □

Lemma 6. *If Assumption 1 is satisfied. Then,* $M_{h,n} = o_p(1)$.

Proof of Lemma 6. : First, we can write

$$\begin{aligned} M_{h,n} &= \sum_{k=1}^{N(n)} f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right]\cdot\left[B_h(t_k) - B_h(t_{k-1})\right] - \int_0^1 f\left[B(t), t\right]\, dB_h(t) \\ &= \sum_{k=1}^{N(n)} \int_{t_{k-1}}^{t_k} \left(f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right] - f\left[\mathbf{B}(t), t\right]\right) dB_h(t) \end{aligned}$$

Therefore,

$$E\left(M_{h,n}^2\right) \;=\; E\left(\sum_{k=1}^{N(n)} \int_{t_{k-1}}^{t_k} \left(f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right] - f\left[\mathbf{B}(t), t\right]\right) dB_h(t)\right)^2 \;=\; \sum_{k=1}^{N(n)} \int_{t_{k-1}}^{t_k} E\left(A_k^2\right)\, dt$$

where $A_k = f\left[\mathbf{B}(t), t\right] - f\left[\mathbf{B}(t_{k-1}), t_{k-1}\right]$. Again, following Chan and Wei (1988), if

$$E\left(A_k^2\right) \leq \bar{K}\cdot(t_k - t_{k-1}) \tag{44}$$

where $\bar{K}$ is a positive constant, then

$$E\left(M_{h,n}^2\right) \leq \bar{K}\cdot\sum_{k=1}^{N(n)} \int_{t_{k-1}}^{t_k} (t_k - t_{k-1})\, dt \leq \bar{K}\cdot\max\{(t_k - t_{k-1})\} = o(1)$$

from equation (38). The desired result will then follow from the Chebyshev's inequality. In other words, to prove Lemma 6 it is sufficient to show (44). Now, applying (32) in Lemma 2 to A_k^2, we obtain that

$$A_k^2 \leq 2(h+1)\sum_{r=1}^{h+1} K_r^2 \left[1 + (h+1)\sum_{j=1}^{h+1} 2^{(2\alpha_{r,j}-1)}\left(C_j^{2\alpha_{r,j}} + D_j^{2\alpha_{r,j}}\right)\right] D_r^2$$

where, for $j \in [[1, h]]$

$$C_j = B_j(t_{k-1}) \qquad \text{and} \qquad D_j = B_j(t) - B_j(t_{k-1})$$

whereas $j = h+1$ gives terms related to the deterministic terms

$$C_{h+1} = t_{k-1} \qquad \text{and} \qquad D_{h+1} = t - t_{k-1}$$

Therefore,

$$
\begin{aligned}
E\left(A_k^2\right) &\leq 2(h+1)\sum_{r=1}^{h+1} K_r^2 \cdot E\left\{\left[1+(h+1)\sum_{j=1}^{h+1} 2^{(2\alpha_{r,j}-1)}\left(C_j^{2\alpha_{r,j}}+D_j^{2\alpha_{r,j}}\right)\right] D_r^2\right\} \\
&\leq 2(h+1)\sum_{r=1}^{h+1} K_r^2 \left(E\left[1+(h+1)\sum_{j=1}^{h+1} 2^{(2\alpha_{r,j}-1)}\left(C_j^{2\alpha_{r,j}}+D_j^{2\alpha_{r,j}}\right)\right]^2 \cdot E\left[D_r^4\right]\right)^{1/2}
\end{aligned}
$$

from the Cauchy-Schwarz inequality.

Exploiting the properties of the Brownian motion, both expectations are found to be finite. In particular it holds, for $r \in [[1,h]]$,

$$
E\left(D_r^4\right) = E\left([B_r(t_{k-1}) - B_r(t)]^4\right) \leq K_B^2 \cdot (t_{k-1}-t)^2
$$

where $K_B > 0$ is a constant, and, when $r = h+1$, $D_{h+1}^4 = (t - t_{k-1})^4$.

Finally, noting that $(t - t_{k-1}) \leq (t_k - t_{k-1}) \leq 1$, we obtain (44). □

References

[1] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[2] Billingsley, P. (1995). *Probability and Measure.* Wiley, New York.

[3] Chan, N.H. and Wei, C.Z. (1988). Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes. *Annals of Statistics* 16, 367-401.

[4] Ibragimov, R. and Phillips, P.C.B. (2004). Regression Asymptotics Using Martingale Convergence Methods. Cowles Foundation Discussion paper No. 1473.

[5] Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 2nd Edition. Wiley, New York.

[6] Skorokhod, A.V. (1956). Limit theorems for stochastic processes. *Theory Probab. Appl.* 1 261-290.

[7] Strasser, H. (1986). Martingale difference arrays and stochastic integrals. *Probability Theory and Related Fields* 72, 83-98.

[8] White, H. (1980). A Heteroskedasticity Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*, 48, 817-838.

[9] White, H. (2000). *Asymptotic Theory for Econometricians.* Academic Press.

[10] Wooldridge, J.M. (1999). Asymptotic properties of some specification tests in linear models with integrated processes. In R.F. Engle and H. White, *Cointegration, Causality and Forecasting: Festschrift in Honour of Clive W.J. Granger*. Oxford University Press.